Article

Soil Organic Carbon Estimation in Croplands by Hyperspectral Remote APEX Data Using the LUCAS Topsoil Database

Fabio Castaldi [1,*]**, Sabine Chabrillat** [2] iD **, Arwyn Jones** [3]**, Kristin Vreys** [4]**, Bart Bomans** [4] **and Bas van Wesemael** [1] iD

[1] Georges Lemaître Centre for Earth and Climate, Earth and Life Institute, Universite Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; bas.vanwesemael@uclouvain.be

[2] Helmholtz-Zentrum Potsdam—Deutsches GeoForschungsZentrum GFZ, 14473 Potsdam, Germany; chabri@gfz-potsdam.de

[3] European Commission, Directorate General Joint Research Centre (JRC), 21027 Ispra, Italy; arwyn.jones@ec.europa.eu

[4] Flemish Institute for Technological Research, VITO, 2400 Mol, Belgium; kristin.vreys@vito.be (K.V.); bart.bomans@vito.be (B.B.)

* Correspondence: fabio.castaldi@uclouvain.be

Abstract: The most commonly used approach to estimate soil variables from remote-sensing data entails time-consuming and expensive data collection including chemical and physical laboratory analysis. Large spectral libraries could be exploited to decrease the effort of soil variable estimation and obtain more widely applicable models. We investigated the feasibility of a new approach, referred to as bottom-up, to provide soil organic carbon (SOC) maps of bare cropland fields over a large area without recourse to chemical analyses, employing both the pan-European topsoil database from the Land Use/Cover Area frame statistical Survey (LUCAS) and Airborne Prism Experiment (APEX) hyperspectral airborne data. This approach was tested in two areas having different soil characteristics: the loam belt in Belgium, and the Gutland–Oesling region in Luxembourg. Partial least square regression (PLSR) models were used in each study area to estimate SOC content, using both bottom-up and traditional approaches. The PLSR model's accuracy was tested on an independent validation dataset. Both approaches provide SOC maps having a satisfactory level of accuracy (RMSE = 1.5–4.9 g·kg^{-1}; ratio of performance to deviation (RPD) = 1.4–1.7) and the inter-comparison did not show differences in terms of RMSE and RPD either in the loam belt or in Luxembourg. Thus, the bottom-up approach based on APEX data provided high-resolution SOC maps over two large areas showing the within- and between-field SOC variability.

1. Introduction

Many soil chemical components (chromophores) interact with the electromagnetic radiation within the visible-near infrared (VNIR: 400–1300 nm) and short-wave infrared (SWIR: 1300–2500 nm) spectral regions. Soil spectroscopy exploits the correlation between spectral features and chromophores in order to estimate soil variables. The strength of this relationship is not the same for all soil variables, and also the shape of the spectral features can be very different. For example, the metal–OH bending and O–H stretching related to a clay lattice cause a narrow and shallow absorption feature at 2207 nm [1,2]. Overtones of the O–H and H–O–H stretch vibrations in the free water produce two deep absorption features at 1455 nm and 1915 nm, while organic matter has a strong relationship

with electromagnetic radiation in the visible region, due to a wide spectral feature centred around 664 nm related to the chlorophyll pigment [3,4]. However other narrow peaks have a strong relation with organic carbon and organic matter in the SWIR, especially between 2100 nm and 2300 nm [5–7]. In order to exploit spectral features, a suitable spectral resolution is necessary, especially for soil variables affecting spectra in narrow spectral regions (e.g., clay and organic matter). Most of the chemometric approaches to estimate soil variables use the whole spectrum; for example, the partial least square regression (PLSR) method exploits both the wavelengths correlated to the target variable (i.e., absorption features) and other bands indirectly related to the target variable [8]. Thus, in addition to band width, the spectral range is very important to ensure an accurate soil variable estimation by chemometric approaches. PLSR allows summarizing of the spectral information and, therefore, in some cases a very narrow band width is not strictly necessary. Castaldi et al. [9] showed that reducing the spectral resolution from 1 nm to 40 nm did not cause a significant decrease in the estimation accuracy for the SOC, clay, sand and silt content. Spectroscopy, applied in the laboratory, allows estimating soil variables only at the sampling points, while both airborne and satellite data provide quantitative soil information over large areas. As the band width of the main airborne sensors is smaller than 40 nm, their spectral range allows for most of the soil chromophores to be exploited. There is an increasing number of papers concerning the estimation of SOC exploiting airborne hyperspectral data [10–17].

Hyperspectral satellite data could be more attractive than airborne data as satellites, having a short revisit time, allow for the collection of a number of images during a year for the same area from the same sensor under same angle of observation. Unlike airborne data, satellite acquisition ensures a cheap way to obtain spectral data over very large areas. However, the quantitative prediction of soil properties using the first generation of hyperspectral satellite sensors is hampered by the very low signal-to-noise ratio (SNR) in the SWIR region for Hyperion imagers on board of the NASA EO-1 platform [9,18,19], or by the restricted spectral range (415–1050 nm) for the Compact High Resolution Imaging Spectrometer (CHRIS) on the European Space Agency's PROBA platform [9,20–22]. Because of the low SNR, Zhang et al. [23] obtained unreliable PLSR predictions for soil moisture and clay content using Hyperion image reflectance spectra. In order to decrease the SNR of the Hyperion data, Castaldi et al. [19] used the bands retrieved from the minimum-noise fraction transformation (MNF) as regressors in the multivariate models for the clay, sand and organic matter estimation; in this case, the authors managed to improve the estimation accuracy as compared to multispectral data obtained by the Advance Land Imager (ALI). In the near future, at least five satellites equipped with hyperspectral imagers are due to be launched: the German Environmental Mapping and Analysis Program (EnMap) [24], the Italian PRecursore IperSpettrale della Missione Applicativa (PRISMA) [25], the U.S. NASA Hyperspectral Infrared Imager (HyspIRI) [26], the Japanese Hyperspectral Imager Suite (HISUI) [27], the Israeli Hyperspectral imager (SHALOM) [28], and the China Commercial Remote-sensing Satellite System (CCRSS).

Most authors directly calibrate the airborne or satellite spectra on soil samples collected from the corresponding pixels [16,19,21–23,29]. This involves collection of soil samples within bare fields, laboratory measurement of the target variable, and calibration of a multivariate model linking the quantity of the target variable with the spectra extracted from the remote-sensing instrument at the sampling points. This procedure generally ensures a high estimation accuracy within the investigated area, but it entails time-consuming and expensive data collection (sampling, chemical or physical laboratory analysis). The use of other covariates can increase the efficiency of soil-estimation models provided by sampling that is not intensive (i.e., the collection of a few samples in a large area); for example, Casa et al. [21] showed the advantage of merging remote sensing and geophysical data sources using hybrid methods, i.e., statistical techniques combining regression and geostatistical analysis. However, the calibration of empirical models allows only local efficiency and their application to different conditions or areas is often difficult or impossible, even for neighboring fields in cropland as shown by Casa et al. [22] for clay, silt and sand estimation using airborne (MIVIS) and satellite (CHRIS-PROBA) hyperspectral data and PLSR models. The authors compared two different validation

strategies: the random selection of the calibration and validation dataset from two neighboring fields, and the spatially independent validation; i.e., using data from one field for calibration and data from another field for validation, Casa et al. [22] showed a lower accuracy of the spatially independent models as compared to the random strategies, confirming the issues about the extrapolation from empirical models.

In this context, national, continental and global soil spectral libraries could be exploited to decrease the efforts concerning the soil-variable estimation and obtaining more widely applicable models. A consistent number of large soil spectral libraries are available at national [30–33] and continental scale [34,35]. Moreover, a global soil spectral library was created connecting spectral libraries from all over the world [36]. All these libraries provide, in addition to spectral data, analytical data on a number of soil variables, allowing the calibration of multivariate models covering larger soil variability than the models calibrated using local libraries. However, often the large spectral libraries are built collating local libraries that were collected under differing conditions and using different protocols and instruments. Thus, a spectral alignment between libraries is necessary. Recently, Castaldi et al. [37] used the pan-European LUCAS topsoil database to accurately estimate SOC content in two local spectral libraries. They did not need to apply any spectral alignment, because both LUCAS and local spectra were scanned with the same model of spectrometer (FOSS NIR Systems Inc., Laurel, MD, USA), which provided very stable measurements. This approach could allow estimating SOC or other soil variables in croplands in Europe exploiting the LUCAS database and without chemical laboratory analyses. The satisfactory exploitation of the LUCAS database on local spectral libraries was made possible by the high stability of the laboratory spectral measurements. Unfortunately, remote-sensing data are not stable because of the change of the atmospheric and soil conditions between two different acquisition times and the uncertainty of the pre-processing procedure (radiometric, atmospheric and geometric correction). This instability hampers the use of remote-sensing data to build a reliable spectral library and, consequently, the application of a single calibration model on spectral data acquired at different times. At the same time, the future availability of hyperspectral satellite data entails the need for an automated approach that allows estimation of soil variables over large areas, avoiding the calibration of predicted models using remote-sensing spectra of soil samples collected at the location of the image pixels. For this purpose, large spectral libraries could be exploited in synergy with remote-sensing data, but the comparison between remote and laboratory data is hampered by several factors: the difference in support size, the georeferencing error, the spectral characteristics, and the atmospheric effects. Although the remote-sensing data are atmospherically corrected, this correction introduces an uncertainty in the measurements, causing a non-systematic or random difference between laboratory and remote spectra.

