QUANTITATIVE RETRIEVAL OF SOIL ORGANIC CARBON USING LABORATORY SPECTROSCOPY AND SPECTRAL INDICES

Harm Bartholomeus[a,*], Michael Schaepman[a], Lammert Kooistra[a], Antoine Stevens[b], Willem Hoogmoed[c], Otto Spaargaren[d]

[a]Centre for Geo-Information, Wageningen University,Droevendaalsesteeg 3, NL 6708 PB, Wageningen, The Netherlands,
[b] Université Catholique de Louvain, 3 Place Pasteur, 1348, Louvain la Neuve, Belgium
[c]Farm Technology Group, Wageningen University, Bornsesteeg 59, NL 6708 PD, Wageningen, The Netherlands
[d]ISRIC - World Soil Information, Duivendaal 9, NL 6701 AR, Wageningen, The Netherlands

KEY WORDS: Soil Organic Carbon, Spectroscopy, Spectral Indices

ABSTRACT:

Soil Organic Carbon (SOC) has been identified as one of the major C sinks in the global carbon cycle, of which the exact size and spatial distribution are still difficult to determine quantitatively. Estimation of the amount of SOC present using remote sensing is mostly based on the overall decrease in reflectance in the solar reflective part of the electromagnetic spectrum. However, moisture content and soil roughness result in a comparable decrease, resulting in ambiguous identification of a specific soil type. Depending on the decomposition stage, SOC contains biochemical constituents like lignin and cellulose. Absorption features related to these constituents can be used to determine the SOC content of the soil. We investigated nine different soil types (n=40), originating from a wide range of climatic zones and a large variety in SOC content (0.06 – 45.1%). Spectral measurements of all soil samples were performed in a controlled laboratory environment. The ability of several spectral indices related to biochemical constituents' detection towards the quantification of SOC were tested. Good relations were found for indices based on the visible part of the spectrum and for the absorption features related to cellulose. Cross validation was used to evaluate the predictive capacity of the spectral indices. The results demonstrate that it is feasible to use spectral indices derived from laboratory measurements to predict SOC in various soil types. The results allow establishing a perspective towards spatial distributed mapping of SOC using imaging spectroscopy.

1. INTRODUCTION

Soil Organic Carbon (SOC) has been identified as being one of the major C sinks in the global carbon cycle. The exact size of this sink and its spatial distribution are difficult to determine. Remote Sensing is one of the techniques that can be used for a better estimation of the amount of SOC. Estimation of SOC with remote sensing is mostly based on the overall decrease in reflectance in the reflective part of the spectrum ((Chen et al., 2000; Irons et al., 1989)). However, existing statistical models are in general useable in areas with a limited geographical extent. Furthermore, published relationships are mostly applicable only to datasets with a limited variance in SOC concentration and on a limited number of soil types.

The biochemical composition of SOC depends on the source material and its decomposition stage. Since the major source for SOC is vegetation, biochemical constituents present in the vegetation can also be present within the soil. Some biochemical constituents are easily dissolvable and decompose fast (e.g. chlorophyll and anthocyanins). Other biochemical constituents are more resistant to decomposition and therefore remain in the soil longer (e.g. lignin and cellulose).

Absorption features in vegetation are summarized by Curran (1989). Several vegetation analysis indices related to lignin and cellulose have been published. The Cellulose Absorption Index (CAI) (Daughtry, 2001; Daughtry et al., 2004) uses the cellulose absorption feature around 2100 nm. The Normalized Difference Lignin Index (NDLI) (Fourty et al., 1996; Melillo et al., 1982; Serrano et al., 2002) makes use of the fact that reflectance at 1754 nm is influenced by lignin concentration of leaves, as well as by the overall foliage biomass of the canopy.

In this paper we show that the amount of SOC can be related to the cellulose and lignin influenced wavelength ranges and detected with reflectance spectroscopy. The dataset contains samples from nine soil types and represents a large variance in SOC content. For comparison, the relationship between SOC and the overall reflectance in the visible wavelengths will be discussed as well.

