This is the peer reviewed version of the following article: [Fernández, S. et al. Land Degradation & Developmen, 27(5), p. 1479–1488 (2016)], which has been published in final form at [http://dx.doi.org/10.1002/ldr.2452]. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
SPATIAL MODELLING OF ORGANIC CARBON IN BURNED MOUNTAIN SOILS USING HYPERSPECTRAL IMAGES, FIELD DATASETS, AND NIR SPECTROSCOPY (CANTABRIAN RANGE; NW SPAIN)

Susana Fernández¹,²*, Juanjo Peón²,³, Carmen Recondo²,³, Javier F. Calleja⁴, César Guerrero⁵

¹Department of Geology, University of Oviedo, C/ Jesús Arias de Velasco s/n, 33005 Oviedo, Asturias, Spain

²Institute of Natural Resources and Spatial Planning (INDUROT), University of Oviedo, C/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain

³Area of Cartographic, Geodesic and Photogrammetric Engineering, Department of Mining Exploitation and Prospecting, University of Oviedo, C/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain

⁴Department of Physics, Polytechnic School of Mieres, University of Oviedo, C/ Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Asturias, Spain

⁵Departament of Agrochemistry and Environment, University Miguel Hernández, Avda. Universidad s/n, 03202 Elche, Alicante, Spain

Corresponding author:
Susana Fernández. Department of Geology, University of Oviedo, C/ Jesús Arias de Velasco s/n, 33005 Oviedo, Asturias, Spain. Tel: +34 985103180. Fax: +34 985458110
E-mail: fernandezmsusana@uniovi.es
Abstract

Soil organic matter is seriously affected by fires and suffers changes in stock, composition, and distribution. In the North-West side of the Cantabrian Range (northern Spain) fires are very common. In order to develop a cartographic technique to map areas with high carbon stocks caused by fire, we test a technique based on calibrated VIS-NIR soil organic carbon models and hyperspectral images. Total (TOC) and oxidizable carbon (OC) were measured in 89 soil samples. The samples were scanned with VIS-NIR spectrometer (400-2500 nm) and the spectra were resampled to the hyperspectral image channels. Spectroscopic models for TOC and OC were fitted ($R^2 > 0.81$) using partial least squares regression (PLSR). The predictions were regionalized to the hyperspectral image and the results validated with a new soil population consisting of twelve Valeri plots collected in burned slopes of the study area under heather vegetation. In soil samples, TOC and OC values are highly correlated ($R = 0.92$) and the coefficients of the PLSR models have a similar pattern, which suggests similar organic components. Nevertheless, there are significant differences in the values of the regression coefficients, much higher in the TOC model except at 560 and 2054 nm that might be interpreted as labile carbon components, and at 1590 nm. At this wavelength the coefficient of TOC is positive and OC is negative, and it could be interpreted as hydrocarbons components present in the TOC model.

**Keywords:** Total Organic Carbon, Oxidizable Organic Carbon, Soil Organic Carbon Mapping, VIS-NIR Spectroscopy, Hyperspectral Remote Sensing
INTRODUCTION

Soil organic carbon (SOC) is an essential component of organic matter (SOM) that plays a key role in soil chemical and physical properties, affecting bulk density, nutrient availability, structural stability, hydraulic conductivity, and soil biodiversity. SOC is the largest terrestrial carbon pool with an estimated content of 1600 Pg for the first meter of soil depth (Novara et al., 2011). Soil organic carbon pools are essential to climate control through the regulation of C and N fluxes and support primary production of the ecosystems (Brevik et al., 2015). In an undisturbed ecosystem such as natural forest, soil carbon is stable over time (steady-state) and soil carbon mineralization is balanced by organic matter production (Novara et al., 2011), and most of the carbon is stored as partially decomposed organic matter (Schulze et al., 2000). The remaining non-mineralized carbon undergoes slow oxidation processes and is stabilized as humic substances. Changes suffered by organic matter are the result of microbial activity (Guénon et al., 2013; Wang et al., 2015), but also can be originated by abiotic transformations induced by external factors, among which fire is a major one (Hatcher & Spiker, 1988). Some authors have reported an increase of the SOM content in areas affected by fire (Czimczik & Masiello, 2007; Fernández-Menéndez et al., 2011; Johnson & Curtis, 2001; Knicker, 2007; Santín et al., 2015; Santin et al., 2008). It is widely accepted that fire causes alterations in physical, chemical and biological soil properties (Mataix-Solera & Guerrero, 2007; Neary et al., 1999; Úbeda & Outeiro, 2009). In soil organic matter fires lead not only to the formation of new thermally-condensed products, but also to the transformation of labile compounds into recalcitrant organic forms (Gonzalez-Perez et al., 2004), all of them contributing to long-term C and N sequestration in soils (Kuhlbusch & Crutzen, 1995; Schmidt & Noack, 2000). In this way, Hajdas et al. (2007) found biochar from Pinus sp. in the B-C soil horizons boundary of an Australian soil dated 11·180±75 and 11·270–10·980 BP. Therefore, burned soils could represent a carbon sink of great importance and with potential impact on the global carbon cycle (Kuhlbusch, 1998).