Although a consistent number of spectral-transfer procedures were tested between two datasets both composed of laboratory spectra [38–43], the transfer between spectra acquired in laboratory conditions to those acquired by remote sensors is a still rare [44,45]. Moreover, in order to find a reliable transfer function, it is mandatory to collect and scan soil samples. In fact, a transfer function can only be obtained by the comparison between spectra acquired using both the master and slave instrument. The transfer function can be stable in the case of two sensors working in the same conditions, like two laboratory spectrometers, while for the remote data the change in atmospheric conditions between two different acquisition times involves different spectral responses at the same points. This difference entails a new transfer function, but the available bare soil surfaces are not always the same, thus implying the need for new sampling for each remote acquisition.

Here we aim to test a new approach, named bottom-up, that allows avoidance of the aforementioned issues related to the spectral transfer between laboratory and remote-sensing data, while at the same time prepares for an automated approach to estimate soil variables over large areas, as will be required once the hyperspectral satellites will be launched. This approach started from the methodology proposed by [37] to predict the SOC values at sampling points based on the LUCAS spectral library; then these values were linked to the airborne spectra building a PLSR model; and in

the final step the PLSR model was applied to all bare soil pixels of the airborne image producing SOC maps with the same spatial resolution as the airborne data. Thus, this approach allows laboratory analysis of the target variable to be avoided. The new method was compared with the traditional approach, i.e., the calibration of a multivariate model linking remote spectra and the quantity of the target variable measured in the laboratory.

2. Study Area

We chose two study areas with a fair amount of cropland having different soil characteristics: the loam belt in Belgium and the Gutland–Oesling region in Luxembourg.

The northern part of Wallonia is mostly characterized by Quaternary and Tertiary geological formations. Iin particular, the Belgian loam belt is mainly characterized by niveo-eolian sediments dominated by cropland with productive silt loam soils and a rolling topography [46]. The Belgian research area consists of a 7 km-wide and 66 km-long SW–NE strip (NW corner: 50°46′N, 5°5′E; SE corner: 50°32′N, 4°41′E) from Gembloux (Walloon region) to Sint Truiden (Flanders region). The Belgian loam belt has a temperate oceanic climate, the mean annual rainfall is 790 mm, the lowest monthly mean temperature is in January (2.3 °C), and the highest monthly mean in July (17.8 °C). The Luxembourg research area is located in the central and northern part of the Grand Duchy of Luxembourg within the Gutland and Oesling regions (Figure 1) along a 5 km-wide and 29 km-long North–South strip (NW corner: 49°57′N, 6°49′E; SE corner: 49°42′N, 6°11′E). The climate in Luxembourg is temperate oceanic. The mean temperatures ranging from 0.7 °C in winter (January) to 17 °C in summer (July) [46]. The mean temperatures in the north of the country (Oesling region) are slightly lower both in winter and summer. The mean annual rainfall is approximately 800 mm in the central and southern Luxembourg regions and 1000 mm in the north. The Gutland region is located in central Luxembourg and, in the southern part of the image area, it consists of secondary marls, limestone, sandstone and dolomites with a rolling topography. The Gutland region can be divided in two agro-pedological zones: the Redange–Diekirch area with clay, clay-loam and sandy-loam soils and the Central area with clay, sandy-loam and sandy soil [14]. The Oesling region is located in the north of the Grand Duchy of Luxembourg and it is mainly characterized by shallow stony loam soils derived from Devonian slate, quartzite and sandstone substrates on a sub-horizontal plateau with a mean altitude of 450 m incised by the main rivers [47].
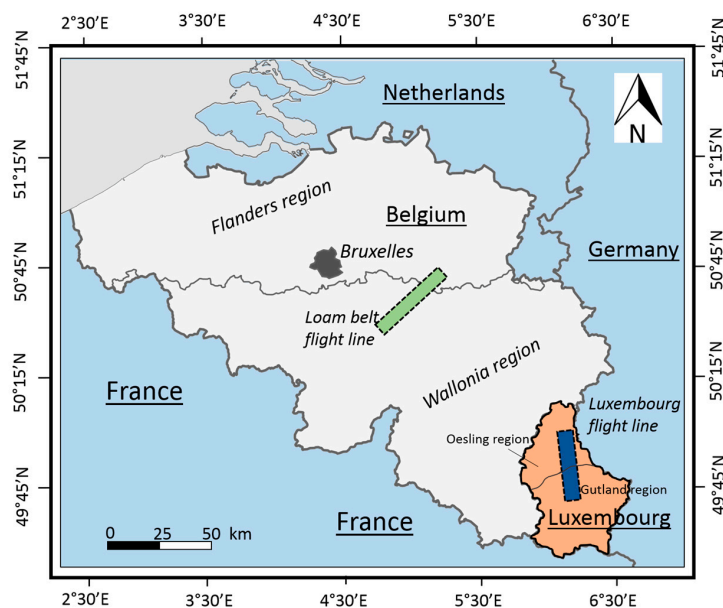

Figure 1. Location of the two study areas in Belgium and Luxembourg over which Airborne Prism Experiment (APEX) data were acquired.

3. Materials and Methods

3.1. Lucas Topsoil Database

The topsoil dataset provided by the Land Use/Cover Area frame statistical Survey (LUCAS) is a pan-European spectral library, composed of 19,967 topsoil samples collected in 2009 throughout 25 countries [35]. The soil samples were air-dried and sieved (2 mm) before the spectral acquisition by means of an XDS Rapid Content Analyzer (FOSS NIR Systems Inc., Laurel, MD, USA) following the protocol described by the manufacturer and the Soil Spectroscopy Group [36]. The diffuse high-resolution reflectance was continuously measured from 400 nm to 1300 nm (VNIR) and from 1300 nm to 2500 nm (SWIR) with a spectral sampling interval of 0.5 nm. In addition to spectral measurements, the LUCAS dataset includes 12 chemical and physical soil properties. The total carbon was measured by dry-combustion using a CN analyser (VarioMax, Elementar Gmbh, Hanau, Germany) and SOC content was obtained subtracting the inorganic carbon by means of measuring the CO_2 pressure upon addition of concentrated HCl [48].

A subset of the LUCAS dataset was created selecting only the sample collected in croplands [37]. This subset, named LUCAS_crop, consisted of 12,128 samples, which were split into seven classes according to a cluster analysis using the k-means algorithm and the Euclidean distance as a metric [49]. The LUCAS clusters were obtained exploiting 12 soil variables: clay, sand, silt, and coarse fragments content, SOC, carbonate content ($CaCO_3$), pH (in H_2O and $CaCl_2$), cation-exchange capacity (CEC), total nitrogen (N), soluble phosphorus (P), and potassium (K). Some of the LUCAS soil variables were correlated among each other, thus a principal component analysis (PCA) was carried out after the standardization of the variables, obtaining uncorrelated principal components (for more details about the cluster analysis on LUCAS_crop data, see [37]).

The LUCAS_crop classes differ in pH, soil texture, SOC and macroelements: the soils of class A are loamy and rich in N, P, K and SOC (29.7 g·kg^{-1}); class B is characterized by acid soils having a sandy loam texture and an average SOC value of 21.5 g·kg^{-1}; class C includes slightly acid soils having a silt loam texture, the mean SOC value is 25.7 g·kg^{-1}; the soils of class D are organic, with a very high SOC content; the soil of class E are basic with a clay loam texture and are rich in $CaCO_3$; class F consists in slightly basic soils with a sandy loam texture and poor in nutrients; the soils of class F are silty clay loam with high K content (326.4 mg·kg^{-1}). The average values of the main soil variables of the seven LUCAS_crop classes are shown in Table 1.

