Mapping Soil Organic Carbon for Airborne and Simulated EnMAP Imagery Using the LUCAS Soil Database and a Local PLSR

Kathrin J. Ward [1,*], **Sabine Chabrillat** [1], **Maximilian Brell** [1], **Fabio Castaldi** [2], **Daniel Spengler** [1] **and Saskia Foerster** [1]

[1] Helmholtz Center Potsdam, GFZ German Research Center for Geosciences, Section 1.4 Remote Sensing and Geoinformatics, Telegrafenberg A17, 14473 Potsdam, Germany; chabri@gfz-potsdam.de (S.C.); maximilian.brell@gfz-potsdam.de (M.B.); daniel.spengler@gfz-potsdam.de (D.S.); saskia.foerster@gfz-potsdam.de (S.F.)

[2] Research Institute for Agriculture, Fisheries and Food—ILVO, Burgemeester Van Gansberghelaan 92 box 1, 9820 Merelbeke, Belgium; Fabio.Castaldi@ilvo.vlaanderen.be

* Correspondence: ward@gfz-potsdam.de

check for updates

Abstract: Soil degradation is a major threat for European soils and therefore, the European Commission recommends intensifying research on soil monitoring to capture changes over time and space. Imaging spectroscopy is a promising technique to create spatially accurate topsoil maps based on hyperspectral remote sensing data. We tested the application of a local partial least squares regression (PLSR) to airborne HySpex and simulated satellite EnMAP (Environmental Mapping and Analysis Program) data acquired in north-eastern Germany to quantify the soil organic carbon (SOC) content. The approach consists of two steps: (i) the local PLSR uses the European LUCAS (land use/cover area frame statistical survey) Soil database to quantify the SOC content for soil samples from the study site in order to avoid the need for wet chemistry analyses, and subsequently (ii) a remote sensing model is calibrated based on the local PLSR SOC results and the corresponding image spectra. This two-step approach is compared to a traditional PLSR approach using measured SOC contents from local samples. The prediction accuracy is high for the laboratory model in the first step with R^2 = 0.86 and RPD = 2.77. The HySpex airborne prediction accuracy of the traditional approach is high and slightly superior to the two-step approach (traditional: R^2 = 0.78, RPD = 2.19; two-step: R^2 = 0.67, RPD = 1.79). Applying the two-step approach to simulated EnMAP imagery leads to a lower but still reasonable prediction accuracy (traditional: R^2 = 0.77, RPD = 2.15; two-step: R^2 = 0.48, RPD = 1.41). The two-step models of both sensors were applied to all bare soils of the respective images to produce SOC maps. This local PLSR approach, based on large scale soil spectral libraries, demonstrates an alternative to SOC measurements from wet chemistry of local soil samples. It could allow for repeated inexpensive SOC mapping based on satellite remote sensing data as long as spectral measurements of a few local samples are available for model calibration.

Keywords: soil organic carbon; mapping; hyperspectral; EnMAP; LUCAS; localPLSR

1. Introduction

Soil degradation is a serious concern worldwide with implications for climate change and food security [1]. Soil organic carbon (SOC), especially, is an important soil property as its presence is a precondition for the formation of soils and it plays an essential role in soil fertility and soil water characteristics [2,3]. Soils are one of the largest carbon sinks and hence variations in SOC content have effects on the climate [1]. In the framework of the Sustainable Development Goals (SDGs) listed

by the United Nations, SOC is regarded as the most relevant soil property with regard to climate regulations [2] and as such is listed as one of the GCOS (Global Climate Observing System) ECV (Essential Climate Variables) for which monitoring of the SOC status should be improved [4]. In order to prevent further soil degradation, there is a recognized need for up-to-date, spatially accurate and inexpensive soil maps to monitor soils and their processes (e.g., [5–7])

As traditional soil mapping involves intense soil sampling and wet chemistry analyses, it is time and cost intensive and the spatial accuracy is dependent on the density of collected soil samples. An alternative to wet chemistry is soil spectroscopy which uses spectral information to derive key soil properties. It is a well-established method under laboratory conditions using point spectrometers [8] which measure the radiation reflected by an object at a single point. Therefore, soil properties need to be spectrally active to produce a specific spectral signature or alternatively they need to be correlated to a soil property that fulfils this requirement (e.g., [5]). SOC shows strong interactions with the electromagnetic radiation in the spectral region visible for human eyes and hence, soils rich in SOC appear darker as a general decrease in reflectance is observed with an increasing SOC. Furthermore, biochemical components of soil organic matter like chlorophyll, cellulose, starch, oil, pectin, lignin and humic acids [9] have impacts on the reflectance in from the visible (400–700 nm) to the near-/shortwave infrared (700–2500 nm) regions [10] of the electromagnetic spectrum. As a first effect, a general flattening of the reflectance spectral curve in the visible (VIS) is typical of SOC in agricultural soils, related to the decomposition of chlorophyll pigments which originally causes a wide spectral absorption feature around 664 nm [10,11]. Furthermore, there are strong and narrow absorptions caused by SOC in the short-wave infrared (SWIR) between 2100–2300 nm that can be observed in lignite-rich soils [10–12]. To handle the broad, weak and superimposing absorption features of soils, most studies use partial least squares regression (PLSR) or similar multivariate modelling techniques as they cope well with a high collinearity and a large number of predictor variables [13,14].

Additionally, imaging spectroscopy, primarily from airborne platforms, is already widely used in many local studies and leads to good prediction accuracies, with overviews provided in: [5,15]. Many studies have been conducted to estimate the SOC content mainly at a local farm or field scale using hyperspectral imagery and soil samples collected at the study site, which were later analyzed for their SOC content at a laboratory [13]. Using airborne hyperspectral data, good prediction accuracy can be achieved: [16] used the airborne HyMAP sensor to estimate the SOC content in northern Germany using PLSR at field scale and received an $R^2 = 0.9$. Lower prediction accuracies were reported using the first generation of spaceborne hyperspectral imagers: [17] explain their SOC prediction accuracy using Hyperion data and PLSR models of $R^2 = 0.51$ and RPD = 1.43 in Australia by a low signal-to-noise ratio (SNR) and a spatial resolution of 30 m.

When moving from the laboratory to an airborne and spaceborne scale, new challenges arise for soil property quantification, e.g., weather conditions and atmospheric attenuation, soil conditions (mainly moisture content, roughness and crusts), internal pixel variations of the soil property or even fractions of other land cover types within the pixel [5,18]. There is ongoing research to meet these challenges and improve model accuracy (overview in [5]). Scale is also an important component influencing model accuracy. On larger scales, soils are more inhomogeneous leading to increasing variances of soil properties but also to non-linear relationships between soil property and spectrum which complicates model calibration [14,19,20]. Furthermore, the availability of local or regional soil databases for model calibration and validation is an issue. A reduced need for soil samples for model calibration would be desirable, as sampling campaigns and wet chemistry analyses are time and cost intensive, especially with regard to applications at a global/regional scale.

With the recent and upcoming availability of hyperspectral satellite data with high signal to noise ratio (e.g., PRISMA (PRecursore IperSpettrale della Missione Applicativa) [21], EnMAP (Environmental Mapping and Analysis Program) [22]) research also focusses on the potential for the quantification of soil properties from these new platforms and for larger areas. Chabrillat et al. [23] demonstrated that no simple best-practice rule exists concerning the method to be used for certain soil properties and sensors.

