Soil organic matter and texture estimation from VIS-NIR-SWIR spectra in areas of land cover changes using correlated component regression (CCR)

Olga A. Rosero-Vlasova[1,2*], Lidia Vlassova[3]

Fernando Pérez-Cabello[1,2], Raquel Montorio[1,2], Estela Nadal-Romero[2,4],

[1]*GEOFOREST Group, Environmental Sciences Institute (IUCA), University of Zaragoza, Spain*

[2]*Department of Geography and Land Management, University of Zaragoza, Spain*

[3]*Department of Environmental Sciences, Technical State University of Quevedo, Ecuador*

[4]*Pyrenean Institute of Ecology, Geo-environmental Processes and Global Change, IPE-CSIC, Zaragoza, Spain*

Corresponding author:

Olga Alexandra Rosero Vlasova

Address:

Department of Geography and Land Management

Faculty of Arts, University of Zaragoza

Pedro Cerbuna, 12, E-50009, Zaragoza

Tel: +34-876554058

E-mail:

*oarosero@unizar.es

Running Title:

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ldr.3250

SOM and texture estimation from VIS-NIR-SWIR spectra using CCR

Abstract

Land use changes due to natural and human-related factors, which include wildfires and crop abandonment, are among the most important drivers of soil degradation and demand regular monitoring. Proximal soil sensing in VIS-NIR-SWIR spectral regions could offer a solution. However, to become operational optimal combination of data and technique have to be defined. Thus, the purpose of this study was (i) to predict the soil organic matter (SOM) content and soil texture in areas of wildfire burns and crop abandonment in Aragón province, northern Spain, from their laboratory reflectance spectra using novel correlated components regression with a step-down variable selection algorithm (CCR-SD) and (ii) to compare the CCR-SD and the PLSR methods. The results obtained by the tested methods were similar. CCR-SD models showed high predictive capacity with coefficients of determination (R^2) in the range of 0.80–0.86 and 0.70–0.87 for calibration and validation datasets, respectively, and the highest R^2 value was attained in the SOM estimation. Moreover, the CCR-SD models stand out for the superior accuracy-parsimony relationship: the number of predictors varied from 16 (silt models) to 49 (SOM models).On average, the CCR-SD calibrations needed less than a half of the predictors employed in PLSR models. This research confirmed that CCR-SD can be used for monitoring SOM content and texture of soils from VIS-NIR-SWIR spectra in the study area and, probably, in other areas of land use/land cover change and that CCR-SD can create highly parsimonious models that achieve results comparable with the commonly used PLSR method.

Keywords: soil organic matter, soil texture, land cover change, VIS-NIR-SWIR spectroscopy, Correlated Components Regression

1. INTRODUCTION

Deterioration of soil conditions closely related to land use and land cover (LULC) changes is a threat to human well-being (Lal & Stewart, 2010; Pimentel, 2006). The scope and environmental effects of LULC changes caused by natural and human factors vary in space and time (García-Ruiz, 2010; Lu, Li, Valladares, & Batistella, 2004). In Mediterranean ecosystems, regular wildfires, which can trigger soil erosion (Cerdà & Robichaud, 2009), have contributed to landscape formation for several centuries (Pausas, Llovet, Rodrigo, & Vallejo, 2009), while the spread of cropland abandonment in the region is an example of a more recent LULC phenomenon linked to the soil conditions (Nadal-Romero, Cammeraat, Pérez-Cardiel, & Lasanta, 2016).

The urgent need for action to reduce soil degradation is recognized in several sustainable development goals (SDGs) formulated in the United Nations 2030 Agenda for Sustainable Development (United Nations, 2015). SDG 15.3 is specifically dedicated to land degradation and mentions the restoration of degraded land and soil and the achievement of a land degradation-neutral world. To evaluate the movement towards SDGs it is necessary to establish, register and compare relevant soil characteristics (Montanarella & Panagos, 2018). Moreover, in spite of the seriousness of the problem, our knowledge on soil degradation and the scope and effects of mitigation strategies are incomplete (Assessment, 2005). Thus, monitoring the soil status on a regular basis is imperative (Tóth, Hermann, da Silva, & Montanarella, 2018), especially in areas of natural and anthropogenic disturbances, such as Mediterranean (Merino, Moreno, Navarro, & Gallardo, 2016). Examples of successful systems for monitoring soil variables (including SOC and texture) at a regional level exist in

Europe (soil monitoring network in Slovakia) (Kobza, 2015) and Australia (New South Wales Monitoring, Evaluation and Reporting (MER) Program) (Chapman et al., 2011).