Table 1. Textural class and average values of the main soil variables for the seven classes of the LUCAS_crop dataset.

LUCAS_Crop Class	Textural Class	pH	SOC g·kg^{-1}	N g·kg^{-1}	P mg·kg^{-1}	K g·kg^{-1}	CaCO$_3$ g·kg^{-1}
A	Loam	7.2	29.7	2.8	106.5	764.2	54.5
B	Sandy loam	5.6	21.5	1.8	43.8	111.9	0.4
C	Silt loam	6	25.7	2.3	33	168.1	1.2
D	Organic soil	6	318.6	18	34.4	193.5	11.3
E	Clay loam	8	16.9	1.5	21.5	283.4	391.8
F	Sandy loam	7.4	15.8	1.5	27.9	165.8	30.9
G	Silty clay loam	7.5	25.1	2.4	25.8	326.4	60.8

3.2. Airborne Images

Airborne imagery was acquired with the Airborne Prism Experiment (APEX) sensor mounted on a non-pressurized Dornier 228 aircraft of the German Aerospace Center (DLR). The APEX is an airborne (dispersive push broom) imaging spectrometer recording hyperspectral data in 313 bands (in binned mode) in the wavelength range 400–2500 nm [50]. APEX spectrometer is developed by a Swiss–Belgian consortium on behalf of the European Space Agency. Since 2010 several flight campaigns have been

completed across Europe, covering about 160 areas. The large amount of these airborne data have allowed addressing a wide range of research topics related to remote sensing in land and water [50]. The APEX data were recorded on 1 July 2015 in Belgium (flying height 3900 m) and on 20 March 2015 in Luxembourg (flying height 4700 m). The location of the two study areas is shown in Figure 1. The APEX data were pre-processed at the VITO Remote Sensing Department, using the Central Data Processing Center (CDPC) for airborne Earth observation data [51]. The APEX radiometric and spectral calibration is performed by means of calibration cubes generated from data measured and collected on the APEX Calibration Home Base (CHB) hosted at DLR Oberpfaffenhofen, Germany [52]. The geometric correction of the APEX data is based on direct georeferencing [53]. GPS/IMU data registered during the flight were post processed with Applanix POSPac MMS software, thereby using GPS base stations data. The post-processed position and orientation data, together with boresight correction data, the detailed sensor model and the SRTM DEM were used in the ortho-rectification process. Finally, the images were spatially resampled to an output resolution of 2.5 m and projected to the geographic coordinate system (WGS84). The atmospheric correction of the APEX data is performed using the MODTRAN4 radiative transfer code and reflectance retrieval algorithms given in [54,55]. As a result, the pre-processed image data were provided as geo-referenced at surface reflectance.

3.3. Soil Organic Carbon (SOC) Estimation

SOC estimation using the traditional approach entails merging a soil-variable value obtained by analytical measurements in the laboratory and the spectrum extracted from the APEX imagery at the sampling point, where the former is the dependent variable, while the bands of the electromagnetic spectra are the independent variables [56]. These two types of data are used to build a multivariate model to estimate soil variables in other bare soil pixels where we do not have information about the dependent variable (Figure 2). This approach entails the collection of soil samples, laboratory analysis for each variable, the calibration of a multivariate model using a consistent number of samples, and finally the validation of the model in some points not included in the calibration dataset. We used this traditional approach in order to compare it with an alternative approach, the bottom-up, which we tested, to the best of our knowledge, for the first time.
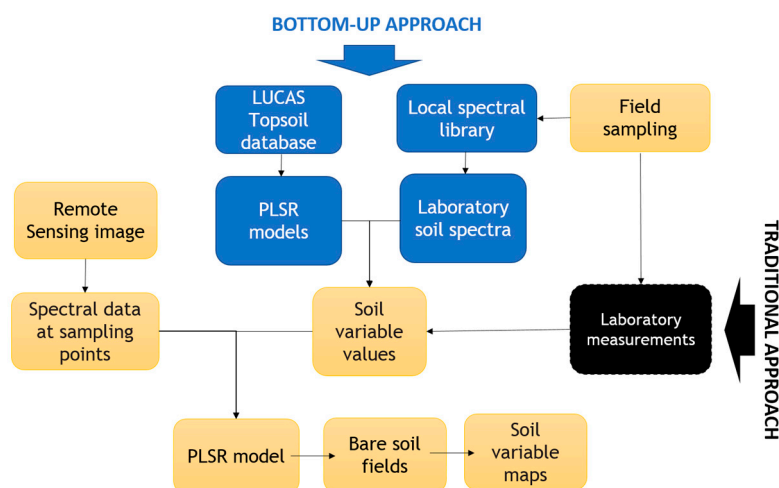

Figure 2. Flow chart concerning the two soil organic carbon (SOC) estimation approaches used in this work: traditional and bottom-up.

The main difference between the traditional and bottom-up approaches is that the latter does not require analytical laboratory measurements. Instead, different soil variables are estimated exploiting laboratory spectral data. The bottom-up approach consists of two main steps: (i) the estimation of a

soil variable at sampling points exploiting the LUCAS spectral library and its ancillary data; and (ii) the mapping of the soil variables using remote-sensing data (Figure 2).

The first step consists in splitting the LUCAS_crop dataset into classes using a cluster analysis (see Section 3.1 and [37] for details). Furthermore, some soil samples are collected within the investigated area. For each sample, the spectrum is acquired using the FOSS spectrometer (calibration spectral dataset). Later, the LUCAS dataset classified according to the cluster analysis is used to train an artificial neural network (ANN), which is used to allocate to each sample of the calibration dataset one of the LUCAS classes (see Table 1). The ANN is a multilayer perceptron feed-forward neural network; the input layer consists of the LUCAS spectral data, while the output layer contains seven neurons, one for each of the detected classes. The number of neurons in the only hidden layer were set observing the cross-validation results of classification of 500 iterations. Then, PLSR models are calibrated for each LUCAS class and they were used to estimate the soil variable on the soil spectra of the calibration datasets of the same class.

In the second step of the bottom-up approach, the previously estimated values were used to calibrate the SOC concentration of the pixels in the APEX imagery corresponding to the sampling points. Thus, a PLSR model was created for the bare soil fields in each area and was subsequently applied to create SOC maps.

In order to reduce propagation of the georeferencing error, the spectra were extracted from APEX images using a bi-linear interpolation algorithm for both the traditional and bottom-up approaches. This method provides the spectrum of an interpolated pixel, taking into account the weighted average of the four pixels nearest to the center of the sample area. Moreover, pre-processing of the spectra consisted of removing the water bands (1349–1446 nm and 1795–1972 nm) and applying a Savitzky–Golay smoothing filter [57] with a second-order polynomial fit and a window size of 11 data points.

3.4. The Calibration and the Independent Validation Datasets

Two soil spectral datasets were collected for calibration in July 2016, consisting of 84 samples in Luxembourg and 54 in Belgium (Figure 1). The calibration samples were collected within fields that were bare during the over flight. We first checked the state of the soil surface during a field survey one day after the flight campaign. Then we also computed the NDVI on the APEX images in order to locate the remainder of the bare soil cropland fields. A sampling campaign was carried out in July 2016 in order to collect at least one sample for each selected field, and the sample location was randomly chosen in a homogeneous area away from field borders. Each calibration sample consists of five sub-samples collected from the 0–10 cm layer taken with a gouge auger within an area of 3 m radius. Soil samples were air dried, gently crushed, and passed through a 2 mm sieve in the laboratory. Each sample was scanned and SOC values were measured (for the traditional approach only) using the same instruments (FOSS spectrometer, model XDS rapid content analyzer and Variomax CN analyzer), and protocols were used to compile the LUCAS dataset (see Section 3.1).

The SOC maps of all bare soils obtained using the bottom-up approach were divided into seven classes based on the standard deviation of the SOC concentration. Then, we selected some fields both for the loam belt and Luxembourg areas (Figure 3) showing a good variability in terms of SOC content. Within these fields, we applied a stratified random strategy to select at least one random point for each of the seven standard deviation classes. The same within-field sampling procedure used for the calibration datasets was adopted for the validation points. All samples were air-dried and sieved (2 mm) and the SOC content was measured by dry-combustion using the same protocol used for the LUCAS dataset and the calibration datasets (see Section 3.1).