Annually 0.6 M ha of forest ecosystems are affected by fires in Europe (FAO, 2001). In the western part of the Cantabrian Mountain Range, northern Spain, the use of fire for millennia has modified the vegetation cover (Díaz González & Fernández Prieto, 1994) and promoted intense soil transformations. Former research in the area concluded that fires influence the rates of SOC sequestration, carbon stocks of 32 Mg/ha and 90 Mg/ha were measured in unburned and burned soils, respectively (Fernández-Menéndez et al., 2011).

Fires alter SOC (Czimczik & Masiello, 2007; Gonzalez-Perez et al., 2004; Knicker, 2007; Santin et al., 2008) but the resulting SOM quantity and distribution in the landscape depend not only on losses by combustion and inputs of charred materials (Knicker et al., 2005) but also on post-fire erosion and sedimentation processes and landscape redistribution (Rumpel et al., 2006). In these processes the presence of ash layer plays an important role because it provides substantial SOM protection after fire episode. In this way, Cerdà & Doerr (2008) obtained that a 34 mm ash cover alone can reduce surface runoff from small plots by more than one, and sediment yield by more than two orders magnitude for a 1-hour, 55 mm h⁻¹ rainstorm. Nevertheless in steep slopes the ash spatial variability increased with time as a result of water erosion (Pereira et al., 2013). Knowledge of soil carbon distribution in landscape using conventional analytical methodologies is further complicated by its high spatial variability (Ellert et
al., 2001). Development of methods for soil carbon analysis that address and minimize the uncertainties associated with conventional methodologies are important for improving estimates of terrestrial carbon inventories and fluxes (Burres et al., 2001). In this way, remote sensing offers the possibility of spatial and temporal estimates of land cover and land management which impact soil carbon dynamics (Gehl & Rice, 2007). Nevertheless, research reporting the direct quantification of SOC or SOM content in the upper layer of soils using remote sensing has been limited to agricultural or bare soil areas (Ben-Dor et al., 2002; Gomez et al., 2012; Patzold et al., 2008; Selige et al., 2006; Uno et al., 2005). Relationships between visible reflectance, soil colour and organic carbon are the most common direct way to predict the organic carbon content in soils using satellite images from optic sensors (Chen et al., 2000; Steinhardt & Franzmeier, 1979). But when soils are partially covered by vegetation, as it is the case of burned slopes covered by heather, the spectral signature in each pixel of image corresponds to mixed signal from soil and vegetation.

We understand that separating the vegetation and soil contributions to the signal detected by the sensor is a key question for building valid soil carbon models using satellite imagery with optical sensors. However, in optical images, a complete removal of the vegetation signal is not possible. Nevertheless, the use of soil lab optical spectra (400-2500 nm) together with hyperspectral images can help to minimize the vegetation signal. This is possible because, unlike multispectral imagery, hyperspectral remote sensing provides a quasi-continuous spectrum for each pixel, which enables the spectral identification of minerals, rocks, or soils with similar laboratory reflectance spectral properties at the remote sensing scale (Chabrollat et al., 2002).

On one hand, reflectance spectroscopy laboratory analysis provides a rapid and non-destructive method for soil carbon measurement based on diffusely reflected radiation of an illuminated soil sample (Gehl & Rice, 2007; Guénon et al., 2013). Diffuse reflectance spectroscopy in the visible and near infrared region (VIS-NIR, 400-2500 nm) has been widely applied to quantify soil carbon successfully, often with reported R² values greater than 0.80 (Ben-Dor & Banin, 1995; Chang & Laird, 2002; Chang et al., 2001; Janik et al., 1998; McCarty et al., 2002; Morra et al., 1991). Calibration methods are used to relate lab-spectra to soil analyses, including multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR) (Viscarra Rossel & Behrens, 2010; Viscarra Rossel et al., 2006). Of special interest are recent works in which SOM was partitioned into fractions with different time of residence in soils. In this way, Vasques et al. (2009) obtained labile and recalcitrant soil organic carbon forms and used PLSR techniques to construct accurate models to predict these fractions. On the other hand, the application of satellite hyperspectral imagery to measure soil properties has rarely been reported (Croft et al., 2012; Gomez et al., 2008; Lu et al., 2013), and only data from the Hyperion sensor (400-2500 nm) on board the Earth Observing-1 (EO-1) satellite have been used. Several studies used data from airborne hyperspectral sensor HyMap (400-2500 nm) to generate PLSR models for SOC: Hbirkou et al. (2012), Patzold et al. (2008), and Selige et al. (2006) in agricultural fields in Germany (R² of 0.83, 0.74, and 0.90 respectively). Data from the airborne sensor AHS (430-2540 nm) were also used to calibrate SOC models (Stevens et al., 2010; Stevens et al., 2008). These models had limited predictive ability, but the results showed the potential of imaging spectroscopy as an alternative to conventional analytical techniques. A review of several studies dealing with the
estimation of SOC from airborne and space-borne sensors and with challenges in achieving accurate predictions is given by Croft et al. (2012).