For each combination the best method needs to be defined by further studies. Steinberg et al. [24] obtained good PLSR prediction accuracies for selected soil properties using simulated EnMAP images for sites in Spain and Luxembourg. Generally, the prediction accuracy decreases from airborne to simulated satellite observations with decreasing spatial resolution [23,24]. Currently, efforts are being made to create large scale soil spectral libraries that can be used for model calibration to develop more robust models that are applicable for larger areas in order to create global maps with less need for local soil sampling and analyses [5]. Wetterlind et al. [25] found that using the national soil spectral library of Sweden spiked with local samples from their Swedish study site, can lead to comparable results as if PLSR models were calibrated at a farm-scale. Castaldi et al. [26] used the samples of pan-European topsoil database LUCAS (land use/cover area frame survey) located within a selected scene of the multispectral Sentinel-2 mission to fit exponential regression models based on band combinations. When applied to the Sentinel-2 bands, they received good validation results by using the red and red edge bands. Castaldi et al. [27] obtained good results by also using LUCAS in a bottom-up approach using PLSR to predict SOC for hyperspectral airborne data acquired in Belgium and Luxembourg without a spectral transfer function between laboratory and airborne data. Tziolas et al. [28] estimated the SOC content for a study site in Israel applying a spiked bottom-up approach with a local Gaussian regression on airborne imagery. They improved the modelling accuracy by spiking the GEOCRADLE soil spectral library with a varying number of local samples. Additionally, they spectrally resampled the airborne imagery to the spectral resolutions of EnMAP and Sentinel-2 and showed that this has minor effects on the prediction accuracy.

In this study we quantified and mapped the SOC content, to our knowledge for the first time, based on fully simulated hyperspectral spaceborne images by making use of a large-scale soil spectral library. This approach was demonstrated on laboratory data but has never been tested for hyperspectral satellite based remote sensing data. Previously, Castaldi et al. [26] used the LUCAS database to map SOC using multispectral Sentinel-2 data. Castaldi et al. [27] applied a methodology similar to our approach but using airborne hyperspectral imagery and an approach with pre-defined clusters based on soil properties from the LUCAS database developed in [29]. Recently, [28] also used a similar methodology as in [27], but used a soil spectral library spiked with local samples and applied it to airborne data, which was spectrally but not spatially resampled to EnMAP's resolution.

Here, we aim to build models based on the LUCAS soil spectral library and on few spectral measurements of local samples in order to develop an approach which is applicable to remote sensing imagery. We are apply a two-step approach by: (i) using a local PLSR approach [30] tested in [31] and the LUCAS database to quantify the SOC content for a range of field calibration soil samples; and subsequently (ii) we use the estimated SOC content of these field calibration samples to calibrate models for both airborne and simulated EnMAP imagery. The benefit of this approach is the replacement of expensive wet chemistry analyses of soil samples from the local study site, needed to calibrate remote sensing models, by soil spectroscopy. We assess and compare the prediction accuracy for both airborne and simulated satellite imagery for the same study site. Furthermore, we assess the uncertainty added by the proposed two-step approach by comparing it with the traditional approach which uses the SOC content measured by wet chemistry.

2. Materials and Methods

2.1. Data

2.1.1. Study Site

The study site is located in north-eastern Germany near the town of Demmin in the federal state of Mecklenburg-Western Pomerania. This study site belongs to the observatory Northeastern German Lowland as part of the TERrestrial ENvironmental Observatories (TERENO) which is a long term terrestrial environmental monitoring site initiated by the Helmholtz Association [32] and is also part of the Durable Environmental Multidisciplinary Monitoring Information Network—the German

JECAM site DEMMIN operated by the German Aerospace Center (Deutsches Zentrum fuer Luft- und Raumfahrt, DLR) and the German Research Centre for Geosciences (GFZ) Potsdam [33]. The area is located in the young morainic soil landscape of northern Germany [34] and the morphology of this area was formed by periodic glacial processes during the Pleistocene (Weichselian Glaciation). The dominant soils in this area are Haplic Luvisols, Eutric Podzoluvisols and Stagnic Luvisols from boulder clay [35]. Soil collection was performed in the sandy loam area characteristic for a high variability in SOC at small scale and large crop fields adequate for remote sensing studies (Figure 1).
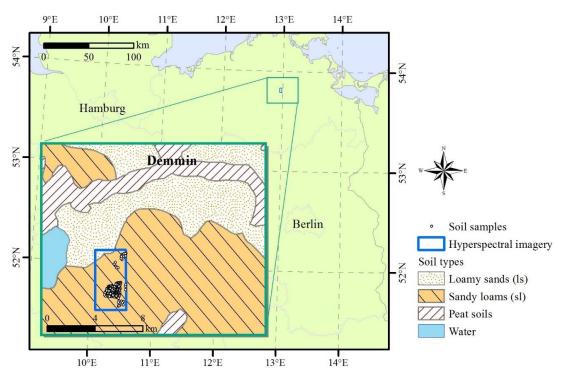

Figure 1. Study site Demmin in the north east of Germany. Data source soil types: BOART1000OB V2.0, (C) BGR, Hannover, 2007.

2.1.2. LUCAS Soil Database

In this study we use the European Land Use/Cover Area Frame Survey (LUCAS) topsoil database which is managed by EUROSTAT together with the European Commission's Directorates-General for Environment and the Joint Research Centre at Ispra, Italy [36,37]. LUCAS is currently the most consistent soil database at a continental scale and was created to support policy making. We used data from the survey that took place in 2009 in 25 member states of the European Union and which includes 19,967 top-soil samples (0–20 cm) collected on different land use types [37]. The database includes 12 different soil properties and spectral measurements. The SOC content has been measured by dry combustion using a VarioMax CN Analyzer (Elementar Analysensysteme GmbH, Germany). Before taking spectral measurements, samples were dried at 40 °C, crushed and sieved (< 2 mm). The absorbance spectra were measured using a FOSS XDS Rapid Content Analyzer (FOSS NIR Systems Inc., Laurel, MD, USA). The XDS data were acquired within a range of 400.0–2499.5 nm with a spectral resolution of 0.5 nm, resulting in 4200 wavelengths. We subset the LUCAS database to the land use and land cover class "agriculture" consisting of ~8000 data points, as in this paper the focus is placed on agricultural soils and associated SOC and spectral properties. Furthermore, agricultural areas that are temporarily free of vegetation and exposed bare fields after harvest can be used as soil surfaces for the subsequent mapping of soil properties from air- and spaceborne platforms. The SOC content in the LUCAS agricultural subset ranges between 0–194 g kg^{-1} and is highly skewed including many low and few high values (median 14.4 g kg^{-1}).

2.1.3. Field Soil Samples

We collected 37 top soil samples (0–10 cm depth) within five exposed bare fields in the study site in October 2013, 2016 and 2017 which were used for model calibration and validation. Each sample consists of a mixture of five sub-samples taken within a radius of 5 m around a central location. These samples were dried, crushed and sieved (<2 mm). The SOC content was extracted for all samples by dry combustion using a VarioMax CN Analyzer (Elementar Analysensysteme GmbH, Germany). It ranges between 7.6–134.1 g kg^{-1} with a median of 9.9 g kg^{-1} and a mean of 20.3 g kg^{-1} (Figure 2a, Table 1). As for the LUCAS database, the spectral absorbance was measured for the calibration samples with a FOSS XDS Rapid Content Analyzer in the REQUASUD network laboratories of the Province of Liege [38]. The XDS data were acquired within a range of 400.0–2499.5 nm with a spectral resolution of 0.5 nm, resulting in 4200 wavelengths.
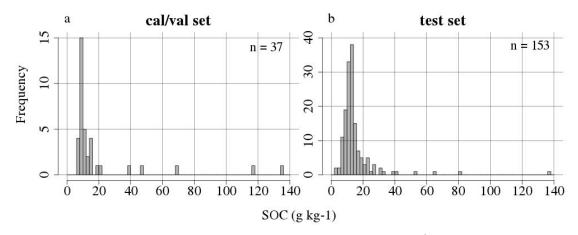

Figure 2. Frequency of the measured soil organic carbon (SOC) content (g kg^{-1}) for the top soil sampling in the Demmin study site area used for model calibration and validation (**a**) and for model testing (**b**).

Table 1. Statistical summary of the three datasets.