The root mean square error (RMSE; Equation (1)) and the ratio of performance to deviation (RPD; Equation (2)) were used as measures for the estimation accuracy.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_o - y_p\right)^2}{n}}, \tag{1}$$

$$\text{RPD} = \frac{std}{\text{RMSE}}, \tag{2}$$

where y_o is the observed and y_p is the predicted value, n the number of the samples, and std the standard deviation.
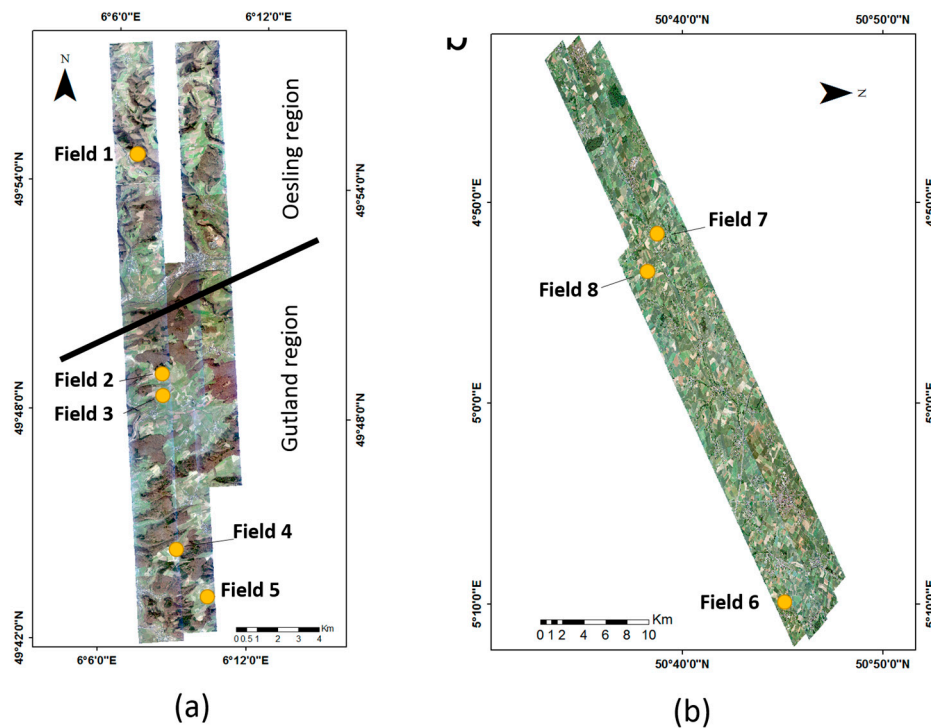

Figure 3. RGB (red: 661 nm; green: 545 nm; blue: 450 nm) airborne images acquired by APEX sensor in Luxembourg (**a**), and Belgium (**b**), and the location of the eight validation fields.

4. Results

4.1. Calibration and Validation Datasets

The traditional and bottom-up approaches were tested in two areas having different geological and pedological characteristics: the loam belt area in Belgium and the Gutland–Oesling region in Luxembourg. The loam belt is quite homogenous in terms of substrate, soil texture and organic carbon content. The niveo-eolian loess deposit uniformly covered all the area, resulting in well-drained silt loam soils [14]. According to the classification of the LUCAS_crop dataset [37], the samples collected in this area for the calibration dataset belong to class F, which is characterized by a low SOC content (Table 1). Actually, both the samples of the calibration and validation dataset in Belgium have a SOC content that ranges between 6.1–15.5 g·kg^{-1} (Table 2).

The Luxembourg area is less homogenous than the loam belt from a geological and geomorphological point of view. The SOC values range from 6.6–49.8 g·kg^{-1} and standard deviation value is considerably higher than in Belgium (Table 2). The measured SOC values of the validation datasets ranges between 6.9–36.0 g·kg^{-1} (Table 2).

Table 2. Descriptive statistics of the soil organic carbon content in the calibration and validation dataset of the loam belt and Luxembourg areas.

Dataset	N°	Minimum /g·kg^{-1}	Maximum /g·kg^{-1}	Mean /g·kg^{-1}	Std /g·kg^{-1}
Calibration					
Luxembourg	84	6.6	49.9	23.4	11.1
Loam belt	54	7.1	15.5	10.2	1.7
Validation					
Luxembourg	43	6.9	36.0	18.9	8.6
Loam belt	45	6.1	14.7	10.2	2.1

4.2. SOC Estimation

The first step of the bottom-up approach consisted of the SOC estimation at sampling points exploiting the LUCAS topsoil database. The SOC content at sampling points in Luxembourg was estimated with a RMSE of 5.4 g·kg^{-1} and a RPD of 2.1 i.e., the calibration (Table 3). In the second step of the bottom-up approach, the estimated SOC values were coupled with the corresponding APEX spectra to calibrate a PLSR model for each flight area. The PLSR models were applied to all bare fields obtaining SOC maps for each study area (Figure 4). The differences between the two study areas in terms of range and standard deviation of SOC values (Table 2) result in different RMSE and RPD values for the independent validation, both for traditional and bottom-up approaches. The independent validation in Luxembourg provided a RMSE values of 4.9 g·kg^{-1} and a RPD of 1.7 for both approaches, while in Belgium, both traditional and bottom-up approach gave a RMSE of 1.5 g·kg^{-1} and a RPD of 1.4 (Table 3; Figure 5).

Table 3. Estimation accuracy in terms of root mean square error (RMSE) and ratio of the performance to deviation (RPD) obtained from the validation datasets using the traditional and bottom-up approaches.

	Traditional Approach		Bottom-Up Approach			
	Validation Dataset		Calibration Dataset		Validation Dataset	
	RMSE g·kg^{-1}	RPD	RMSE g·kg^{-1}	RPD	RMSE g·kg^{-1}	RPD
Luxembourg	4.9	1.7	5.4	2.1	4.9	1.7
Loam belt	1.5	1.4	1.2	1.4	1.5	1.4
Total	3.6	2.1	4.3	2.5	3.6	2.1

The SOC maps in Luxembourg highlighted a much higher SOC content in the northern area than in the central and southern areas (Figure 6). The SOC ranges between 19.7–51 g·kg^{-1} with a mean value of 31 g·kg^{-1} in the north (Figure 6a), while the minimum and maximum values in the central area are, respectively, 2.8 g·kg^{-1} and 39.3 g·kg^{-1} and the mean SOC content is 19.4 g·kg^{-1} (Figure 6b). This north–south gradient in SOC content is in agreement with the findings in Stevens et al. [14] and it is probably due to wetter and colder conditions in the northern plateau which allow a general higher SOC content as compared with the values of the central and southern area. Most of the calibration samples located in the northern area were classified in group A according to the LUCAS_crop classification; this class is mainly characterized by quite high SOC content (Table 1). The substrate of the central and southern area (Gutland region) consists of marls and limestone and most of the calibration samples of this region were classified as B and F according to the LUCAS_crop dataset. These two types of soils are poorer than those of class A, their texture is sandy loam with high sand content and low SOC concentration. The SOC maps obtained by the bottom-up approach in Belgian loam belt confirmed a greater soil homogeneity of this area as compared to Luxembourg (Figure 4b).
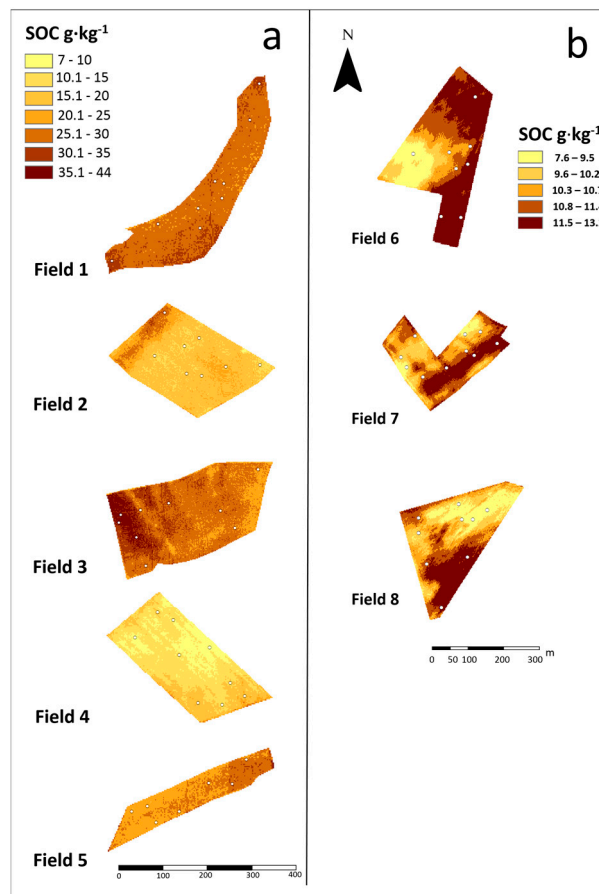
Figure 4. SOC maps of the validation fields in Luxembourg (**a**), and Belgium (**b**), using the bottom-up approach. The white points in the fields correspond to the validation dataset.
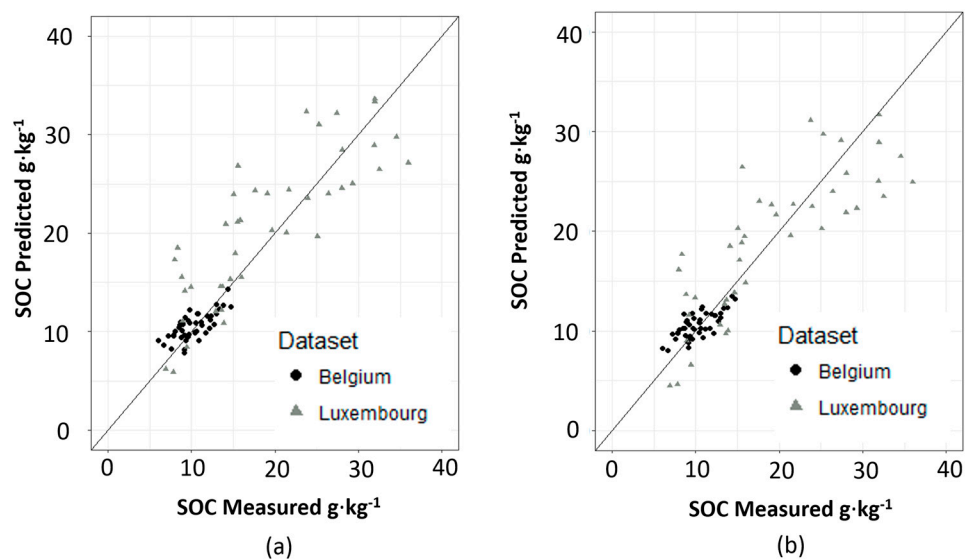