2. METHODOLOGY

A selection of 40 soil samples, originating from a) ISRIC's World Soil Reference Collection, b) the LOSOC-AHS2005 field campaign and c) the MIES2004 field campaign was used for analysis. The samples represent nine soil types and cover a wide spatial range and a large variance in SOC content (0.06 to 45.1 %). All soil samples were sieved over a 200 µm sieve, keeping the smaller fraction, and analysed for SOC content using the Walkley Black method. The samples were divided into two sets, one for calibration and the other for validation. Both sets contain 20 samples, which represent all nine soil types and practically the entire range of used SOC concentrations. An overview of used samples, the different soil types and the SOC content is given in Table 1.

* Corresponding author: Tel: +31 317 474 578, Fax: +31 317 419 000, E-Mail: Harm.Bartholomeus@wur.nl

	Soil Unit	SOC (%)	Code	Setup
Set 1	Andosol[1]	0.1	CR930063	ASD + Lamp
	Andosol[1]	1.6	CR930061	ASD + Lamp
	Andosol[1]	6.4	CR930060	ASD + Lamp
	Cambisol[1]	0.2	NI930010	ASD + Lamp
	Cambisol[1]	1.0	NI930008	ASD + Lamp
	Chernozem[1]	0.2	HU9700178	ASD + Lamp
	Chernozem[1]	1.4	HU9700175	ASD + Lamp
	Fluvisol[1]	0.2	NI930006	ASD + Lamp
	Sand on Peat[2]	4.0	RING47	ASD + CP
	Sand on Peat[2]	10.7	RING05	ASD + CP
	Sand on Peat[2]	16.0	RING15	ASD + CP
	Histosol[1]	22.9	RING27	ASD + CP
	Histosol[1]	45.1	RING02	ASD + CP
	Sandy Loam[2]	0.9	G1_08	ASD + CP
	Sandy Loam[2]	1.5	P2_01	ASD + CP
	Sandy Loam[2]	2.2	P2-15	ASD + CP
	Luvisol[1]	0.8	NI930020	ASD + Lamp
	Luvisol[1]	3.0	NI930444	ASD + Lamp
	Suelo Neguev[2]	0.9	CR930065	ASD + Lamp
	Suelo Neguev[2]	4.0	CR930067	ASD + Lamp
	Soil Unit	**SOC (%)**	**Code**	**Setup**
Set 2	Andosol[1]	0.2	CR930059	ASD + Lamp
	Andosol[1]	1.8	CR930057	ASD + Lamp
	Andosol[1]	4.0	CR930056	ASD + Lamp
	Cambisol[1]	0.6	NI930009	ASD + Lamp
	Cambisol[1]	2.9	NI930007	ASD + Lamp
	Chernozem[1]	0.3	HU9700177	ASD + Lamp
	Chernozem[1]	2.1	HU9700174	ASD + Lamp
	Fluvisol[1]	0.3	NI930003	ASD + Lamp
	Sand on Peat[2]	4.0	RING50	ASD + CP
	Sand on Peat[2]	10.7	RING49	ASD + CP
	Sand on Peat[2]	16.1	RING28	ASD + CP
	Histosol[1]	23.8	RING43	ASD + CP
	Histosol[1]	40.2	RING01	ASD + CP
	Sandy Loam[2]	1.0	G1_06	ASD + CP
	Sandy Loam[2]	1.5	P3_08	ASD + CP
	Sandy Loam[2]	2.2	P5_08	ASD + CP
	Luvisol[1]	1.1	NI930014	ASD + Lamp
	Luvisol[1]	2.2	NI930445	ASD + Lamp
	Suelo Neguev[2]	0.7	CR930069	ASD + Lamp
	Suelo Neguev[2]	3.8	CR930064	ASD + Lamp

[1] Soil Unit according to FAO classification
[2] Soil Unit according to Local Classification

Table 1: Overview of selected soil samples: Soil unit = type of soil according to FAO or Local classification; SOC(%) = Soil Organic Carbon Content in %; Code = the label of the sample in the dataset; Setup = measurement method with either the ASD and lamp or the ASD with the Contact Probe (ASD + CP).