The soil status can be evaluated through a set of soil attributes/indicators. According to a published review (Bünemann et al., 2018), a minimum set of soil attributes/indicators should include chemical, physical and biological indicators. A large number of studies reviewed by Bünemann et al. (2018) include soil organic matter (SOM) and texture among the most important indicators of soil quality. SOM, which is one of the main sources of soil carbon and plant nutrients, determines soil fertility and plays an important role in both water cycle (infiltration and runoff) and quality (Tóth et al., 2018). On the other hand, land productivity is directly impacted by soil erosion (Troeh & Thompson, 2005), with texture being one of the basic indicators of soil erodibility (Goldman, Bursztynsky, & Jackson, 1986) and other hydraulic properties (Tóth et al., 2018).

Conventional methods for the estimation of soil properties require important investments of time and effort, which motivate the search for alternatives. Spectral sensing methods, such as VIS-NIR-SWIR spectroscopy, may be one of the time and cost-effective solutions (Demattê et al., 2016). Based on results of previous research, which demonstrated that soil characteristics correlate with their spectral signatures (Demattê & da Silva Terra, 2014; Stevens, Nocita, Tóth, Montanarella, & van Wesemael, 2013), this technique uses electromagnetic spectra in visible (VIS), near infrared (NIR) and shortwave infrared (SWIR) spectral regions to estimate soil properties. Because soil spectra are obtained by sensors located near (< 2 m) the soil surface, this method is sometimes referred to as proximal sensing.

Spectroradiometers produce more accurate results than satellite and airborne instruments because of the high-resolution spectra (contain > 2000 of narrow (up to 1 nm) bands) obtained in controlled environmental conditions (Ben-Dor & Demattê, 2016). Soil VIS-NIR-SWIR spectra have been successfully applied to estimate soil carbon, SOM and texture (Conforti, Matteucci, & Buttafuoco, 2018; Lugassi, Ben-Dor, & Eshel, 2014; Mouazen, Karoui, De Baerdemaeker, & Ramon, 2005). High-resolution spectra of soil surface samples obtained under laboratory conditions serve as a standard in spectral unmixing of remote sensing images acquired by sensors on airborne and satellite platforms and are used for digital mapping of soils and other environmental variables (Demattê et al., 2016; Ben-Dor & Demattê, 2016).

Since soils are mixtures of organic and inorganic particles with highly variable proportions of each substance and particle size, their spectra present overlaps of spectral features corresponding to specific soil constituents (Ben-Dor & Demattê, 2016). Soil variables are not directly calculated from the spectra; instead, they are related to a set of known reference samples representative of the soil variation in the study area through the development of multivariate statistical models. Models calibrated for a certain area are not usually transferable to another area (Grunwald, Thompson, & Boettinger, 2011).

Extracting information from a large number of highly correlated spectral bands is a challenging task. There is a wide range of statistical tools available for multivariate modelling of soil properties. Ongoing research is continuously evaluating new tools at the same time striving to clarify the viability of application of VIS-NIR-SWIR soil spectroscopy in specific scenarios (e.g., Gholizadeh, Saberioon, Carmon, Boruvka, & Ben-Dor, 2018; Ogen, Neumann, Chabrillat, Goldshleger, & Ben-Dor, 2018; Ostovari et al., 2018; Terra, Demattê, & Viscarra Rossel, 2018; Viscarra Rossel & Brus, 2018).

Methods based on linear models, especially partial least squares regression (PLSR), are among the most popular (Mouazen, Kuang, De Baerdemaeker, & Ramon, 2010; Vasques, Demattê, Viscarra Rossel, Ramírez-López, & Terra, 2014; Viscarra Rossel, McGlynn, & McBratney, 2006), although latest comparative studies report on successful applications of data mining techniques and tools, such as artificial neural networks (ANN) (Mouazen et al., 2010), support vector machines (SVM) (Viscarra Rossel & Behrens, 2010) and memory-based learning (MBL) (Gholizadeh, Borůvka, Saberioon, & Vašát, 2016).