The objectives of this research were: i) test the use of hyperspectral images and VIS-NIR lab-spectra of soils to map oxidizable and total organic carbon in hard to reach areas partially covered by vegetation and affected by fires, and ii) by comparing OC and TOC PLSR models approaching the molecular composition of both soil organic matter components in terms of residence times.

MATERIALS AND METHODS

Location and geological setting

The study area is located in the Cantabrian Mountains (NW of Spain), in the western part of the mountain range which runs parallel to the northern coast of Spain (Figure 1). The climate is oceanic, with mean annual rainfall of 1500 mm, and moderate temperatures, with mean annual values ranging between 6.5 and 10.7°C. The orography corresponds to low-middle mountain areas, with altitudes between 200 and 1200 m and average slopes between 10º and 30º. In this part of the Cantabrian Range an area of approximately 60 km² running SE-NW (Figure 1) was selected to conduct a flight campaign to obtain hyperspectral images. The geological and geomorphological characteristics of this area, together with the large history of forest fires, make it highly suitable for this study. The bedrock is mainly formed by sandstone, quartzite, and slates, giving rise to a high degree of lithological homogeneity. A large part of this substratum is covered by a coarse regolith exhibiting a grain-supported fabric with continuous soil cover. The soils developed over these bedrocks are Histosols and Regosols (WRB, 2014) with poorly evolved profile, sandy and stony, as well as very rich in organic matter (Fernandez et al., 2005) (Figure 2).

Forest fires and soils

Historically, livestock farming has given rise to the repeated burning of extensive areas, which has caused severe degradation of soil and natural vegetation. At present, the majority of the slopes are deforested, alternating gorse-briar scrubland with stony areas where the soil has completely disappeared. The dominant vegetation is oceanic heaths formations (Erica australis subsp. aragonensis (Willk.) Cout. and Calluna vulgaris (L) Hull), a result of the degradation suffered by the natural vegetation caused by forest fires (Díaz González & Fernández Prieto, 1994). As for the natural forest formations, at present, these have been reduced to different types of oak groves (Quercus petraea subsp. Petraea (Matt.) Liebl., Quercus robur L., and Quercus pyrenaica Willd.) (Figure 2).

During the first half of the 20th century forest fires occurred with extreme frequency in the north of the Iberian Peninsula, and in recent years fires are still very frequent in the area. Recent studies of the area and neighbouring locations show alterations of properties of the burnt soils, like significant loss in clay contents and textural and structural degradation which affect the hydrologic behaviour of burned soils, increased infiltration rates (0.0107 to 0.1070 cm/s, Fernandez et al., 2005), and enhanced soil erosion, with average values of soil losses, measured with 137Cs on burned slopes around 6.5 t ha⁻¹ year⁻¹ (Duarte et al., 2008). Fires also affected soil organic matter. Santin et al. (2008) measured TOC concentration in the surface layer (0-5 cm depth) ranging from 390.2 mg g⁻¹ for soil in the burned slopes covered by heaths to 157.1 mg
g$^{-1}$ for soil in the non-burned forest. In the subsurface layer (5-10 cm depth), TOC accumulation is significantly lower, with values ranging from 289.5 mg g$^{-1}$ for soil in the burned slopes to 70.7 mg g$^{-1}$ in non-burned soils.