Dataset	Minimum	Mean	Median	Maximum	Standard Deviation	Number of Samples
calibration	8.57	29.24	15.82	117.77	31.61	13
validation	7.58	15.50	8.97	134.05	26.02	24
test	3.37	15.37	12.40	136.60	13.78	153

As a test set, another 153 independent soil samples were used which were collected in 2013 and 2016 on exposed bare soils and as a mixture of five sub-samples taken within a radius of 5 m around a central location. They were dried, crushed and sieved (<2 mm) and their SOC content was measured in the laboratory of the Technical University Berlin by loss of ignition (LOI). To transform the LOI into the SOC content, the following transformations were performed: SOC = (LOI—0.1*clay content)/1.72 [39]. The outcome was multiplied by 10 to transform from % SOC to g kg^{-1}. The resulting SOC contents ranged between 3.4–136.6 g kg^{-1} with a median of 12.4 g kg^{-1} and a mean of 15.4 g kg^{-1} (Figure 2b, Table 1). For this set of samples, no FOSS spectra were scanned. Due to the SOC content being measured by a different method compared to the calibration and validation samples, we intended to include these samples as test set samples where they are used for independent model testing.

2.1.4. Remote Sensing Imagery

The airborne hyperspectral imagery was acquired using the HySpex imaging system owned by GFZ Potsdam which was mounted on the Cessna 207T aircraft of the Free University Berlin within the frame of an EnMAP flight campaign [40,41] The NEO HySpex system consists of two separated

push-broom hyperspectral cameras (visible to near infrared (VNIR)-1600: 400–1000 nm and short-wave infrared (SWIR) 320m-e: 1000–2500 nm) with a spectral resolution of 3.7 nm (VNIR-1600) and 6.0 nm (SWIR-320m-e), resulting in 416 spectral bands [42].

On 1 October 2015 on a clear day, three flight lines were recorded using the HySpex VNIR and SWIR systems. The airborne campaign was conducted at a mean altitude of 2500 m which resulted in a mean resampled ground sampling distance of 4 m. The georectification, co-registration and adaptation of the SWIR to the VNIR sensor was performed with the GFZ in-house processing chain HyPrepAir based on the procedures described in [43]. Subsequently, the data cube was atmospherically corrected with the ATCOR-4 software [44], the single flight lines were mosaicked together and a spatial subset of ~13 km^2 was selected. Based on the HySpex image, the EnMAP end-to-end image simulation software EeteS [45] was used to simulate hyperspectral imagery from the upcoming EnMAP satellite (launch planned for end of 2021). EeteS allows the simulation of EnMAP L0 data using airborne hyperspectral imagery and a complete L1 and L2 pre-processing chain to retrieve spectral reflectance. EnMAP has a spatial resolution of 30 m and consists of two spectrometers with a spectral coverage of 420–2450 nm and a spectral resolution between 5 and 12 nm, resulting in 242 spectral bands.

2.2. Methods

2.2.1. Overview and Concept

In this paper we propose to use the local PLSR approach [30] described previously by [19] and further developed by [31]. It is based on the principle of the spectral neighborhood similarity to define individual "local" spectral clusters for each of the validation samples as basis for PLSR modelling of SOC content. This approach is useful as it would allow deviation from the need for wet chemistry analyses by spiking the LUCAS soil spectral library, but with spectral analyses of local samples only. One challenge when moving from the laboratory to remote sensing data, is that spectra are measured under very different conditions resulting in spectral differences that hamper a combined use of the spectra in prediction models. Due to the difficulties in matching laboratory spectra with remote sensing spectra, we propose and test in this paper a two-step approach close to the concept proposed by [27]. In the first step (i) the SOC content of soil samples collected from the area of interest was estimated using laboratory models based on the LUCAS database using the local PLSR approach. In the second step (ii) a remote sensing model was created using the image spectra at the location of the soil samples and the corresponding predicted SOC values from the laboratory model (not from wet chemistry). This approach was compared to a traditional approach using image spectra and SOC contents measured by wet chemistry. Subsequently, the two-step approach was tested on the HySpex 4 m imagery and on the simulated EnMAP 30 m data. Most calculations were completed in R 3.5.0 [46]. An overview of the processing scheme is given in Figure 3.

2.2.2. Data Pre-Processing

The following pre-processing steps were performed for the LUCAS spectral library as well as for the laboratory spectra of the soil samples belonging to the calibration subset. We removed the wavelength < 500 nm due to instrumental artefacts, as observed by [20], and reduced the spectral resolution from 0.5 nm to 2 nm by keeping one in four wavelengths in order to increase the processing time and as they contain highly redundant information. Based on absorption spectra, we performed a Savitzky–Golay smoothing with a 2nd order polynomial and a window size of 11, corresponding to 22 nm, and calculated the first derivative using the R function sgolayfilt (package: signal) [47]. Furthermore, due to a strongly skewed SOC variable (skewness = 4.6), the SOC content was transformed towards a normal distribution (skewness = 0.12) using the natural logarithm.
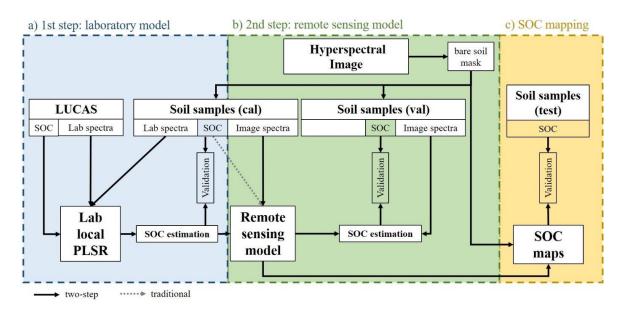
Figure 3. Overview of the processing scheme with (**a**) laboratory model, (**b**) remote sensing model and (**c**) SOC mapping. LUCAS: land use/cover area frame survey; PLSR: partial least squares regression.

In the remote sensing images, the bands ≤ 500 nm and ≥ 2300 nm were removed due to noise, and the following bands were excluded (i) 1093.62–1231.67 nm, 1351.72–1501.77 nm and 1855.9–2095.98 nm for HySpex, respectively (ii) 1085.4–226.7 nm, 1358.5–1499.4 nm and 1803.5–2097.6 nm for EnMAP. A Savitzky–Golay smoothing with a 2nd order polynomial and a window size of (i) 21 bands for HySpex, corresponding to 76 nm in the VNIR and 126 nm in the SWIR and (ii) 13 bands for EnMAP, corresponding to 65 nm–156 nm, was performed on the mosaic. Erroneous pixels in some wavelengths caused by the sensor were replaced by averaging the neighbouring wavelengths. Then, a soil mask was created following the methods from the HYSOMA/EnSoMAP algorithms [48,49] using a set of spectral indices and thresholds in order to ensure using only bare soil spectra for model calibration, validation and mapping. Bare soil pixels were selected with the normalized soil moisture index (NSMI) [50] < 0.27, the normalized difference vegetation index (NDVI) < 0.3 and the normalized cellulose absorption index (nCAI) < 0.03 as defined and tested in [48,49].

$$NSMI = \frac{r1800 - r2119}{r1800 + r2119}, \tag{1}$$

$$NDVI = \frac{NIR - red}{NIR + red} = \frac{r660 - r800}{r660 + r800}, \tag{2}$$

$$nCAI = \frac{0.5 * (r2000 + r2200) - r2100}{0.5 * (r2000 + r2200) + r2100}, \tag{3}$$

A bare soil mask was derived based on the indices and thresholds described above, setting all non-bare soil pixels to NA. For the HySpex image the bare soil mask was complemented by a land use classification applying the classification workflow with the random forest classifier in the EnMAP-Box 3.3 [51] as QGIS 3.4.8 plugin to exclude artificial surfaces present in the villages. This was not necessary in the EnMAP images as due to the coarser spatial resolution no pure spectra of artificial materials could be identified and additionally, the bare soil masks based on indices already excluded most pixel related to human settlements. After creating the bare soil mask, the following procedure was performed for all bare soil pixels: the original reflectance spectra were transformed to absorbance spectra using A = log(1/reflectance), smoothed with the same parameters used for smoothing before calculating the bare soil mask and the first derivative was calculated.