The frequent choice of PLSR is explained by its capacity to produce well-fit models from datasets containing a small number of observations characterized by a great number of correlated predictors. Robustness of the models is mainly achieved through reduction of data dimensionality using a set of orthogonal vectors (components) (Wold, Sjöström, & Eriksson, 2001). Still, PLSR models sometimes demonstrate unrealistically high fit due to inclusion of noise variables relevant only for calibration dataset, which is known as overfitting (Babyak, 2004; Esbensen, Guyot, Westa, & Houmoller, 2002).

A recently introduced method of Correlated Components Regression (CCR) is trying to avoid overfitting problem in a different way. It prevents model overfit through application of the regularization process, which involves identification of suppressors and elimination of less relevant predictors (Magidson, 2013). Since development, CCR has been successfully applied to very diverse research areas, such as socio-demography (Alkerwi, Vernier, Sauvageot, Crichton, & Elias, 2015), medicine (Ruiz-Rodado et al., 2014) and logistics research (Garver & Williams, 2018). However, CCR was not used in soil spectroscopic modelling until now, even though the characteristics of CCR are very attractive.

In this context, the study seeks to contribute to the search and assessment of the methods more adequate for modelling soil properties from spectral data in specific scenarios and answer the following research questions (i) is it possible to predict SOM content and texture fractions of soils from wildfire burns and cropland abandonment in Mediterranean environment from VIS-NIR-SWIR spectra using correlated components regression (CCR)? and (ii) what are the advantages of using CCR in simultaneous modelling SOM and soil texture compared to two versions of PLSR? This allows formulating the corresponding research hypotheses: (i) CCR is an adequate tool for monitoring SOM and texture of soils in areas of LULC changes, and (ii) CCR offers several advantages in simultaneous modelling soil texture and SOM compared to PLSR.

2. METHODOLOGY

2.1. Study area and soil sampling

The study area of approximately 310 km^2 (Figure 1) is located in the Aragón region, northern Spain (42°10'-42°37'N, 0°16'-1°17'W), and contains sites affected by (i) wildfire burns and (ii) cropland abandonment. The area of uneven topography (elevations between 450 m and 1300 m) is characterized by a Mediterranean climate with a mean annual temperature of approximately 10°C and a precipitation range of 600-800 mm (Cuadrat & Martín-Vide, 2007).

The mosaic of vegetation covers in the study area is composed of plant communities dominated by *Quercus gr. Cerrioides* (Willk and Costa) and *Quercus ilex* L. and pine forests of *Pinus sylvestris* L., *Pinus nigra* (Arnold), *Pinus halepensis* L., and *Pinus pinaster* (Aiton) interspersed with shrublands dominated by *Buxus sempervirens* L. and *Genista scorpius* L. (Ruiz de la Torre, 1990). In areas affected by wildfires, typical soils that formed on calcaric materials have coarse and medium textures and are classified as Cambisols, and there are some patches of Regosols and Leptosols (Badía-Villas & del Moral, 2016). On the other

hand, in areas previously used for agriculture, thin soils with silt loam texture are classified as Leptic Calcaric Regosols (FAO, 2014).

A total of 113 soil samples were collected from the surface soil layer (0-10 cm) during the 2013 and 2014 field campaigns. Approximately two-thirds of the samples (82) were from wildfire burns that occurred during 1975-2009. The sample site locations were determined by the spatial pattern of the burned areas throughout that period. Wildfire perimeters were identified using databases of the Aragón Government (Service for Management of Wildfires and Coordination, Head Office for Forest Management), as well as mapping products produced in the context of the research project "Forest fires and predictive models of ecologic vulnerability to fire: restoration management activities and application of climate change scenarios" GA-LC-042/2011 (Caixa-DGA). Within the wildfire boundaries, the precise location of the samples is a function of accessibility factors, plant-community variability in the context of Aragón and wildfire size. In the absence of pre-fire soil data, a paired-samples approach (Novara, La Mantia, Barbera, & Gristina, 2012) was applied: at each of the 41 selected sites, a pair of samples was obtained–one sample of the burned soil and a reference sample of the same/similar unburned soil. The unburned soils are located in areas near the outer perimeter of the burned lands that have not been affected by fire, which are representative of large areas with similar physical conditions. The SOM content in these samples was estimated with UV-visible spectrophotometry. The soil texture, i.e., the relative proportion of sand, silt and clay (%) in the dry and sieved (< 2 mm) samples, was determined using the standard particle size-distribution analysis (USDA, 1996).