*Field soil sampling*

Soil sampling was carried out in the summer of 2013. Two types of samples were collected and its geographical position was recorded by a Global Positioning System (GPS) device. 89 surface soil samples from the upper 5 cm were collected randomly from three areas located in the south, north, and in the middle of the study area trying to cut the spatial gradient associated to the abundance of heaths, which are clearly higher in the north of the study area (Figure 2). All the samples were taken in the heathery slopes because this vegetation cover comes from the fire management of the territory (Díaz González & Fernández Prieto, 1994). In addition, for validation and in order to ameliorate the image pixel size effect, 12 Valeri plots (Rossello & Baret, 2007) with 15×15 meters were used to estimate the full spectral variability on soils. Twelve samples were collected by mixing 5 samples gathered at the corners and the centre of each plot.

*Soil carbon analysis*

The soil samples were dried at 70°C, sieved at 2 mm, homogenized and divided into two parts: the first for chemical analysis and the second for spectral analysis in the laboratory. The total C (TOC) concentrations were determined, in duplicate, by means of dry combustion at 950°C (Elemental Vario EL CN-Analyzer). Oxidizable Carbon (OC) was analysed with dichromate oxidation method (Walkley & Black, 1934).

*Laboratory spectroscopy*

The diffuse reflectance spectra of the sieved soil samples were obtained using the VIS-NIR spectrometer LabSpec® 2600 (Analytical Spectral Devices Inc., USA). Samples were scanned using a Hi-Brite Contact Probe with built-in light source and a spot size of 10 mm. A Spectralon® panel was used as white reference. Ten measurements per sample were collected and then averaged in order to obtain a representative spectrum of each soil sample.

*Imaging spectroscopy*

A flight campaign was conducted over the study area by the National Institute for Aerospace Technology (INTA). Airborne hyperspectral data were collected at 11:05 UTC on 10th October 2011 at an altitude of 2987 m above the sea level (9800 ft), using the AHS sensor. This sensor has 80 channels: 63 in the reflective part of the spectrum (visible and near infrared, VIS-NIR), 7 in the 3177-5251 nm region (mid infrared, MIR) and 10 in the 8310-12952 nm range (thermal infrared, TIR). Reflective channels were the only ones used in this study: 20 channels between 443 and 1001 nm with Full Width at Half Maximum (FWHM) ranging from 27 to 29 nm, 1 channel centred at 1590 nm with FWHM of 85 nm, and 42 between 1924 and 2553 nm with FWHM of 14-18 nm. It is worth mentioning that in remote sensing the 1300-2500 nm region is usually referred as short wave infrared (SWIR), while in spectroscopy it is considered as NIR. This last notation was used in this article. At sensor radiance data were corrected by INTA using Atmospheric & Topographic Correction (ATCOR-4) to
obtain Hemispherical Directional Reflectance Factor (HDRF). Then data were georeferenced to the coordinate system Universal Transverse Mercator (UTM), datum World Geodetic System 1984 (WGS-84), zone 30 north; with a spatial resolution of 5 m. Channels of the 1924-2553 nm region that were affected by noise, those with Signal-to-Noise Ratio (SNR) lower than 10, were removed, 24 in total. One channel of the VIS-NIR region was also removed due to an incorrect atmospheric correction. A total of 38 reflective channels were considered for further analysis.

Spectra preprocessing and modelling of soil properties

Lab-spectra of the 89 soil samples were shortened from 350-2500 nm to 400-2450 nm in order to remove noise at the beginning and end of each spectrum. Then they were spectrally resampled to match the response of the AHS sensor. After the resampling process, the 2051 channels of the lab-spectra were resampled to 38 new channels with the spectral resolution of the AHS sensor. Several preprocessing techniques were tested in the spectra before modelling and the best ones were identified by maximizing the ratio of performance to deviation (RPD). Reflectance spectra were converted to log(1/reflectance), since the prediction ability of the models improved with this transformation.

The technique used for modelling TOC and OC was the PLSR method (Wold et al., 2001), a common multivariate statistical technique used in chemometrics, as well as in the quantitative spectral analysis of soils. It allows to construct predictive models when there are many predictor variables that are highly collinear. By fitting a PLSR model, one hopes to find a few PLSR factors that explain most of the variation in both predictors and responses (Haaland & Thomas, 1988; Viscarra Rossel et al., 2006). Two types of PLSR models were calibrated for each soil property using VIS-NIR data, one with the lab-spectra between 400-2450 nm (2051 channels) and another with the resampled spectra to the AHS sensor (38 channels) (Figure 3). The number of factors to include in each model was selected by leave-one-out cross-validation. The optimal number of factors was chosen by minimizing the Akaike Information Criterion (AIC), in order to avoid under- and over-fitting (Akaike, 1969). The lower the AIC is, the better the model is. The regression coefficients of the PLSR models (b) were plotted to identify the most important spectral regions for the prediction of each soil property.