Figure 5. Plot of measured against estimated soil organic carbon content obtained by traditional (**a**), and bottom-up (**b**), approaches on the independent validation datasets in Belgium and Luxembourg.
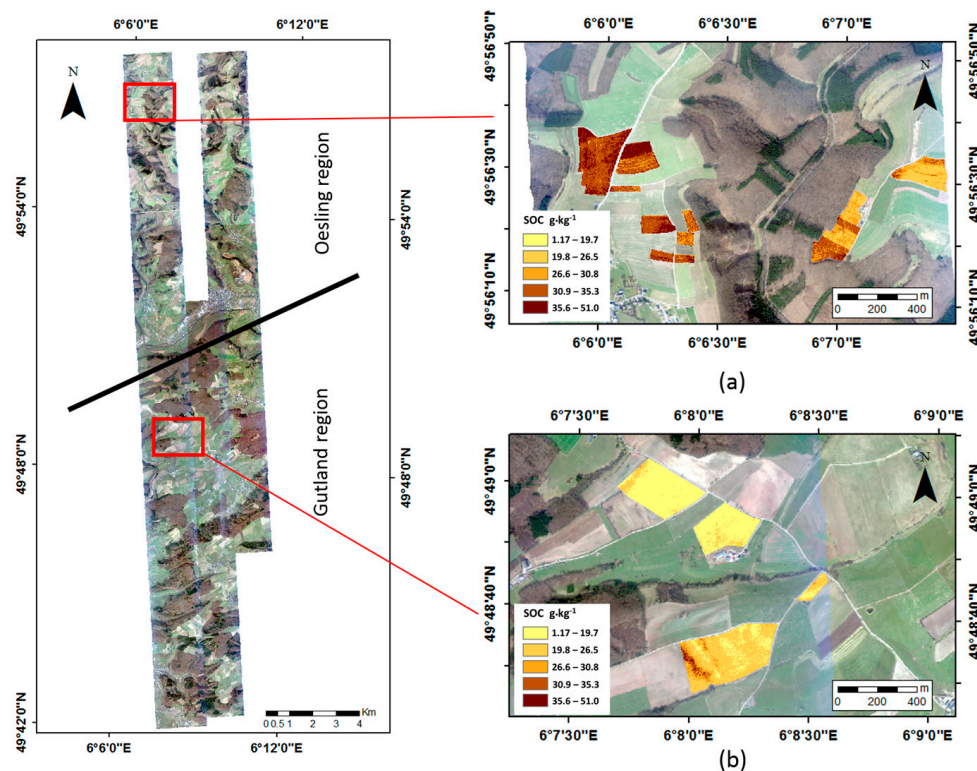
Figure 6. RGB (red: 661 nm; green: 545 nm; blue: 450 nm) airborne images acquired by APEX sensor in Luxembourg (left), and some SOC maps obtained by the bottom-up approach in the northern (**a**), and central (**b**), areas.

The digital terrain model (DTM) of Wallonia (1 m of resolution) and that of Luxembourg (5 m of resolution) allow a better interpretation of the SOC patterns within the fields. In fact, for both study areas the intra-field variability is mainly due to erosion and sedimentation of topsoil. Higher SOC content can be detected along the downslope field border in field 1 in Luxembourg (Figures 4a and 7a) and in most of the fields in the loam belt (Figure 4b). Although the slope is gentle, the higher SOC values in field 7 are close to the bottom of the field along a strip quite parallel to the contours (Figure 8a); and also for the field in Figure 8b, the highest SOC values were detected in the southern part of the field bordering a small river valley. SOC content is higher to the top and bottom of field 4 (Figure 7b). Field 5 is located in a very flat area and this entails a very low in-field SOC variability; however, there is a clear-cut area in the north-eastern part of the field showing higher SOC content, and this is probably because the neighboring woodland extended in the past in the field. Although most of the SOC maps confirm the relevance of the slope for the within-field variability, some differences, especially between contiguous fields, are due to historical differences in management or land use, as can be clearly seen from the rectangular patterns in the fields in the eastern part of Figure 6a.
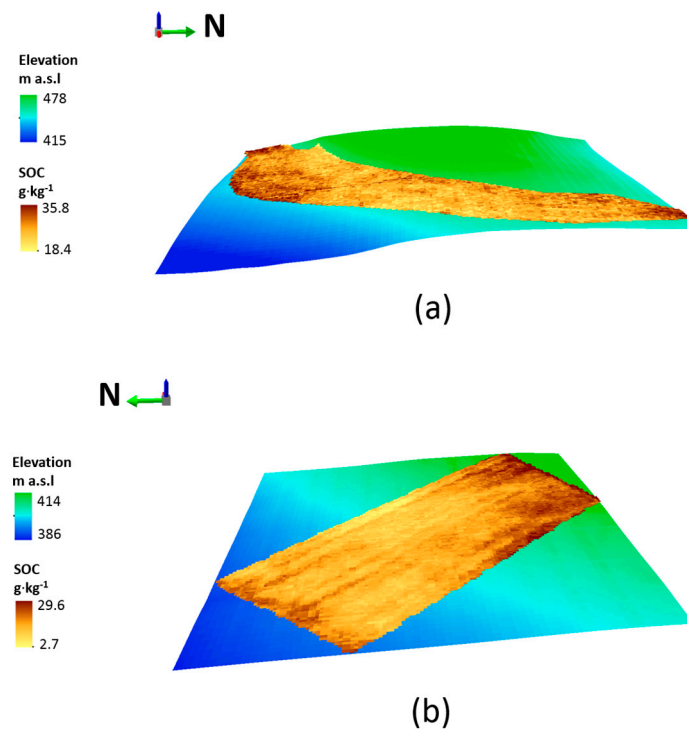
Figure 7. Soil organic carbon maps of the field 1 (**a**) and field 4 (**b**) obtained by the bottom-up approach in the loam belt area laid on top of the digital terrain model (DTM) of the Luxembourg region with a 5 m resolution.
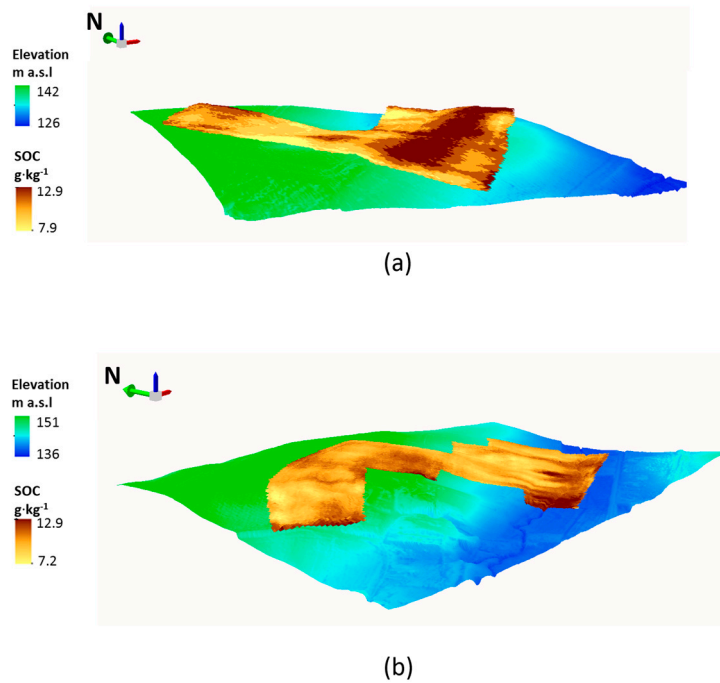