Spectral measurements were performed using air-dried samples, with an ASD Fieldspec Pro FR in laboratory setup. Six soils (Table 1) were measured using an ASD lamp for illumination at a distance of 55 cm in nadir position and viewing angle of 30° of nadir, with a 1 degree aperture angle, at a distance of 40 cm. The setup results in a Ground Projected Field of View of 0.38 cm^2. The three other soils were measured with an ASD Contact Probe, which results in a FOV of 3.14 cm^2.

First, the direct relation between reflectance values and SOC content was investigated. This was performed for each wavelength individually, the summed reflectance in the visible wavelength range (400-700 nm) and the slope for several ranges in the VIS. Following this, the usability of spectral indices related to lignin and cellulose absorption for SOC prediction, was investigated. For the lignin related absorption, the reflectance around 1754 nm was used. For the cellulose related absorption, the reflectance around 2100 nm was studied. First, the general reflectance pattern was investigated. Next, based on the reflectance pattern, spectral indices were developed. The first type of index uses the area of the absorption feature, which was defined as the sum of the total reflectance minus the continuum removed function. This was done for several wavelength ranges and widths of the absorption feature. The second type of index uses the slope of the spectral signature corresponding to the higher wavelength part of the absorption feature.

Both linear and curvilinear regression functions were plotted. The quality of the fit is assessed using the R^2-value with a confidence level of 0.95. Furthermore the results were cross-validated by applying the found relations for Set 1 on Set 2 and vice versa. Quality of the SOC content prediction is expressed in Standard Error of Calibration (SEC) and Standard Error of Performance (SEP).

3. RESULTS AND DISCUSSION

For individual wavelengths, the highest correlations appear in the visible part of the spectrum. Due to the large variance in SOC content, the relation is not linear. Curvilinear relations are more often reported in literature (Baumgardner et al., 1985), (Schreier, 1977). The relation can be made practically linear by calculating the inverse of the reflectance (1/reflectance), which reveals highest correlation between 640 and 690 nm ($R^2_{set\ 1}$ = 0.8).

Figure 2A shows that the relation between the 'Summed Reflectance 400-700 nm' and SOC, can be considered linear in the SOC range from 0 to 5 % and from 5 % and higher. A combination of samples with high and low SOC content results in a non-linear relation. Again, the relation can be made linear by taking '1/Summed Reflectance 400-700 nm', as shown in Figure 2B.

Not only is there an overall decrease in reflectance in the visible part of the spectrum, but the slope in this spectral region also varies with the SOC content (Figure 3). The correlation of SOC and the slope in the VIS was tested for several spectral ranges, but '1/Slope 400 - 600 nm' appeared to correlate best ($R^2_{set\ 1}$ = 0.93).

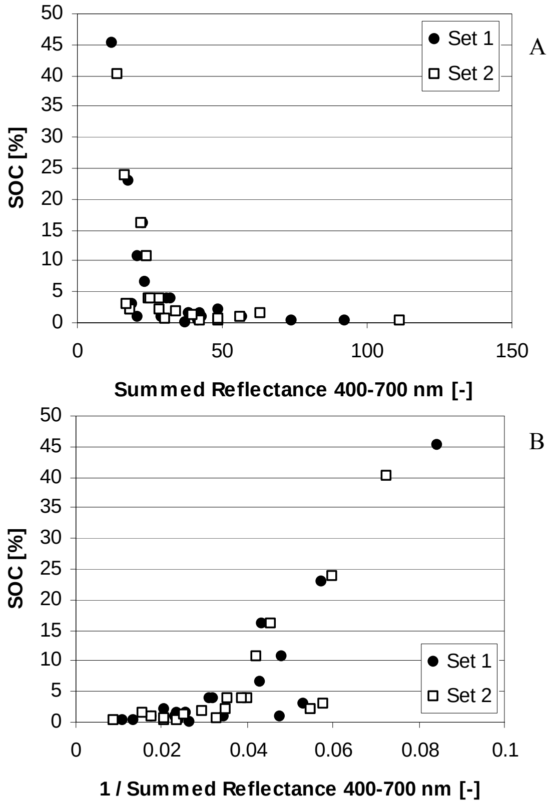

Figure 2: SOC content plotted against the sum of the reflectance in the visible wavelengths (graph A) and the SOC content plotted against '1/the sum of the reflectance' in the visible wavelengths (graph B).