In order to obtain spatially continuous maps of both soil properties, the PLSR models calibrated with the lab-spectra resampled to the AHS were applied pixel by pixel to the AHS image, after the geometric and atmospheric correction. Digital maps of both soil properties were obtained with a spatial resolution of 5 m (Figure 3).

Validation and scaling methods

The quality of the PLSR models was evaluated by means of several parameters obtained in the leave-one-out cross-validation: the coefficient of determination ($R^2_{cv}$), the root mean square error (RMSE$_{cv}$), and the ratio of performance to deviation (RPD). RPD is the ratio of the standard deviation of the reference data to the RMSE of cross-validation. Chang et al. (2001) defined three categories of models based on the value of RPD: models in category I (RPD>2) are considered to accurately predict the soil property, models in category II (2>RPD>1.4) belong to an intermediate class and have a limited predictive power, and models in category III (RPD<1.4) have no prediction ability.

On the other hand, the predictive power of the TOC and CO maps was evaluated using the independent validation dataset consisting of 12 Valeri plots. A mean filter
with kernel size 3×3 was applied to the maps of soil properties in order to obtain a representative value of the 15×15 m Valeri plot, and then compared to the values of reference. The parameters used to evaluate the difference between the reference and the predicted values by the maps were the $R^2$, residual standard error (RSE), root mean square error of prediction (RMSEP), and Bias for the mean of the residuals. The results of the validation were satisfactory in terms of $R^2$ and RSE, but not in terms of RMSEP and Bias. The overestimation of each map was corrected by means of a linear fit, adjusted using the estimated and reference values of two randomly chosen plots. Finally, the remaining 10 plots were used for an independent validation of the scaled maps.

**RESULTS AND DISCUSSION**

*Soil samples descriptive analysis*

The results of the chemical analysis obtained in the lab for the 89 soil samples (individual samples) and for the 12 Valeri plots are summarized in Table I. A strong positive correlation was observed between both soil properties, TOC and OC, with a Pearson correlation coefficient ($R$) of 0.92 in the individual soil samples and 0.97 in the plot samples. The range of both TOC and OC are high but in accordance with the results obtained by Fernandez et al. (2005) and Santin et al. (2008) in the same area. The maximum values of 63.2% (TOC) or 37.9% (OC) could have a geomorphological origin related to the accumulation of extensive mantles of organic soil in remaining paleo surfaces, located mainly in the north of the study area. On the other hand, the lowest values of TOC and OC (4.3% and 2.7%, respectively) are related to the erosive degradation suffered by burned slopes in which the soil parent material corresponds mainly to coarse regolithic with grain-supported fabric. Frequent fires in these slopes caused very intense soil degradation and as a result these soils have extremely thin organic horizons (Duarte et al., 2008; Fernandez et al., 2005).

*PLSR models of soil organic carbon*

Several preprocessing techniques were tested in the spectra before the PLSR modelling. The best preprocessing for both soil properties was the transformation from reflectance (R) to log(1/R), and was applied to all the spectra. These results are in accordance with Vasques et al. (2009) and Viscarra Rossel et al. (2006), who also obtained the best models for TOC prediction using this transformation. PLSR settings and model performance statistics are summarized in Table II, and the PLSR regression coefficients obtained using the resampled spectra are plotted in Figure 4. Accurate PLSR models were obtained for both soil properties and spectral resolution (entire lab-spectra or lab-spectra resampled to AHS), with $R^2_{cv}$ values between 0.78 and 0.89, and RPD ranging from 2.13 to 2.97. According to the model classification based on RPD by Chang et al. (2001), models provide accurate predictions. The best predictions were obtained for OC, with RMSE$_{cv}$ of 3.5% and an RPD of ~3; and slightly worse for TOC, with RMSE$_{cv}$ of ~6.5% and an RPD of ~2.2. The prediction ability of the models was similar to that obtained in other works in which soil organic carbon was also modelled using lab-spectra and PLSR: Chang & Laird (2002) ($R^2 = 0.89$), Martin et al. (2002) ($R^2 = 0.75$, RPD = 1.97), and McCarty et al. (2002) ($R^2 = 0.82$).

Despite the high contrast in the number of channels of both spectral resolutions in Table II, no significant differences were observed between models using the entire
lab-spectra (2051 channels) or the lab-spectra resampled to the AHS sensor (38 channels). Slightly better results were even obtained using the resampled spectra. These results show the high potential of hyperspectral remote sensing for organic carbon mapping, and are in accordance with the results obtained by Gomez et al. (2008), who found that the spectral resolution did not significantly change the accuracy of the PLSR models when the lab-spectra were resampled to the hyperspectral sensor Hyperion, regardless of the organic carbon ranges and the number of samples. Lu et al. (2013) evaluated the performance of PLSR models for SOC prediction and other soil properties using three spectral settings: lab, Hyperion image, and lab resampled to Hyperion. Their model for SOC based on resampled spectra was slightly less accurate than that based on the entire lab-spectra, but still provided a moderately satisfactory result ($R^2 = 0.74, RPD = 1.98$).