For the two-step approach the soil samples were divided into 1/3 calibration and 2/3 validation samples. Therefore, the fields in the study site in which the soil samples have been collected were

divided into calibration and validation fields to allow for an independent validation (Figure 4). The split was performed in such a way that both calibration and validation datasets contain a similarly high variability in SOC. The calibration field in particular was chosen based on expert knowledge of the area. It is large in size and encompasses a high variability in SOC content representative for the area. In order to improve the applicability of the approach, we wanted to keep the number of calibration samples as low as possible. Additionally, using only one representative field for model calibration reduces the logistic efforts during sampling. The collection of validation samples is optional and necessary only if the model accuracy is to be accessed.
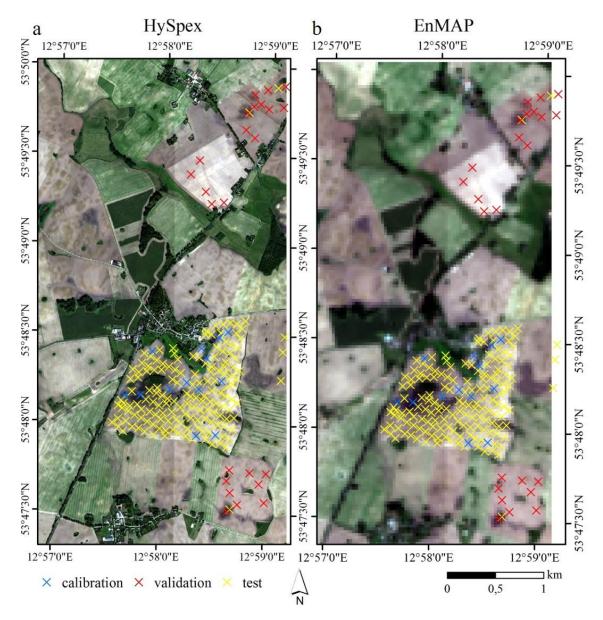

Figure 4. HySpex (**a**) and simulated EnMAP (**b**) images of the study sites as RGB composites. Locations of soil samples for model calibration (blue), validation (red) and test (yellow) are shown as crosses.

2.2.3. First Step: Laboratory SOC Modelling

The smoothed 1st derivative of the agricultural subset of the LUCAS database was used for model calibration to predict the SOC content for the calibration subset of the soil samples collected in the study site. A local PLSR [30] was used for the prediction of SOC as described in [19,31]. It calibrates

a separate PLSR model for each validation sample based on a set of most similar spectra out of the calibration samples using the pls package in R [52]. The most similar samples were chosen using the Mahalanobis distance that was calculated using the function fDiss (package: resemble) [53] and based on the first seven principal components which explained more than 99.9% of the spectral variance. A threshold of 0.19 in the Mahalanobis distance was used to select the most similar LUCAS calibration samples and to allow for varying numbers of calibration samples but with a minimum of 200 samples which proved best in [31] also using the LUCAS database.

2.2.4. Second Step: Remote Sensing SOC Modelling

Airborne spectra were extracted at the location of the soil samples by applying a buffer with a certain radius around the soil sampling locations. The spectra of all pixels which have their center within this buffer radius were averaged using the function extract (package: raster) [54]. For HySpex we used the average raster values within a 10 m buffer around the soil samples' locations as this is the typical maximum spatial inaccuracy of the GPS device specified by the manufacturer. For EnMAP with its 30 m pixel size the 10 m buffer is not meaningful and was therefore extended to 30 m to include the nearest 3–4 pixels. We selected this method to ensure that all soil sampling locations especially those located at the borders of a pixel were well represented. In case of the EnMAP image, two samples were removed, because they were too close to the image border (Figure 4). Both samples belonged to the validation samples which reduced the number from 24 to 22 samples. The measured and predicted SOC values of the soil samples were highly skewed and were therefore transformed towards a normal distribution using the inverse value 1/SOC.

A PLSR model was calibrated for HySpex respectively simulated EnMAP based on the first derivative image spectra and the SOC values predicted in the laboratory model in the first step. To select the optimal number of latent variables (LV) in the PLSR, we used the smallest number of LV within one standard deviation of the minimal error. The prediction results of the remote sensing models were compared with the measured SOC contents of the independent validation samples and assessed using the root mean squared error (RMSE), coefficient of determination (R^2), the ratio of performance to deviation (RPD) and the bias:

$$RMSE = \left(\sum_{i=1}^{n} (yp_i - yo_i)^2 / n \right)^{1/2}, \tag{4}$$

$$R^2 = 1 - \sum_{i=1}^{n} (yp_i - yo_i)^2 / \sum_{i=1}^{n} (yo_i - \overline{yo})^2, \tag{5}$$

$$RPD = sd(yo) / RMSE, \tag{6}$$

$$bias = \sum_{i=1}^{n} (yp_i - yo_i) / n, \tag{7}$$

with yo_i as the observed SOC value of sample i and yp_i as the predicted SOC value of sample i. $\overline{yo}$ is the mean of the observed SOC values, n is the number of samples and sd is the standard deviation.

Finally, the remote sensing SOC models were directly applied to the corresponding bare soil pixels of the HySpex resp. EnMAP images and SOC maps were created. The accuracy of the SOC map was assessed with a test set of 153 soil samples collected in 2013 and 2016. The SOC content was extracted from the maps using again a 10 m buffer for the HySpex map and a 30 m buffer for the EnMAP map and compared to the measured SOC content of the test set samples.

3. Results

3.1. Spectral Data

Reflectance spectra of the airborne calibration and validation sets are given in Figure 5 which clearly shows the differences between laboratory and image spectra. Due to the relatively dark soil surfaces and low solar illuminations conditions for high latitude northern Germany in autumn, the acquisition conditions were non-optimal which resulted in a relatively low signal-to-noise ratio (SNR). This can be seen in the absence of very shallow clay features in some spectra in the SWIR which could also be caused by spectral mixing at the observation scale compared to laboratory sampling. Nevertheless, data are of adequate quality considering the acquisition conditions.

Figure 5. Reflectance spectra of laboratory spectrometer FOSS (left; **a**), airborne HySpex (middle; **b**,**d**) and simulated satellite EnMAP (right; **c**,**e**) of the calibration samples (top; **a**–**c**) and validation samples (bottom; **d**,**e**). FOSS laboratory spectra were used for laboratory modelling while HySpex/EnMAP spectra were used for the remote sensing model. n is the number of soil samples.

3.2. Laboratory SOC Modelling

The best number of LV in the local PLSR of the laboratory model was chosen based on a 10-fold cross validation that was used to calculate the RMSE for a range of LV. This implies a random selection of the validation subsets in the cross-validation part and hence, a variance in the selected number of LV and SOC contents. Therefore, we used an ensemble model by calibrating 100 models for each sample and averaging the predicted SOC contents. We received good validation results (Table 2, top row) with an R^2 of 0.86, RPD of 2.77 and a RMSE of 11.41 g kg^{-1} and a very high Pearson's correlation coefficient between observed and predicted SOC contents (Figure 6 top, left) of 0.96.

Table 2. Model results of the first step (laboratory, row 1), airborne HySpex (10 m buffer) with traditional and two-step approach (row 2–3) and simulated satellite EnMAP (30 m buffer) with traditional and two-step approach (row 4–5). RMSE: root mean squared error.