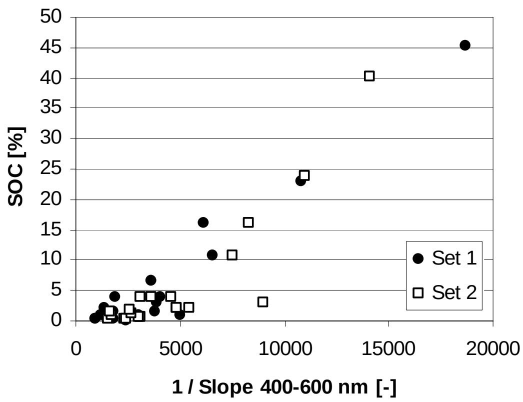

Figure 3: SOC content plotted against '1/Slope 400-600 nm'.

The spectral response from 1600 to 1800 nm for three different SOC contents is shown in Figure 4. Although the overall reflectance varies, there is no difference in reflectance pattern. Therefore, it is not possible to derive an index based on this spectral region, which in vegetation studies is used as an indication for lignin content. A slight variance in slope is visible in Figure 4. However, this does not show a relationship with SOC content.

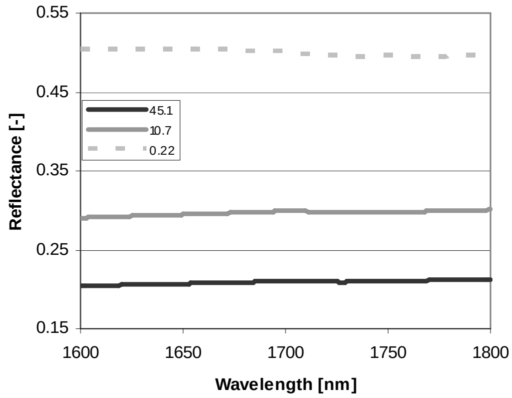

Figure 4: Spectral signature (1600 – 1800 nm) of three soils with different SOC content. This spectral range is influenced by the presence of lignin.

Figure 5 shows the spectral response from 2000 to 2200 nm for three different SOC contents. Besides the difference in overall reflectance, the reflectance pattern also varies. In contrast to vegetation, where the presence of cellulose results in a clear absorption feature, the presence of cellulose does not lead to an absorption dip in this case. Nonetheless, the spectral profile flattens when the SOC content increases. This can be interpreted as an absorption feature, although the reflectance does not show a dip as such.

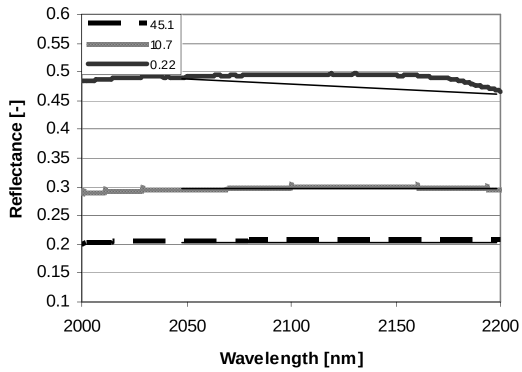

Figure 5: Spectral signature (2000 - 2200 nm) of three soils with different SOC content. This spectral region is influenced by the presence of cellulose.