Figure 8. Soil organic carbon maps of the field 7 (**a**) and another bare field (**b**) obtained by the bottom-up approach in the loam belt area laid on top of the digital terrain model (DTM) of the Wallonia region with a 1 m resolution.

5. Discussion

In the first step of the bottom-up approach we obtained the estimation of SOC values at sampling points with a sufficient level of accuracy (i.e., calibration in Table 3) and without new analytical SOC measurements. The SOC estimation was gained from calibration models obtained by subsetting the LUCAS soil database into classes based on 12 soil variables, and the comparison between measured and estimated SOC values showed an overall RMSE of 4.3 g·kg^{-1} and a RPD of 2.5 for the calibration datasets (Table 3). Nocita et al. [58] and Clairotte et al. [59] showed similar results in terms of RPD subsetting large spectral library by means of a local PLSR models; however, the validation datasets are not completely independent for both studies, because the samples used for the validation were selected from the same library as that used for the calibration, i.e., the LUCAS dataset for [58] and the French national spectral library for [59].

Remote-sensing data (airborne and satellite) are used to obtain quantitative soil variables estimation for a large area [29,46,56]. For this purpose, most studies report that a calibration dataset is compiled consisting of measured soil values collected at points corresponding to the remote spectral data [56]. The calibration dataset is generally used to build a multivariate calibration model to estimate the soil properties across the remote images (traditional approach). Theoretically, large spectral libraries, such as the European LUCAS soil database, could be exploited not only to estimate point data, but also to build estimation models of the soil variables at large scale, thus avoiding the collection of a calibration dataset. Unfortunately, the existing large spectral libraries provide soil spectra acquired in laboratory conditions, thus using spectrometers conceived to provide high-resolution spectral measurements in a controlled environment. Therefore, the main challenge is to find a method to align laboratory and remote spectral data [45,60]. Many issues need to be addressed for this goal: first of all, it is necessary to consider the different technical characteristics of the two sensors [40], in particular the full width at half maximum (FWHM) response at each band, the spectral sampling interval and the spectral range, which entail a systematic differences between sensors [7,45]. However, when moving from laboratory to field conditions the radiometric resolution and the SNR hamper the alignments of dissimilar spectral outputs, and the atmospheric correction introduces non-systematic or random differences between laboratory and remote spectra. Other important factors concerning the alignment between laboratory and remote-sensing data are the different spatial resolution, the effect of the georeferencing accuracy, and the different soil conditions at acquisition time (soil moisture, roughness) [61]. The uncertainty of the pre-processing procedure, and the changing atmospheric conditions at each acquisition time [45], do not allow stable spectral data to be obtained at the same location. In other words, the remote-sensing spectra extracted in the same geographical point but at different dates will never be identical. In order to verify the uncertainty of the spectral data from remote sensors and, consequently, their instability, we compared the spectra acquired over a dark (Figure 9a,b) and bright (Figure 9c,d) reference surface both by Analytical Spectral Devices (ASD) Field Spec Fr Pro spectroradiometer (ASD Inc., Boulder, CO, USA) (Figure 9a,c) in the field and by the APEX airborne sensor (Figure 9b,d). The measurements were carried out in 2013 and 2015 in the Belgian loam belt area, and highlight a consistently greater measurement stability using ASD as compared to remote acquisition (Figure 9).
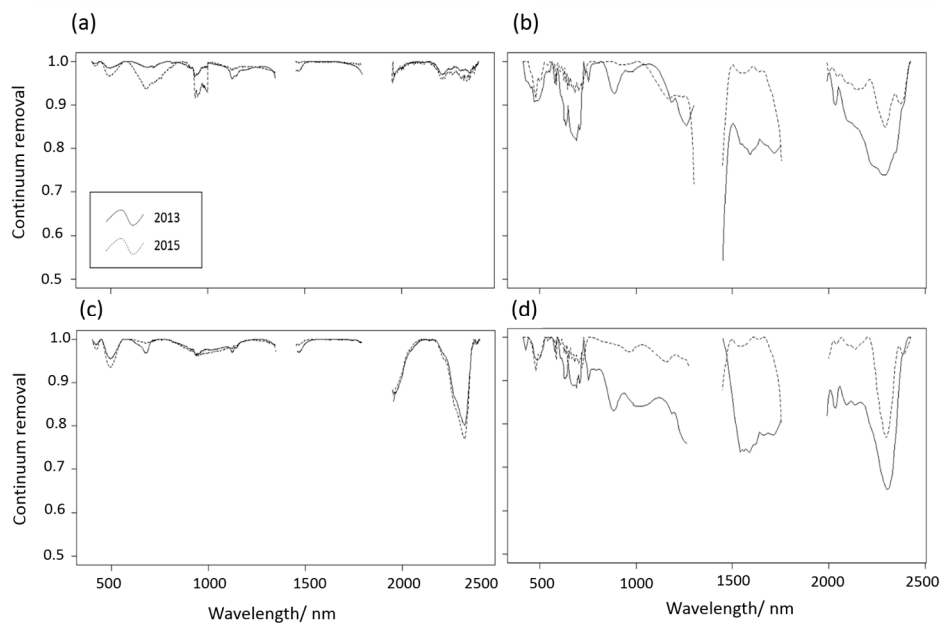
Figure 9. Spectral comparison between Analytical Spectral Devices (ASD) (**a**,**c**) and APEX (**b**,**d**) measurements acquired on dark (**a**,**b**), and bright (**c**,**d**), reference surfaces in the loam belt area in 2013 and 2015.

In light of the results of this comparison, a different approach is necessary to exploit the spectral libraries over a large area. In this work we proposed the bottom-up approach to bypass all the issues that hamper the stability of the remote-sensing data and, consequently, the alignment between laboratory and remote spectra. The main advantage of the bottom-up as compared to the traditional approach is that the chemical analyses in the calibration dataset are replaced by spectral measurements. Thus, the traditional approach uses the measured soil variable values, while the bottom-up approach uses the estimated values obtained by means of the LUCAS library (Figure 2). Another advantage in connection with the use of the bottom-up approach is that it could estimate different soil variables with the same spectral data; in fact, the LUCAS dataset includes 12 soil variables, most of them linked with spectral features in the VNIR region. Thus, the bottom-up approach could be useful for obtaining maps of different soil variables at regional scale on bare fields without chemical or physical laboratory measurements and without any spectral transfer between laboratory and remote spectra. The application of a spectral-transfer function entails the collection of a specific spectral dataset, where each sample, named standard, is measured both with master and slave instruments. Although Nouri et al. [45] successfully applied some transfer methods from laboratory to airborne data to map clay content, their results were dependent on the quantity of the standards, in other words the best estimation performances were obtained with the maximum number of standards available (64 samples). Moreover, the collection of new standards is necessary for each remote acquisition because of the change in atmospheric conditions between two different acquisition times.