Method	Approach	Sensor	RMSE (g kg^{-1})	R^2	RPD	Bias	LV	n cal/val	SOC Range val (g kg^{-1})
local PLSR	lab	FOSS	11.41	0.86	2.77	−4.69	14	8294 [1]/13	9–118
PLSR	traditional	HySpex	11.88	0.78	2.19	6.19	1	13/24	8–134
PLSR	two-step	HySpex	14.56	0.67	1.79	2.15	1	13/24	8–134
PLSR	traditional	EnMAP	12.63	0.77	2.15	2.01	1	13/22	8–134
PLSR	two-step	EnMAP	19.22	0.48	1.41	−0.96	1	13/22	8–134

[1] subset varying in size was used during the local PLSR with a minimum of 200 samples.
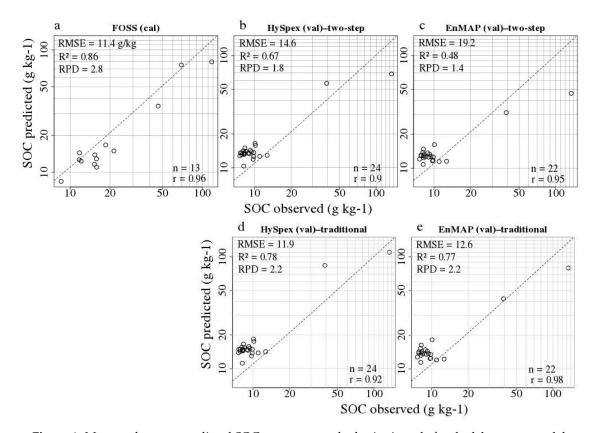

Figure 6. Measured versus predicted SOC contents on a logharitmic scale for the laboratory model (airborne calibration subset), airborne HySpex and simulated EnMAP imagery (airborne validation subsets). Left (**a**): Predicted SOC contents of the laboratory model result from the application of the local PLSR; middle (**b**,**d**): HySpex; and right (**c**,**e**): EnMAP. Top panel (**a**–**c**): SOC predictions result from the two-step approach using PLSR models; and bottom panel (**d**,**e**): SOC predictions result from the traditional approach using PLSR models. For HySpex using the mean spectra within a 10 m buffer and for EnMAP 30 m buffer. n is the number of soil samples and r is Pearson's correlation coefficient.

3.3. Remote Sensing SOC Modelling

For HySpex we compared the performance of the traditional and the two-step approach to test the validity of the new approach. The traditional approach uses measured SOC values and image spectra, whereas the two-step approach uses predicted SOC values (Figure 3) and image spectra. For HySpex both approaches gave good results with R^2 = 0.78, RPD = 2.19 for the traditional and R^2 = 0.67, RPD = 1.79 for the two-step approach (Table 2, rows 2–3). For EnMAP the prediction accuracy is very

similar to HySpex for the traditional approach but lower for the two-step approach with $R^2 = 0.48$, RPD = 1.41 (Table 2, rows 4–5). A random forest approach was tested simultaneously to the PLSR approach for the remote sensing models but results were poor and are subsequently not shown in this study.

The Pearson's correlation coefficients between observed and predicted SOC contents are high (r > 0.9) for both HySpex and EnMAP and both PLSR approaches. The prediction of the SOC content of soil samples with high amounts of SOC is prone to larger errors while SOC contents of the soil samples in the lower SOC range tend to be slightly over-predicted (Figure 6).

Important wavelengths for remote sensing SOC modelling are depicted in Figure 7 which shows the coefficients of the first latent variable of the PLSR models. As a threshold, one standard deviation is plotted according to [55]. The patterns were very similar for both HySpex and EnMAP and most important wavelengths were located roughly at 500–600 nm, 700–750 nm, 850–950 nm, around 1050 nm and for HySpex additionally at 2300 nm.
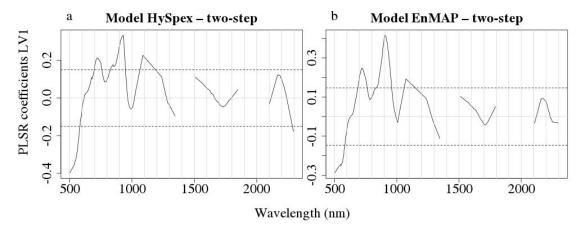

Figure 7. PLSR coefficients for the airborne HySpex model (**a**) and simulated satellite EnMAP model (**b**) using the two-step approach. Dotted lines represent one standard deviation of the coefficients.

Resulting SOC maps (Figure 8) were generated using the remote sensing PLSR models based on the two-step approach. These models were applied to the bare soils within the images. The few pixels with negative SOC predictions were set to zero and values above 200 g kg^{-1} were set to 200 g kg^{-1}. Figure 8 shows that SOC maps for both HySpex and simulated EnMAP show similar spatial patterns. In the HySpex map with its finer spatial resolution more sharp contours are visible and some sandy paths remained within the map after applying the bare soil mask. Additionally, areas with very high SOC content were correctly depicted. In the EnMAP map the areas with very high SOC contents are smaller while the transition areas with intermediate SOC contents are larger. In both maps some of the field's borders were predicted to have a higher SOC content which is more pronounced for EnMAP with its larger pixel size.

The SOC maps were validated with a test set using independent soil samples. These samples were not used for the model training and validation as their SOC content was measured in the laboratory by a different method and this would add uncertainty concerning the measured SOC content. Figure 9 shows a good correlation between SOC values measured in the laboratory and SOC values predicted for both remote sensing sensors. The correlation for HySpex is higher with r = 0.88 and slightly lower for EnMAP with r = 0.72. The calculated model performances based on the test set samples are high for the HySpex data ($R^2 = 0.76$, RPD = 2.1) and medium–high with EnMAP data ($R^2 = 0.46$, RPD = 1.4). For both sensors especially predictions at locations with SOC contents > 40 g kg^{-1} were with a higher uncertainty. Using HySpex they were both under- and overestimated and for EnMAP they were all underestimated. At the lower edge of the SOC range, samples collected at locations with a SOC content < 10 g kg^{-1} were overestimated by both models.
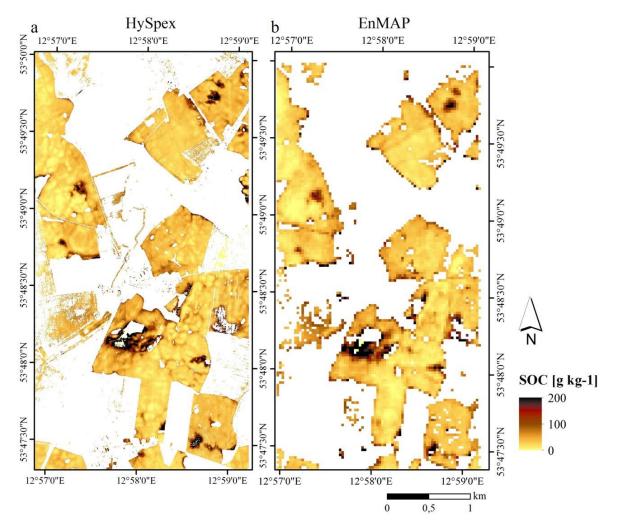
Figure 8. SOC maps for airborne HySpex (**a**) and simulated satellite EnMAP imagery (**b**) using the two-step PLSR approach applied to all bare soils.
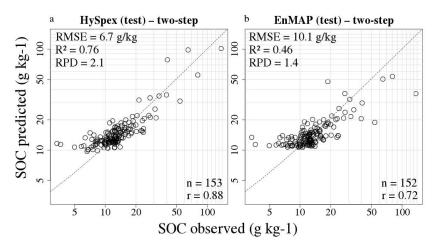

Figure 9. Observed vs. predicted SOC content on a logarithmic scale for the test set of soil samples used for SOC map assessment for the HySpex (**a**) and EnMAP sensors (**b**). n is the number of soil samples while r is Pearson's correlation coefficient.