The changes of the spectral profile from 2000 to 2200 nm are quantified in two ways. First, the total decrease in reflectance, compared to the continuum removed values, between 2050 and 2200 nm, is calculated and related to SOC. This results in a value that describes the area of the absorption feature. Since there is no actual absorption dip, the area shows a negative curvilinear relation with SOC (Figure 6A). More SOC leads to a higher amount of cellulose, which results in more absorption and a flatter spectral signature. This is shown by a lower value for the area. To create a linear relation, the inverse of the area is taken (Figure 6B) ($R^2_{set\ 2} = 0.87$). As a second index, the slope of the spectral signature is calculated, which yielded the best results for '1/Slope 2138-2209 nm' ($R^2_{set\ 2} = 0.98$). This corresponds most with protein and nitrogen related absorption features in vegetation, as reported by (Curran, 1989).

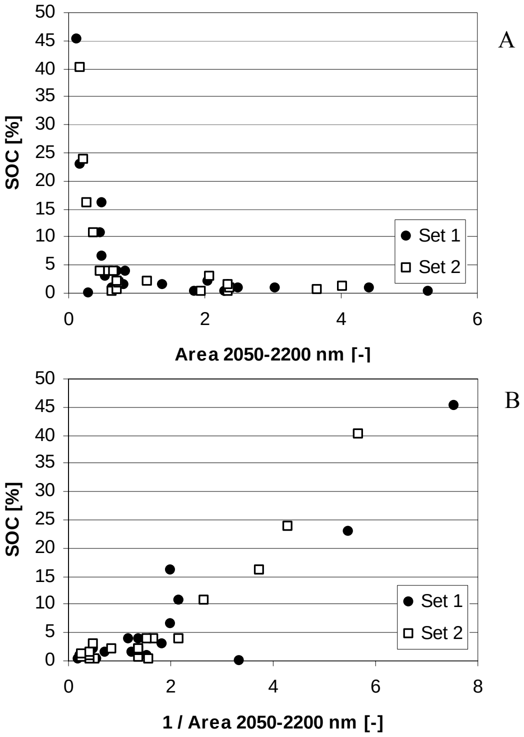

Figure 6: SOC content plotted against the Area from 2050-2200 nm (A) and the SOM content plotted against '1/Area 2050-2200 nm' (B).

The SEC and SEP values (Table 7) show a pattern comparable to the R^2-values. In general SEC and SEP values are highest for '1/Summed Reflectance 400-700 nm'. Mostly, SEP values are somewhat higher than SEC values. The cellulose absorption feature based indices show a larger SEP value for Set 1. Especially Set 1 of '1/Slope 2138-2209 nm' shows a very high SEP value, which can be reduced by removing point 'Ring02'. This sample has a SOC content that lies outside the SOC range of Set 2. Therefore, calculating the SEP for Set 1 implies an extrapolation outside the range of the training set. However it is impossible to draw firm conclusions based on a single observation, '1/Slope 2138-2209 nm' appears to be very sensitive to extrapolation of the found relation beyond the SOC range for calibratoin. Other indices do not show this strong sensitivity. The results show that the SWIR spectral region can be used for the determination of SOC, as was also reported by Morra et al. (1991) and Henderson et al. (1992).

4. CONCLUSION

It is possible to use spectral indices to estimate SOC for a dataset composed of nine soil types. All investigated indices show a curvilinear relationship with SOC, due to the large range in SOC content. Indices based on the presence of cellulose yield higher R^2-values than indices based on the visible part of the spectrum, but SEC and SEP values are not necessarily lower. Some indices show a significant sensitivity to estimation of the SOC content outside the SOC range of the training set. Further research will therefore also rely on broader ranges of soil samples containing even larger variation of SOC.

	R^2-value		Skewedness		SEC		SEP	
	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2	Set 1	Set 2
1/Summed Reflectance 400-700 nm	0.70	0.48	1.22	0.38	5.85	7.07	6.24 / 4.56[a]	7.41
1/Slope 400-600 nm	0.93	0.79	2.57	1.41	2.85	4.51	3.70 / 3.73[a]	5.05
1/Area 2050-2200 nm	0.81	0.87	2.14	1.54	4.67	3.56	5.14 / 5.27[a]	3.93
1/Slope 2138-2209 nm	0.87	0.98	-3.96	-2.57	3.81	1.22	15.14 / 3.84[a]	6.01
			Skewedness		SD		SD	
SOC-content	-	-	2.82	2.61	10.92	10.05	10.92	10.05

[a] For Set 1 the SEP was calculated two times. The first value represents the SEP for the entire set of 20 samples. Because this implies an extrapolation outside the SOC range of Set 2, the SEP was calculated also for the interpolated range of SOC content only. The second number represents this non-extrapolated SEP value

Table 7: Performance of spectral indices towards the prediction of SOC. R^2-values are based on linear relations.