The traditional and bottom-up approaches provided very similar estimation accuracy on the independent validation datasets (Table 3; Figure 5). Both approaches allowed estimating SOC with a sufficient degree of accuracy (RMSE = 1.5–4.9 g·kg^{-1}; RPD = 1.4–1.7) and using the same number of samples for the calibration datasets. The optimum number of samples for the bottom-up approach, or more in general for all remote-sensing applications, depends on the characteristics of the investigated area and the range of the target variable, but usually the information about the range or the variability of the target variable are not available in advance, making it difficult to set the optimum number of samples for the bottom-up approach.

Many authors report SOC estimation from airborne data, especially exploiting the VNIR–SWIR region [10–13,16,62,63], but also the long-wave infrared region (LWIR; 8–14 μm) [15]. However, all of them build a calibration dataset for the airborne spectra on samples analysed in the laboratory (i.e., the traditional approach). The HyMap hyperspectral sensor (400–2450 nm, with a spectral resolution of 12–17 nm) has been tested for SOC estimation, with mitigated results: Gomez et al. [12] gained very poor results in southern France (RPD = 0.99; RMSE = 2.6 g·kg^{-1}) as a result of the very low SOC concentrations. Hbirkou et al. [13] obtained a RPD of 2.32 and a RMSE of 1.1 g·kg^{-1} in northern Germany with higher and more variable SOC concentrations. The aisaDUAL hyperspectral scanner (367 bands between 400 nm and 2500 nm) provided good cross-validation results in Germany [63]; in this case, the RMSE was 2.7 g·kg^{-1}, but the standard deviation was quite high (8.2 g·kg^{-1}). The full VNIR–SWIR data (400–2500 nm) from the airborne hyperspectral scanner (AHS) was employed by Stevens et al. [11] in Belgium, obtaining a RPD of 1.47; while Stevens et al. [46] in Luxembourg discarded the SWIR region because of the low SNR, thus using only 20 bands from 442 nm and 1019 nm, and they obtained a RPD of 1.9. Most of the aforementioned papers provided cross-validation statistics [10–12,63], and only Hbirkou et al. [13] and Stevens et al. [46] tested their models on a validation dataset. In both cases, the entire dataset was split into calibration and validation ensuring a representative subset of the data both in the calibration and validation datasets. Here, we tested the PLSR estimation models on a completely independent validation dataset, obtaining an estimation accuracy similar to those obtained by most of the authors cited above who used the traditional approach. In particular, the results obtained for the loam belt area (RPD = 1.40) are very similar to those obtained by Stevens et al. [11] from leave-one-out cross validation in Belgium (RPD = 1.47), while the results that we gained in Luxembourg (RPD = 1.70) are slightly lower to those obtained by [46] in the same area (RPD = 1.90).

6. Conclusions

We investigated the feasibility of a new approach, named bottom-up, which allows SOC maps to be provided over a large area by APEX airborne data without recourse to chemical analyses or any spectral transfer between laboratory and remote spectra, by exploiting the spectra and ancillary data of the European LUCAS topsoil database. This approach was tested on fields that were bare during the over flight in two areas with different soil characteristics i.e., the loam belt in central Belgium and a north–south flight strip across Luxembourg. The results were compared to those provided by a direct calibration of the airborne spectra. The comparison between the independent validations of these two approaches did not show differences in terms of RMSE or RPD. Thus, the new approach tested in this work (bottom-up) allowed SOC maps to be obtained without chemical laboratory analyses over two large areas with the same satisfactory level of accuracy (RPD $\geq$ 1.4) obtainable by a traditional approach. Moreover, the two steps of the bottom-up approach allowed bypassing for the first time (to our knowledge) the issues and restrictions related to the spectral transfer between laboratory and remote-sensing data, in this case between LUCAS spectral data obtained by FOSS spectrometer and APEX spectra.

The loam belt area, due to its homogenous geological substrate, did not show any SOC gradient along the flight line (SW–NE). While clear differences between the northern and southern parts of the flight line exist in Luxembourg, this is due to wetter and colder conditions in the northern plateau, which allow a generally higher SOC content as compared with the values of the central and southern area. The within-field variability could be nearly always explained by the geomorphology in these two areas; in fact, by overlaying SOC maps on a DTM it has was possible to show that the areas along downslope field borders or in zero-order valleys are characterized by higher SOC content than the top of the fields, also with gentle slopes, as is the case in the loam belt. We conclude that the bottom-up approach provided the same SOC estimation accuracy as the traditional approach in the loam belt and Luxembourg areas, and that it is able to explain the SOC variability both at field and regional scale, increasing the level of knowledge of SOC availability for agricultural and environmental management.

In light of these promising results, further research will need to focus on the accuracy level of the bottom-up approach using simulated or real satellite data achievable from the new generation of hyperspectral imagers, and also to investigate an algorithm that allows the optimum number of samples for the calibration dataset to be set. Since the efficiency of the bottom-up approach is strongly affected by the LUCAS dataset, it will be appropriate to test the method in different soil and climate regions, covered by the LUCAS dataset, thereby eventually investigating the potential for predicting other soil variables such as clay content or CEC linked with spectral features in the VNIR region.

Author Contributions: Fabio Castaldi, Bas van Wesemael and Sabine Chabrillat conceived, designed and performed the experiments. Sabine Chabrillat, Arwyn Jones and Fabio Castaldi analyzed the data. Bart Bomans organized the flight campaign. Kristin Vreys performed the image processing. Fabio Castaldi performed statistical analysis and modeling. All authors contributed to writing the paper.

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

References

1. Chabrillat, S.; Goetz, A.F.; Krosley, L.; Olsen, H.W. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. *Remote Sens. Environ.* **2002**, *82*, 431–445. [CrossRef]
2. 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]
3. 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]
4. He, T.; Wang, J.; Lin, Z.; Cheng, Y. Spectral features of soil organic matter. *Geo-Spat. Inf. Sci.* **2009**, *12*, 33–40. [CrossRef]
5. Mouazen, A.M.; Maleki, M.R.; De Baerdemaeker, J.; Ramon, H. On-line measurement of some selected soil properties using a VIS–NIR sensor. *Soil Tillage Res.* **2007**, *93*, 13–27. [CrossRef]
6. Rossel, R.A.V.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]
7. Ben-Dor, E.; Chabrillat, S.; Dematte, J.A.M.; Taylor, G.R.; Hill, J.; Whiting, M.L.; Sommer, S. Using Imaging Spectroscopy to study soil properties. *Remote Sens. Environ.* **2009**, *113*, S38–S55. [CrossRef]
8. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
9. 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]
10. 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]
11. Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma* **2008**, *144*, 395–404. [CrossRef]
12. 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*, 176–185. [CrossRef]
13. Hbirkou, C.; Pätzold, S.; Mahlein, A.-K.; Welp, G. Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale. *Geoderma* **2012**, *175*, 21–28. [CrossRef]
14. Stevens, A.; Miralles, I.; van Wesemael, B. Soil Organic Carbon Predictions by Airborne Imaging Spectroscopy: Comparing Cross-Validation and Validation. *Soil Sci. Soc. Am. J.* **2012**, *76*, 2174–2183. [CrossRef]

15. Pascucci, S.; Casa, R.; Belviso, C.; Palombo, A.; Pignatti, S.; Castaldi, F. Estimation of soil organic carbon from airborne hyperspectral thermal infrared data: A case study. *Eur. J. Soil Sci.* **2014**, *65*, 865–875. [CrossRef]

16. 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]

17. Liu, H.; Shi, T.; Chen, Y.; Wang, J.; Fei, T.; Wu, G. Improving Spectral Estimation of Soil Organic Carbon Content through Semi-Supervised Regression. *Remote Sens.* **2017**, *9*, 29. [CrossRef]

18. Kruse, F.A.; Boardman, J.W.; Huntington, J.F. Comparison of airborne hyperspectral data and eo-1 hyperion for mineral mapping. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1388–1400. [CrossRef]

19. Castaldi, F.; Casa, R.; Castrignanò, A.; Pascucci, S.; Palombo, A.; Pignatti, S. Estimation of soil properties at the field scale from satellite data: A comparison between spatial and non-spatial techniques. *Eur. J. Soil Sci.* **2014**, *65*, 842–851. [CrossRef]

20. Villafranca, A.G.; Corbera, J.; Martín, F.; Marchán, J.F. Limitations of Hyperspectral Earth Observation on Small Satellites. *J. Small Satell.* **2012**, *1*, 19–29.