In general, between 1000 and 2500 nm there are several absorption peaks in both models corresponding to overtones and combinations of fundamental vibrations of C-O, O-H, C=O, C-H, and N-H bonds that are present in organic compounds such as proteins, starch, cellulose, humic acids, and lignin (Ben-Dor et al., 1997; Huang et al., 2008). These absorptions are however difficult to attribute to a single component since they are greatly overlapping in the VIS-NIR. The coefficients of the PLSR models obtained with the resampled spectra are plotted in Figure 4 as a function of wavelength, and in Table III the positive PLSR values and the corresponding wavelengths are given. In general, peaks in TOC model are higher than in OC model probably as a consequence of a greater amount of organic components in TOC model. This can be related to the chemical extraction methods (oxidization and dry combustion) more effective in TOC extraction (Tessier et al., 1979). Nevertheless, at 560 nm and 2054 nm there are absorption peaks that show higher value in OC than in TOC model. These peaks correspond to the most labile carbon fractions in Vasques et al. (2009) interpretation and to extractive lignine and carbohydrates in Huang et al. (2008) interpretation. These results seem to highlight the more labile nature of the OC model components. The rest of coefficients are higher in TOC than in OC model showing a greater amount of stable soil carbon components. A negative regression coefficient was obtained at 1590 nm in the OC model. This wavelength is significant in this work because it corresponds to a wideband (85 nm) bringing together the NIR bands information and moreover it was the only NIR band in the range between 1000 and 2000 nm that has been used in the models. The rest of the bands in this interval of the hyperspectral image had to be removed due to noise problems. According to the literature, this band could be associated with crystalline cellulose and lignine (Huang et al., 2008; Viscarra Rossel, 2007) and also could be related to hydrocarbons components (~1647, ~1712 nm) present only in TOC model and associated to fire organic carbon transformations (Okparanma & Mouazen, 2013) but it is not easy to interpret in our results. At 2181, 2196, 2212, and 2227 nm in both models there is an intense negative peak (Figure 4) which could be related to biochar components or intensely mineralized organic components (Bellon-Maurel & McBratney, 2011; Vasques et al., 2009).

Maps of soil organic carbon

PLSR models for TOC and OC were applied to the AHS image in order to obtain maps of both properties in the upper 5 cm of soils. The independent validation dataset consisting of 12 Valeri plots was used to evaluate the accuracy of these maps. The best results of the analysis between the reference and the predicted values were obtained for
OC, with $R^2$ of 0.74 and RSE of 4.5%; and slightly worse for TOC, with $R^2$ of 0.74 and RSE of 6.0%. However, the analysis of the RMSEP and Bias showed that maps tend to strongly overestimate both soil properties, with RMSEP values higher than 40% and Bias of about -40%, probably due to the signal from the vegetation in the spectra from the image, not considered in lab calibrations based on soil spectra. Maps of soil properties were scaled using a linear fit based on data from two plots, and the performance of the maps was evaluated again using the remaining 10 plots (Table IV). The scatter plot of observed against estimated soil properties after the scaling is shown in Figure 5. As expected, very similar $R^2$ and RSE values were obtained after the scaling, and a significant improvement was observed in the RMSEP and Bias. The validation of the scaled maps was satisfactory for both soil properties, TOC ($R^2 = 0.73$, RMSEP = 7.8%) and OC ($R^2 = 0.72$, RMSEP = 5.1%), with an accuracy only slightly lower than that obtained in the lab (see Table II). For illustration purposes, the spatial distribution of the predicted TOC and OC content in a slope located in the south of the study area is shown in Figure 6. Maps in Figure 6 were smoothed using a mean filter of kernel size 3x3 in order to improve the outcomes, and to clearly show the variation in the soil properties within the slope.