Additionally, we tested the influence of different buffer sizes on the prediction accuracies for HySpex PLSR model validation (Figure 10). Spectra were extracted and averaged within a varying

radius (10, 20 and 30 m) around the central location specified by the soil samples. These mean spectra were used for PLSR model calibration and validation and evaluated for both the two-step and traditional approaches. For the two-step approach prediction accuracy decreased with increasing tested buffer size whereas the accuracy for the traditional approach was lowest for the 20 m buffer. The traditional approach gave higher prediction accuracies for all buffer sizes and the difference between the two approaches increased with buffer size. For EnMAP and HySpex the RMSE is very similar for the 30 m buffer using the traditional approach and differs more for the 30 m buffer using the two-step approach.
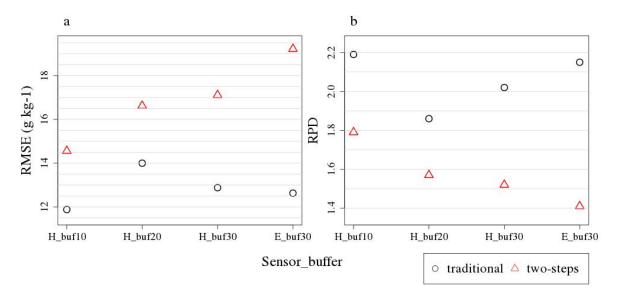

Figure 10. RMSE (**a**) and RPD (**b**) for HySpex (H_) and EnMAP (E_), for the traditional and the two-step approaches and different extraction buffer: buf10 = buffer of 10 m, buf20 = buffer of 20 m, buf30 = buffer of 30 m.

4. Discussion

The accuracy of the laboratory model in the first step using the local PLSR approach with the LUCAS database leads to an $R^2 = 0.86$ and RPD = 2.77 which is better or comparable to the results of studies with similar approaches [19,27,31], all using the LUCAS dataset and [28] using the GEOCRADLE dataset without spiking.

In the second step, the remote sensing model calibration, we showed that PLSR was capable of achieving good prediction accuracies with a low number of samples. This model calibrated on one field was transferable to other fields nearby that have a similar SOC variability and distribution. Using few samples is beneficial in order to reduce time and costs dedicated to create SOC maps and can improve the applicability of the approach, if it does not significantly reduce the prediction accuracy. Here we did not test the effect of varying sizes of calibration subsets, but showed that a low number might be sufficient depending on the local characteristics of the study site. This number is likely to vary for other study sites. Although high performance modelling was observed as given by RPD and R^2, relatively high RMSE values of 11.88 g kg^{-1} for HySpex and 12.63 g kg^{-1} for EnMAP were determined, due to a very large SOC range in the area with values ranging from 8–134 g kg^{-1} in very small spatial scale. Comparable studies which used a study site with a larger SOC range also report higher RMSE, e.g., [13] who acquired an RMSE of 73 g kg^{-1} using an airborne hyperspectral scanner (AHS) imagery for a study site with a SOC range of 27–380 g kg^{-1}. The RPD value is more suitable for comparison between different study sites as it sets the standard deviation of the measured SOC values relative to the RMSE. Our RPD results of 1.79 in the second step using HySpex compare well to the results of [27] using a comparable approach who received an RPD of 1.4 respectively 1.7 for their two study sites.

The R^2 value of 0.67 is comparable to the result of [28] who received an R^2 = 0.65 for their bottom-up approach without spiking the soil spectral library with local samples.

The SOC prediction accuracies of the HySpex traditional approach (R^2 = 0.78, RPD = 2.19) compare well with other studies using the traditional approach [13]. The accuracies for the traditional approach using simulated EnMAP imagery (R^2 = 0.77, RPD = 2.15) are higher than those of [24] who also used simulated EnMAP image data and achieved an accuracy of R^2 = 0.67, RPD = 1.7. [56] and [31] predicted SOC with a similar accuracy, R^2 = 0.67, RPD = 1.80 [56] and R^2 = 0.66, RPD = 1.71 [31], using simulated EnMAP spectra based on laboratory spectra. Studies using Hyperion satellite data for SOC predictions achieved lower prediction accuracies, e.g., [17] with R^2 = 0.51 and RPD = 1.43 or [57] with R^2 = 0.63, RPD = 1.65.

For the Demmin study site and the HySpex imagery the two-step and the traditional approach lead to only small differences in accuracies concerning SOC predictions. The accuracy decreases slightly for the two-step approach which is caused by the additional uncertainty included by using predicted SOC contents for model calibration, and not real SOC contents as provided usually by wet chemistry. Hence, the two-step approach which makes use of a large scale soil spectral library provides a valid alternative to the traditional approach. Castaldi et al. [27] made a similar conclusion for their study sites in Belgium and Luxembourg with a similar approach. For the simulated EnMAP image the prediction accuracy differs more between the two approaches which is probably mainly a consequence of the coarser spatial resolution. This is in line with [28] who used an approach similar to our two-step approach and did not observe a significant decrease in prediction accuracy by spectrally resampling an airborne imagery to EnMAP's spectral resolution. It may be (partly) explained by the fact that in [28] the high spatial resolution was not resampled to EnMAP's coarser 30 m resolution different from the present study that is based on more realistic simulations of future EnMAP data. As Figure 10 shows the difference between the two approaches increases for larger buffer sizes tested for HySpex which is consistent with [17]. For EnMAP the uncertainties resulting from using predicted SOC contents and a coarser spatial and spectral resolution seem to accumulate. Especially the one sample with the highest measured SOC content is predicted less accurately when compared to the HySpex model which is largely influencing the model accuracy (Figure 6).

The spectra extracted from the HySpex and EnMAP images reveal a high similarity (Figure 5). For EnMAP one sample with a high SOC content shows a reflectance spectrum with a higher albedo compared to HySpex which is probably caused by the coarser spatial resolution and hence, averaging of a larger area. Additionally, the important wavelengths shown using the PLSR coefficient (Figure 7) are very similar for HySpex and EnMAP. Larger differences appear comparing laboratory and airborne/simulated spaceborne spectra (Figure 5). There is a reduced overall albedo and additional noise in the airborne/simulated spaceborne spectra. This is induced by effects due to the impact of reduced spatial, spectral and radiometric performances of remote sensing sensors compared to laboratory sensors, due to, e.g., atmospheric absorptions and disturbing effects of changing surfaces for remote sensing observations. It is a consequence of the different conditions during spectral acquisition: for the laboratory measurements soils are dried, crushed and sieved and measured under stable illumination conditions using an artificial light source. Airborne measurements are influenced by many factors such as changing atmospheric conditions, soil roughness, soil crusts, dry residues, moisture and mixed spectral information within one pixel. Therefore, a direct application of the local PLSR approach based on a laboratory spectral library to airborne image spectra or even field spectra is unlikely. The reduced performance of airborne models compared to laboratory models has been demonstrated in many papers, e.g., [5].

The correlation between observed and predicted SOC contents are very high for all modelling approaches (Figure 6), which could indicate that the models are accurate in describing the spatial patterns of SOC. On the other hand, the model accuracy measures are lower compared to the correlation results which shows that the difficulty lies in the accurate quantification of the SOC contents. The laboratory model has a negative bias which influences the prediction accuracy of the second step,

the remote sensing models. The remote sensing models have more difficulties in predicting higher SOC contents and are less valid for samples with high SOC values (Figure 6). One reason is the skewed distribution of SOC contents with a low number of samples with high SOC values in the calibration dataset. Other possible explanations are the different link between SOC and spectra for low and high SOC values and saturation effects [19]. Furthermore, we used the average spectra within a buffer around the sampling locations. Thereby, smaller areas with high SOC contents are potentially mixed with areas with lower SOC contents leading to smoother and less extreme spectra. At the lower end of the SOC range, SOC values tend to be overestimated (Figure 6). This might also be explained by signals being averaged over several pixels. As the frequency of pixels with low SOC values is low, the chance of including pixels with slightly higher SOC contents which have a higher frequency is likely.

When increasing the buffer size around the sampling locations the accuracy of the two-step approach decreases for larger buffers (Figure 10). The traditional approach follows this pattern except for the 30 m buffer for which the accuracy increases. This might be explained by site specific characteristics and concerning HySpex more pixel being averaged leading to smoother spectra. However, this finding should be checked for other study sites. Additionally, the accuracy of the traditional approach using a 30 m buffer for EnMAP is slightly higher than for HySpex using the same buffer radius. Even using the same buffer size for both sensors does not ensure covering exactly the same area. The reason for this is that pixels are only included if their centre is located within the buffer radius. For EnMAP, the number of pixels included is dependent on the location of the soil sample within the certain pixel and is typically 3–4 pixels, and in this case adding or excluding one pixel already makes a big difference. For HySpex with a higher spatial resolution, the location of the soil sample within a certain pixel is less crucial as a larger number of pixels are included within the buffer radius and the area covered can be approximated by a circle.