The R^2-values of the correlation functions are described in Table 7. Because relations based on direct correlations are curvilinear, only the results of inversed relations are shown. The largest R^2-values are found for '1/slope 2138-2209 nm'. Nevertheless, this result is somewhat deceptive. Because the SOC content of the used samples already showed a skewed distribution, a further increase in skewedness is undesired. In this case, all samples with a low SOC content concentrate around 0, while the single soil sample with the largest SOC content shows an extremely large negative value. This results in a R^2-value of 0.98, but the relation is not usable for the prediction of SOC, since it cannot differentiate between smaller SOC contents. The cellulose absorption dip based index '1/Area 2050-2200 nm' and the visible wavelengths based index '1/Slope 400-600 nm', show the overall highest R^2-values, combined with a decrease in skewedness.

5. LITERATURE

Baumgardner, M.F., Silva, L.F., Biehl, L.L. and Stoner, E.R., 1985. Reflectance properties of soils. *Adv. Agron.*, 38, pp. 1-44.

Chen, F., Kissel, D.E., West, L.T. and Adkins, W., 2000. Field-Scale Mapping of Surface Soil Organic Carbon Using Remotely Sensed Imagery. *Soil Science Society of America Journal*, 64, pp. 746-753.

Curran, P.J., 1989. Remote Sensing of Foliar Chemistry. *Remote Sensing of Environment*, 30, pp. 271-278.

Daughtry, C.S.T., 2001. Discriminating Crop Residues from Soil by Short-Wave Infrared Reflectance. *Agronomy Journal*, 93, pp. 125-131.

Daughtry, C.S.T., Hunt Jr., E.R. and McMurtry III, J.E., 2004. Assessing Crop Residue Cover Using Shortwave Infrared Reflectance. *Remote Sensing of Environment*, 90, pp. 126-134.

Fourty, L., Baret, F., Jacquemoud, S., Schmuck, G. and Verdebout, J., 1996. Leaf Optical Properties with Explicit Description of Its Biochemical Composition: Direct and Inverse Problems. *Remote Sensing of Environment*, 56, pp. 104-117.

Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D.E. and Coster, D.C., 1992. High Dimensional Reflectance Analysis of Soil Organic Matter *Soil Science Society of America Journal*, 56, 865-872.

Irons, J.R., Weismiller, R.A. and Petersen, G.W., 1989. Soil Reflectance. In: G. Asrar (Editor), Theory and Applications of Remote Sensing. John Wiley & Sons, Washington D.C., pp. 66-106.

Melillo, J.M., Aber, J.D. and Muratore, J.F., 1982. Nitrogen and Lignin Control of Hardwood Leaf Litter Decomposition Dynamics. *Ecology*, 63, pp. 621-626.

Morra, M.J., Hall, M.H. and Freeborn, L.L., 1991. Carbon and Nitrogen Analysis of Soil Fractions Using near-Infrared Reflectance Spectroscopy. *Soil Science Society of America Journal*, 55, 1, pp. 288-291.

Schreier, H., 1977. Quantitative predictions of chemical soil coniditions from multispectral airborne, ground and laboratory measurements. *Canadian Symposium on Remote Sensing, 4th*, pp. 106-112.

Serrano, L., Penuelas, J. and Ustin, S.L., 2002. Remote Sensing of Nitrogen and Lignin in Mediterranean Vegetation from AVIRIS Data: Decomposing Biochemical from Structural Signals. *Remote Sensing of Environment*, 81, pp. 355-364.