Several studies used data from the AHS sensor to obtain spatially continuous maps of soil organic carbon, but all of them in bare soils. Model for SOC prediction in agricultural fields by Stevens et al. (2008) only explained 52.7% of the SOC variation, whereas Stevens et al. (2010) obtained better results in croplands ($R^2 = 0.75$), but still with limited accuracy (RMSEP of $\sim 3 \text{ g C kg}^{-1}$). Data from the airborne sensor HyMap were also used for SOC prediction in agricultural fields and slightly better results were obtained: Hbirkou et al. (2012) ($R^2 = 0.83$), Patzold et al. (2008) ($R^2 = 0.74$), and Selige et al. (2006) ($R^2 = 0.90$). All these works, and also the one presented here, are limited to the upper layer of soils (0-5 cm), since all of them are based on remotely sensed data collected in the optical region of the spectrum (400-2500 nm), which capacity to penetrate in soil is very limited and restricted to the surface layer. In this work, topsoil organic carbon was modelled in a complex domain far from the bare soil ideal conditions, since soils were partially covered by vegetation. Nevertheless, maps of TOC and OC achieved an acceptable accuracy (Table IV), similar or slightly lower than that obtained in other works for bare soils, and only slightly lower than lab predictions.

**CONCLUSION**

i) The use of PLSR techniques in VIS-NIR spectra of soils is a suitable method to make accurate models to estimate soil organic carbon in the upper 5 cm of soils. The opacity of vegetation cover to optical sensors is a major problem to monitor the organic matter of soils with satellite imagery in wild landscapes, but it can be solved using these techniques applied to mixed pixels which are partially covered by vegetation. It is important to highlight that the accuracy of models is not influenced by the resolution of the spectra, making it possible to resample the spectra to wavelengths in the range of the hyperspectral images.

ii) By comparing the coefficients of absorption bands of OC and TOC PLSR models, we can explore the main molecular components of both organic carbon fractions. The results seem to highlight the more labile nature of the OC model components. The coefficients related to not-labile forms are higher in TOC than in OC model showing a greater amount of stable soil carbon components.
ACKNOWLEDGEMENTS

This work was funded by the Government of Asturias through the project SV-PA-13-ECOEMP-40. J. Peón acknowledges a PhD Grant from the University of Oviedo and a PhD Grant “Severo Ochoa” from the Government of Asturias (BP14-104). We also acknowledge INTA for providing the images and for the geometric and atmospheric correction.

REFERENCES


Tessier A, Campbell PGC, Bisson M. 1979. Sequential extraction procedure for the speciation of particulate trace-metals. *Analytical Chemistry* **51**: 844-851. DOI: 10.1021/ac50043a017.


### Tables

#### Table I. Descriptive statistics of the soil samples

<table>
<thead>
<tr>
<th>Property</th>
<th>n</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOC (%)</td>
<td>89</td>
<td>4·3</td>
<td>63·2</td>
<td>29·4</td>
<td>28·3</td>
<td>14·9</td>
<td>50·7</td>
</tr>
<tr>
<td>OC (%)</td>
<td>12</td>
<td>4·5</td>
<td>37·9</td>
<td>25·7</td>
<td>28·8</td>
<td>11·2</td>
<td>43·6</td>
</tr>
</tbody>
</table>

#### Table II. PLSR settings and performance statistics for prediction models of TOC and OC, based on VIS-NIR spectra

<table>
<thead>
<tr>
<th>Property</th>
<th>n</th>
<th>Factors</th>
<th>AIC</th>
<th>R²cv</th>
<th>RMSEcv (%)</th>
<th>RPD</th>
<th>Factors</th>
<th>AIC</th>
<th>R²cv</th>
<th>RMSEcv (%)</th>
<th>RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOC (%)</td>
<td>89</td>
<td>7</td>
<td>187·3</td>
<td>0·78</td>
<td>7·01</td>
<td>2·13</td>
<td>7</td>
<td>181·2</td>
<td>0·81</td>
<td>6·55</td>
<td>2·28</td>
</tr>
<tr>
<td>OC (%)</td>
<td>12</td>
<td>9</td>
<td>130·2</td>
<td>0·88</td>
<td>3·53</td>
<td>2·93</td>
<td>9</td>
<td>124·8</td>
<td>0·89</td>
<td>3·47</td>
<td>2·97</td>
</tr>
</tbody>
</table>

Reflectance spectra were transformed to log(1/R). n: number of samples; Factors: number of PLSR factors selected by minimizing the Akaike Information Criterion (AIC); R²cv: coefficient of determination of cross-validation; RMSEcv: root mean square error of cross-validation; RPD: ratio of performance to deviation.