The SOC maps (Figure 8) are coherent with the knowledge of the area and previous remote sensing SOC mapping in these fields [58,59]. Additionally, HySpex and EnMAP SOC maps show high coherency, with an aggregation for the EnMAP pixel size. There is a high correlation between the predicted SOC contents extracted from the maps and the additional test samples (Figure 9). Figures 6 and 9 show the same tendency of samples with high SOC contents (> 40 g kg^{-1}) being underestimated by the models and samples with a lower SOC content (<10 g kg^{-1}) being overestimated. Generally, the study site is located in a pedologically more homogeneous area which might facilitate SOC modelling. The terrain is undulated in most fields and SOC accumulates in the depressions which is well visible in the SOC maps. As relicts of the last ice age, which formed this area, there are some kettle holes located in the fields. They mainly appear as depressions with dense vegetation and sometimes including small water bodies. By using the soil mask, they have been excluded and appear as small round, oval or longish excluded areas. Additionally, other areas within field boundaries were excluded due to either the presence of dry or green vegetation or high soil moisture. Some areas were falsely included as some roofing materials have similar spectral characteristics to soils and are therefore classified as the latter, e.g., parts of a village in the very southern part. Especially in the EnMAP map the SOC content at the borders of some fields, is predicted to be higher which might be caused by mixed spectral signals that include, e.g., shadows produced by trees surrounding the fields. Additionally, the direction of ploughing is often different at the border of the fields leading to different bidirectional reflectance distribution function (BRDF) effects.

Future work will focus on testing the proposed two-step local PLSR approach on spaceborne imaging spectroscopy data in areas with varying soil properties. Since we applied this approach on one local study site only, the accuracy for other sites should be investigated. Generally, an investigation could be performed of the areas in terms of size, boundaries and change over time on which remote sensing models are valid.

5. Conclusions

In this study, we aimed to test if the local PLSR approach making use of the LUCAS soil database can lead to SOC prediction accuracies at an airborne scale equivalent to results of established traditional approaches. Furthermore, the goal was to evaluate the SOC prediction accuracy of this two-step approach applied to simulated EnMAP satellite imagery. Therefore, we applied the local PLSR approach in two steps which was necessary to account for differences between laboratory and image spectra. First, we predicted the SOC content of the calibration field samples using the local PLSR and LUCAS, and second, we used these SOC contents predicted in the first step and the corresponding image spectra to calibrate remote sensing models. This approach was compared with the traditional way of SOC quantification using measured SOC contents.

We received high SOC prediction accuracies for the first step using the local PLSR and LUCAS. Despite a medium data SNR, due to the high latitude of the area and late acquisition date (October), we obtained good results for the second step using airborne HySpex imagery with $R^2 = 0.67$, RPD = 1.79. The results were slightly lower than for the traditional approach which leads to the conclusion that the two-step approach provides a beneficial alternative. For the simulated EnMAP images the results of the traditional approach were very similar to those of the HySpex image with the same approach, indicating that the coarser spatial and spectral resolution does not lead to a large difference. When applying the two-step approach on simulated EnMAP images the accuracy decreased more to $R^2 = 0.48$ and RPD = 1.41.

Typically, for airborne SOC quantification there is the need for wet chemistry SOC analyses of soil samples from the study site. The advantage of the two-step approach is the usage of the LUCAS database combined with the local PLSR as a reproducible and less expensive alternative to these wet chemistry SOC analyses. Currently, the only in-situ requirements to apply the two-step approach are the need for spectral laboratory measurements with a FOSS device (for comparison with LUCAS) of a few samples from the study site. We could show in this study that a small number of samples was sufficient for this study site and that the model calibrated on one field was transferable to fields in the vicinity. In future studies—and with respect to current and future spaceborne hyperspectral missions providing data for larger areas—a further reduction in spectral measurements from local samples, the spatial extent to which remote sensing models hold their validity, and how to differentiate between potentially different model areas, could be evaluated.

Author Contributions: K.J.W. developed the overall idea and approach supported by S.C., S.F. and F.C.; D.S. organized the flight campaign; M.B. performed the image pre-processing. K.J.W. analyzed the data, performed statistical analysis, modelling, application, and drafted the manuscript. K.J.W., S.C. and S.F. contributed to data interpretation; S.C., S.F., F.C., M.B., D.S. reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict 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. Lal, R. Soil carbon sequestration impacts on global climate change and food security. *Science* **2004**, *304*, 1623–1627. [CrossRef] [PubMed]
2. Tóth, G.; Hermann, T.; da Silva, M.R.; Montanarella, L. Monitoring soil for sustainable development and land degradation neutrality. *Environ. Monit. Assess.* **2018**, *190*, 57. [CrossRef]
3. Wiesmeier, M.; Urbanski, L.; Hobley, E.; Lang, B.; von Lützow, M.; Marin-Spiotta, E.; van Wesemael, B.; Rabot, E.; Ließ, M.; Garcia-Franco, N. Soil organic carbon storage as a key function of soils-a review of drivers and indicators at various scales. *Geoderma* **2019**, *333*, 149–162. [CrossRef]
4. World Meteorological Organization (WMO). The Global Observing System for Climate: Implementation Needs. WMO Pub No. GCOS—200. 2016. Available online: https://public.Wmo.Int/en/programmes/global-climate-observing-system/ (accessed on 7 October 2020).
5. 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]
6. Conant, R.T.; Ogle, S.M.; Paul, E.A.; Paustian, K. Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation. *Front. Ecol. Environ.* **2011**, *9*, 169–173. [CrossRef]
7. Kibblewhite, M.G.; Miko, L.; Montanarella, L. Legal frameworks for soil protection: Current development and technical information requirements. *Curr. Opin. Environ. Sustain.* **2012**, *4*, 573–577. [CrossRef]
8. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthes, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Adv. Agron.* **2015**, *132*, 139–159.
9. Beyer, L.; Kahle, P.; Kretschmer, H.; Wu, Q. Soil organic matter composition of man-impacted urban sites in north germany. *J. Plant Nutr. Soil Sci.* **2001**, *164*, 359–364. [CrossRef]
10. 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]
11. He, T.; Wang, J.; Lin, Z.; Cheng, Y. Spectral features of soil organic matter. *Geo-Spat. Inf. Sci.* **2009**, *12*, 33–40. [CrossRef]
12. Viscarra Rossel, R.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
13. Peón, J.; Recondo, C.; Fernández, S.; F Calleja, J.; De Miguel, E.; Carretero, L. Prediction of topsoil organic carbon using airborne and satellite hyperspectral imagery. *Remote Sens.* **2017**, *9*, 1211. [CrossRef]
14. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Chapter five-visible and near infrared spectroscopy in soil science. *Adv. Agron.* **2010**, *107*, 163–215.
15. Ben-Dor, E.; Chabrillat, S.; Demattê, J.; Taylor, G.; Hill, J.; Whiting, M.; Sommer, S. Using imaging spectroscopy to study soil properties. *Remote Sens. Environ.* **2008**, *113*, S38–S55. [CrossRef]
16. 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]
17. 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]
18. Ben-Dor, E.; Taylor, R.; Hill, J.; Demattê, J.; Whiting, M.; Chabrillat, S.; Sommer, S. Imaging spectrometry for soil applications. *Adv. Agron.* **2008**, *97*, 321–392.
19. Nocita, M.; Stevens, A.; Toth, G.; Panagos, P.; van Wesemael, B.; Montanarella, L. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. *Soil Biol. Biochem.* **2014**, *68*, 337–347. [CrossRef]
20. 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]
21. Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. Prisma: The italian hyperspectral mission. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 175–178.
22. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C. The enmap spaceborne imaging spectroscopy mission for earth observation. *Remote Sens.* **2015**, *7*, 8830–8857. [CrossRef]

23. Chabrillat, S.; Foerster, S.; Steinberg, A.; Segl, K. Prediction of common surface soil properties using airborne and simulated enmap hyperspectral images: Impact of soil algorithm and sensor characteristic. In Proceedings of the 2014 IEEE Geoscience and Remote Sensing Symposium (IGARSS), Quebec City, QC, Canada, 13–18 July 2014; pp. 2914–2917.