21. Casa, R.; Castaldi, F.; Pascucci, S.; Basso, B.; Pignatti, S. Geophysical and Hyperspectral Data Fusion Techniques for In-Field Estimation of Soil Properties. *Vadose Zone J.* **2013**, *12*. [CrossRef]

22. Casa, R.; Castaldi, F.; Pascucci, S.; Palombo, A.; Pignatti, S. A comparison of sensor resolution and calibration strategies for soil texture estimation from hyperspectral remote sensing. *Geoderma* **2013**, *197*, 17–26. [CrossRef]

23. Zhang, T.; Li, L.; Zheng, B. Estimation of agricultural soil properties with imaging and laboratory spectroscopy. *J. Appl. Remote Sens.* **2013**, *7*, 073587. [CrossRef]

24. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sens.* **2015**, *7*, 8830–8857. [CrossRef]

25. Pignatti, S.; Acito, N.; Amato, U.; Casa, R.; Castaldi, F.; Coluzzi, R.; De Bonis, R.; Diani, M.; Imbrenda, V.; Laneve, G.; et al. Environmental products overview of the Italian hyperspectral prisma mission: The SAP4PRISMA project. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 3997–4000.

26. Houborg, R.; Anderson, M.; Gao, F.; Schull, M.; Cammalleri, C. Monitoring water and carbon fluxes at fine spatial scales using HyspIRI-like measurements. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 7302–7305.

27. Tanii, J.; Iwasaki, A.; Kawashima, T.; Inada, H. Results of evaluation model of Hyperspectral Imager Suite (HISUI). In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 131–134.

28. Staenz, K.; Mueller, A.; Heiden, U. Overview of terrestrial imaging spectroscopy missions. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium—IGARSS, Melbourne, Australia, 21–26 July 2013; pp. 3502–3505.

29. 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]

30. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988–998. [CrossRef]

31. Knadel, M.; Deng, F.; Thomsen, A.; Greve, M. Development of a Danish national Vis-NIR soil spectral library for soil organic carbon determination. In *Digital Soil Assessments and Beyond*; CRC Press: Boca Raton, FL, USA, 2012; pp. 403–408.

32. Peng, Y.; Knadel, M.; Gislum, R.; Deng, F.; Norgaard, T.; Wollesen de Jonge, L.; Moldrup, P.; Humlekrog Greve, M. Predicting soil organic carbon at field scale using a national soil spectral library. *J. Near Infrared Spectrosc.* **2013**, *21*, 213–222. [CrossRef]

33. Guerrero, C.; Wetterlind, J.; Stenberg, B.; Mouazen, A.M.; Gabarrón-Galeote, M.A.; Ruiz-Sinoga, J.D.; Zornoza, R.; Viscarra Rossel, R.A. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? *Soil Tillage Res.* **2016**, *155*, 501–509. [CrossRef]

34. Garrity, D.; Bindraban, P. *ICRAF A Globally Distributed Soil Spectral Library Visible Near Infrared Diffuse Reflectance Spectra*; ICRAF (World Agroforestry Centre)/ISRIC (World Soil Information) Spectral Library: Nairobi, Kenya, 2015.

35. 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]

36. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Demattê, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. A global spectral library to characterize the world's soil. *Earth-Sci. Rev.* **2016**, *155*, 198–230. [CrossRef]

37. Castaldi, F.; Chabrillat, S.; Chartin, C.; Genot, V.; Jones, A.R.; van Wesemael, B. Using LUCAS topsoil database to estimate soil organic carbon content in croplands sampled in Belgium and Luxembourg. *Eur. J. Soil Sci.* **2017**, *19*, 6383.

38. Shenk, J.S.; Westerhaus, M.O. New Standardization and Calibration Procedures for Nirs Analytical Systems. *Crop Sci.* **1991**, *31*, 1694–1696. [CrossRef]

39. Bouveresse, E.; Massart, D.L. Standardisation of near-infrared spectrometric instruments: A review. *Vib. Spectrosc.* **1996**, *11*, 3–15. [CrossRef]

40. Fearn, T. Standardisation and calibration transfer for near infrared instruments: A review. *J. Near Infrared Spec.* **2001**, *9*, 229–244. [CrossRef]

41. Brown, D.J. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma* **2007**, *140*, 444–453. [CrossRef]

42. Minasny, B.; McBratney, A.B.; Bellon-Maurel, V.; Roger, J.-M; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* **2011**, *167*, 118–124. [CrossRef]

43. Kopačková, V.; Ben-Dor, E. Normalizing reflectance from different spectrometers and protocols with an internal soil standard. *Int. J. Remote Sens.* **2016**, *37*, 1276–1290. [CrossRef]

44. Schwanghart, W.; Jarmer, T. Linking spatial patterns of soil organic carbon to topography—A case study from south-eastern Spain. *Geomorphology* **2011**, *126*, 252–263. [CrossRef]

45. 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]

46. Stevens, A.; Udelhoven, T.; Denis, A.; Tychon, B.; Lioy, R.; Hoffmann, L.; van Wesemael, B. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. *Geoderma* **2010**, *158*, 32–45. [CrossRef]

47. Goidts, E.; van Wesemael, B. Regional assessment of soil organic carbon changes under agriculture in Southern Belgium (1955–2005). *Geoderma* **2007**, *141*, 341–354. [CrossRef]

48. Sherrod, L.A.; Dunn, G.; Peterson, G.A.; Kolberg, R.L. Inorganic Carbon Analysis by Modified Pressure-Calcimeter Method. *Soil Sci. Soc. Am. J.* **2002**, *66*, 299–305. [CrossRef]

49. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *Appl. Stat.* **1979**, *28*, 100–108. [CrossRef]

50. Vreys, K.; Iordache, M.-D.; Bomans, B.; Meuleman, K. Data acquisition with the APEX hyperspectral sensor. *Misc. Geogr.* **2016**, *20*, 5–10. [CrossRef]

51. Biesemans, J.; Sterckx, S.; Knaeps, E.; Vreys, K.; Adriaensen, S.; Hooy-berghs, J.; Meuleman, K.; Kempeneers, P.; Deronde, B.; Everaerts, J.; et al. Image processing workflows for airborne remote sensing. In Proceedings of the 5th EARSeL Workshop on Imaging Spectroscopy, Bruges, Belgium, 23–25 April 2007.

52. Gege, P.; Fries, J.; Haschberger, P.; Schötz, P.; Schwarzer, H.; Strobl, P.; Suhr, B.; Ulbrich, G.; Jan Vreeling, W. Calibration facility for airborne imaging spectrometers. *ISPRS J. Photogramm. Remote Sens.* **2009**, *64*, 387–397. [CrossRef]

53. Vreys, K.; Iordache, M.-D.; Biesemans, J.; Meuleman, K. Geometric correction of APEX hyperspectral data. *Misc. Geogr.* **2016**, *20*, 11–15. [CrossRef]

54. De Haan, J.F.; Hovenier, J.W.; Kokke, J.M.M.; van Stokkom, H.T.C. Removal of atmospheric influences on satellite-borne imagery: A radiative transfer approach. *Remote Sens. Environ.* **1991**, *37*, 1–21. [CrossRef]

55. Haan, J.F.; Kokke, J.M.M. Remote sensing algorithm development. In *Operationalization of Atmospheric Correction Methods for Tidal and Inland Waters*; Toolkit, I., Ed.; Netherlands Remote Sensing Board (BCRS): Delft, The Netherlands, 1996; ISBN 9789054112044.

56. Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **2011**, *162*, 1–19. [CrossRef]

57.   Savitzky, A.; Golay, M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]

58.   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]

59.   Clairotte, M.; Grinand, C.; Kouakoua, E.; Thébault, A.; Saby, N.P.A.; Bernoux, M.; Barthès, B.G. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma* **2016**, *276*, 41–52. [CrossRef]

60.   Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Ben Dor, E.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. *Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring*; Academic Press Inc.: Cambridge, MA, USA, 2015; Volume 132, pp. 139–159. ISBN 9780128021354.

61.   Woodcock, C.E. Uncertainty in Remote Sensing. In *Uncertainty in Remote Sensing and GIS*; John Wiley & Sons, Ltd.: Chichester, UK, 2006; pp. 19–24. ISBN 9780470035269.

62.   Bradley, K.C.; Bowen, S.; Gross, K.C.; Marciniak, M.A.; Perram, G.P. Imaging Fourier transform spectrometry of jet engine exhaust with the telops FIRST-MWE. In Proceedings of the 2009 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2009; pp. 1–8.

63.   Kanning, M.; Siegmann, B.; Jarmer, T. Regionalization of Uncovered Agricultural Soils Based on Organic Carbon and Soil Texture Estimations. *Remote Sens.* **2016**, *8*, 927. [CrossRef]