#### Table III. PLSR coefficients of models for TOC and OC prediction using resampled spectra, and molecular interpretations

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>TOC</th>
<th>OC</th>
<th>Difference</th>
<th>Molecular interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>443</td>
<td>173·0</td>
<td>91·4</td>
<td>81·6</td>
<td>Lignine</td>
</tr>
<tr>
<td>560</td>
<td>62·9</td>
<td>65·0</td>
<td>2·1</td>
<td>Lignine and extractives</td>
</tr>
<tr>
<td>591</td>
<td>180·8</td>
<td>138·4</td>
<td>42·4</td>
<td>Labile carbon forms</td>
</tr>
<tr>
<td>620</td>
<td>51·4</td>
<td>47·4</td>
<td>4·0</td>
<td></td>
</tr>
<tr>
<td>885</td>
<td>3·9</td>
<td>2·4</td>
<td>1·5</td>
<td></td>
</tr>
<tr>
<td>914</td>
<td>42·7</td>
<td>27·8</td>
<td>14·9</td>
<td>Organic pigments</td>
</tr>
<tr>
<td>943</td>
<td>88·8</td>
<td>61·6</td>
<td>27·2</td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>135·1</td>
<td>67·1</td>
<td>68·0</td>
<td>Hemicellulose hydroxyls</td>
</tr>
<tr>
<td>1590</td>
<td>26·5</td>
<td>28·0</td>
<td>54·5</td>
<td>Cellulose</td>
</tr>
<tr>
<td>2054</td>
<td>25·8</td>
<td>29·5</td>
<td>3·7</td>
<td>Carbohydrates</td>
</tr>
<tr>
<td>2071</td>
<td>53·7</td>
<td>45·0</td>
<td>8·7</td>
<td>Labile carbon forms</td>
</tr>
<tr>
<td>2086</td>
<td>71·3</td>
<td>52·5</td>
<td>18·8</td>
<td>Cellulose</td>
</tr>
<tr>
<td>2102</td>
<td>88·9</td>
<td>36·4</td>
<td>32·5</td>
<td>Lignine</td>
</tr>
<tr>
<td>2118</td>
<td>103·1</td>
<td>57·3</td>
<td>45·8</td>
<td>Recalcitrant carbon</td>
</tr>
<tr>
<td>2134</td>
<td>112·8</td>
<td>57·6</td>
<td>55·2</td>
<td>Lignine</td>
</tr>
<tr>
<td>2150</td>
<td>85·1</td>
<td>40·5</td>
<td>44·6</td>
<td>Recalcitrant carbon</td>
</tr>
<tr>
<td>2165</td>
<td>1·7</td>
<td>1·7</td>
<td>0·0</td>
<td></td>
</tr>
<tr>
<td>2266</td>
<td>111·8</td>
<td>76·2</td>
<td>35·6</td>
<td>Recalcitrant carbon</td>
</tr>
<tr>
<td>2296</td>
<td>175·0</td>
<td>108·7</td>
<td>66·3</td>
<td>Recalcitrant carbon</td>
</tr>
<tr>
<td>2326</td>
<td>94·9</td>
<td>58·9</td>
<td>36·0</td>
<td>Lignine</td>
</tr>
</tbody>
</table>

#### Table IV. Validation of maps of soil properties with an independent dataset

<table>
<thead>
<tr>
<th>Parameter</th>
<th>n</th>
<th>R²</th>
<th>RSE (%)</th>
<th>a</th>
<th>b</th>
<th>RMSEP (%)</th>
<th>Bias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOC (%)</td>
<td>10</td>
<td>0·73</td>
<td>6·65</td>
<td>1·5 ± 0·3</td>
<td>-11 ± 8</td>
<td>7·80</td>
<td>1·40</td>
</tr>
<tr>
<td>OC (%)</td>
<td>10</td>
<td>0·72</td>
<td>4·98</td>
<td>1·1 ± 0·2</td>
<td>0 ± 4</td>
<td>5·05</td>
<td>0·65</td>
</tr>
</tbody>
</table>

n: number of samples; R²: coefficient of determination; RSE: residual standard error; Observed = a × Estimated + b; RMSEP: root mean square error of prediction; Bias: mean residual.
Figures

Figure 1. Location of the study area (NW Spain).

Figure 2. General view of study area and organic-enriched profiles. In vegetation map, purple represents heaths affected by frequent fires. Pale green represents typical vegetation (chasmophytic) covering skeletal and stony soils originated mainly by post-fire erosion. Forest is represented in vivid green and at least grasslands for livestock use are represented in yellow.
Figure 3. Processing flow chart.

Figure 4. Regression coefficients (b) of PLSR models for TOC and OC prediction using the resampled spectra. Dots represent the 38 AHS channels used in this work.

Figure 5. Scatter plot of observed against estimated soil properties using scaled maps.
Figure 6. AHS true color composite (bands at 650, 560, and 443 nm in RGB) (a), and spatial distribution of the predicted TOC (b) and OC content (c) in a slope partially covered by heath, located in the south of the study area.