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. Wetterlind, J.; Stenberg, B. Near-infrared spectroscopy for within-field soil characterization: Small local calibrations compared with national libraries spiked with local samples. *Eur. J. Soil Sci.* **2010**, *61*, 823–843. [CrossRef]

26. Castaldi, F.; Chabrillat, S.; Don, A.; van Wesemael, B. Soil organic carbon mapping using lucas topsoil database and sentinel-2 data: An approach to reduce soil moisture and crop residue effects. *Remote Sens.* **2019**, *11*, 2121. [CrossRef]

27. 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]

28. Tziolas, N.; Tsakiridis, N.; Ogen, Y.; Kalopesa, E.; Ben-Dor, E.; Theocharis, J.; Zalidis, G. An integrated methodology using open soil spectral libraries and earth observation data for soil organic carbon estimations in support of soil-related sdgs. *Remote Sens. Environ.* **2020**, *244*, 111793. [CrossRef]

29. Castaldi, F.; Chabrillat, S.; Chartin, C.; Genot, V.; Jones, A.; 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]

30. Ward, K.J.; Chabrillat, S.; Foerster, S. LocalPLSR. 2020. Available online: https://github.com/GFZ/LocalPLSR (accessed on 10 October 2020). [CrossRef]

31. Ward, K.J.; Chabrillat, S.; Neumann, C.; Foerster, S. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered lucas soil database. *Geoderma* **2019**, *353*, 297–307. [CrossRef]

32. Zacharias, S.; Bogena, H.; Samaniego, L.; Mauder, M.; Fuß, R.; Pütz, T.; Frenzel, M.; Schwank, M.; Baessler, C.; Butterbach-Bahl, K. A network of terrestrial environmental observatories in germany. *Vadose Zone J.* **2011**, *10*, 955–973. [CrossRef]

33. Spengler, D.; Förster, M.; Borg, E. Editorial. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2018**, *86*, 49–51. [CrossRef]

34. Bundesanstalt für Geowissenschaften und Rohstoffe (BGR). *Bodenübersichtskarte 1:200.000 (bük200)—cc2342 Stralsund*; BGR: Hannover, Germany, 2006.

35. BGR. *Soil Regions Map of the European Union and Adjacent Countries 1:5,000,000 (Version 2.0)*; EU Catalogue Number S.P.I.05.134; Special Publication: Ispra, Italy, 2005.

36. Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O. Lucas soil, the largest expandable soil dataset for europe: A review. *Eur. J. Soil Sci.* **2018**, *69*, 140–153. [CrossRef]

37. Tóth, G.; Jones, A.; Montanarella, L. *Lucas Topsoil Survey: Methodology, Data and Results*; JRC Technical Reports; EUR26102—Scientific and Technical Research Series; Publications Office of the European Union: Luxembourg, 2013; ISSN 1831-9424.

38. Genot, V.; Colinet, G.; Bock, L.; Vanvyve, D.; Reusen, Y.; Dardenne, P. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. *J. Near Infrared Spectrosc.* **2011**, *19*, 117–138. [CrossRef]

39. Udelhoven, T.; Emmerling, C.; Jarmer, T. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. *Plant Soil* **2003**, *251*, 319–329. [CrossRef]

40. Brell, M.; Spengler, D.; Ruhtz, T.; Ward, K.J.; Chabrillat, S.; Segl, K.; Foerster, S.; Itzerott, S. Demmin germany (October 2015)—An enmap preparatory flight campaign. *GFZ Data Serv.* **2020**. [CrossRef]

41. Brell, M.; Spengler, D.; Ruhtz, T.; Ward, K.J.; Chabrillat, S.; Segl, K.; Foerster, S.; Itzerott, S. Demmin, germany (October) 2015—An enmap flight campaign, enmap flight campaigns technical report. *GFZ Data Serv.* **2020**. [CrossRef]

42. Norsk Elektro Optikk. Hyspex. Available online: http://www.hyspex.no/index.php (accessed on 19 May 2015).

43. Brell, M.; Rogass, C.; Segl, K.; Bookhagen, B.; Guanter, L. Improving sensor fusion: A parametric method for the geometric coalignment of airborne hyperspectral and lidar data. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3460–3474. [CrossRef]

44. Richter, R.; Schläpfer, D. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. *Int. J. Remote Sens.* **2002**, *23*, 2631–2649. [CrossRef]

45. Segl, K.; Guanter, L.; Rogass, C.; Kuester, T.; Roessner, S.; Kaufmann, H.; Sang, B.; Mogulsky, V.; Hofer, S. Eetes—The enmap end-to-end simulation tool. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 522–530. [CrossRef]

46. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

47. Signal Developers. Signal: Signal Processing. 2013. Available online: http://r-forge.r-project.org/projects/signal/ (accessed on 5 May 2018).

48. Chabrillat, S.; Eisele, A.; Guillaso, S.; Rogaß, C.; Ben-Dor, E.; Kaufmann, H. *Hysoma: An Easy-to-Use Software Interface for Soil Mapping Applications of Hyperspectral Imagery*; 7th EARSeL SIG Imaging Spectroscopy Workshop: Edinburgh, Scotland, 2011.

49. Chabrillat, S.; Guillaso, S.; Rabe, A.; Foerster, S.; Guanter, L. *From Hysoma to Ensomap-a New Open Source Tool for Quantitative Soil Properties Mapping Based on Hyperspectral Imagery from Airborne to Spaceborne Applications*; (Geophysical Research Abstracts, 18, EGU2016-14697, 2016); General Assembly European Geosciences Union: Vienna, Austria, 2016.

50. 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]

51. EnMAP-Box Developers Enmap-Box 3—A Qgis Plugin to Process and Visualize Hyperspectral Remote Sensing Data. 2019. Available online: www.enmap.org/enmapbox.html (accessed on 5 May 2019).

52. Mevik, B.-H.; Wehrens, R.; Liland, K.H. Pls: Partial Least Squares and Principal Component Regression. R Package Version 2.6-0. 2016. Available online: https://CRAN.R-project.org/package=pls (accessed on 5 May 2018).

53. Ramirez-Lopez, L.; Stevens, A. Resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics. *R Package Version 1.2.2.* **2016**, *1*, 2.

54. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. R Package Version 2.6-7. 2017. Available online: Https://cran.R-project.Org/package=raster (accessed on 5 May 2018).

55. Viscarra Rossel, R.; Jeon, Y.; Odeh, I.; McBratney, A. Using a legacy soil sample to develop a mid-ir spectral library. *Soil Res.* **2008**, *46*, 1–16. [CrossRef]

56. Castaldi, F.; Palombo, A.; Santini, F.; Pascucci, S.; Pignatti, S.; Casa, R. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sens. Environ.* **2016**, *179*, 54–65. [CrossRef]

57. Lu, P.; Niu, Z.; Li, L. Prediction of soil organic carbon by hyperspectral remote sensing imagery. In Proceedings of the 2012 Third Global Congress on Intelligent Systems, Wuhan, China, 6–8 November 2012; pp. 291–293.

58. Blasch, G.; Spengler, D.; Hohmann, C.; Neumann, C.; Itzerott, S.; Kaufmann, H. Multitemporal soil pattern analysis with multispectral remote sensing data at the field-scale. *Comput. Electron. Agric.* **2015**, *113*, 1–13. [CrossRef]

59. Castaldi, F.; Chabrillat, S.; Wesemael, B. Sampling strategies for soil property mapping using multispectral sentinel-2 and hyperspectral enmap satellite data. *Remote Sens.* **2019**, *11*, 309. [CrossRef]

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