Samples of soils affected by cropland abandonment (31) were collected in the Araguás catchment, where cultivation of terraced fields stopped in the 1950s. Subsequent afforestation with *Pinus nigra* (PN) and *Pinus sylvestris* (PS) occurred a decade later, although some areas underwent a process of natural secondary succession with *Genista scorpius* and *Buxus*

sempervirens. Sampling was carried out at sites with five different landcovers typical for the area (bare soil, permanent pasturelands, secondary succession, afforestation with PS and PN) selected based on the analysis of aerial photography, topographic maps and field survey information. The samples were obtained from five 5 m x 5 m plots with a similar topography established at each site. After collecting five surface (0-10 cm) samples from locations at each of the plot diagonals, they were combined into one sample. For these samples, the loss on ignition method was used to determine SOM; soil texture fractions were determined using a particle analyser (Micromeritics, SediGraph 5100, Nocross, USA). Descriptive statistics characterizing the collected samples are presented in Table 1.

Additional details on the study area, as well as the sampling procedure, are available in (Rosero-Vlasova, Pérez-Cabello, Montorio Llovería, & Vlassova, 2016) (wildfire burns) and (Nadal-Romero et al., 2016) (abandoned croplands).

2.2. Soil spectra

For spectral measurements, the fine soil fraction (particle size < 2 mm) of each sample was placed in a Petri dish (90 mm in diameter) and dried in an oven at 105°C for 24 hours.

Soil spectral curves were obtained using an analytical spectral device (ASD) FieldSpec®4spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA) under controlled laboratory conditions, with a setup that included an ASD Illuminator lamp (Analytical Spectral Devices Inc., Boulder, CO, USA) and a pistol grip (Rosero-Vlasova et al., 2016). Figure 2 shows the general view and details of observations geometry. The soil sample area detected by the optic fibre cable (sensor) is determined by the following geometry: an illuminator lamp (field of view (FOV): $\theta = 12°$) is attached to the tripod in a cenital position at a height of H = 42 cm generating a lighted spot 8.82 cm in diameter (D). The setup also includes a pistol grip attached to another tripod at a height of h = 7.5 cm

($FOV_{Bare\ Fibre}$ $\beta = 25°$, diameter SPOT d = 6.99 cm) and an angle $\alpha = 25°$ relative to the vertical axis (Figure 2a). The spectral response of the white reference (WR) panel was obtained with the same viewing geometry (Figure 2b).

Radiances measured in the VIS-NIR (350-1000 nm) and two SWIR regions (1001-1800 nm and 1801-2500 nm) (Castro-Esau, Sanchez-Azofeifa, & Rivard, 2006) were corrected for the baseline electrical signal (dark current) and converted into reflectance values using a calibrated white Spectralon® panel as a reference. The ASD Illuminator halogen lamp was employed as a light source. Previous research (Rosero-Vlasova et al., 2016) has demonstrated that this experimental setup ensures an optimal observation environment resulting in low-noise spectra.

Radiometric jumps, evident at the wavelengths situated at the joins between the detectors (at 1000 nm and 1800 nm), were corrected using a procedure suggested in (Danner, Locherer, Hank, & Richter, 2015), which compensates the difference between the reflectance using the values of the first detector (VIS range) as a baseline. The following formulas were applied:

$$Corr\_val_{1000} = R_{\lambda=1001} - (2 \cdot R_{\lambda=1000} - R_{\lambda=999}) \quad (1)$$

$$Corr\_val_{1800} = R_{\lambda=1801} - (2 \cdot R_{\lambda=1800} - R_{\lambda=1799}) \quad (2)$$

where R_λ is the reflectance at λ wavelength and $Corr\_val_{1000}$ and $Corr\_val_{1800}$ are correction values at the spectral splitting points, which are added to the original values and, depending on their algebraic sign, either increase or decrease reflectances in all further wavelengths.

Moreover, the noisy bands at the extremes of the spectra (< 400 nm and > 2470 nm) were removed, leaving 2071 bands for statistical modelling.

2.3. Statistical modelling

Modelling of SOM content and texture fractions is based on seventy pre-selected spectral bands (11 bands in VIS, 18 bands in NIR and 48 bands in SWIR spectral regions). The importance of this set of wavelengths for soil property detection was previously reported in multiple studies (Ben-Dor, Heller, & Chudnovsky, 2008; Demattê & da Silva Terra, 2014; Demattê et al., 2016; Melendez-Pastor, Navarro-Pedreño, Gómez, & Koch, 2008; Rosero-Vlasova, Borini Alves, Vlassova, Perez-Cabello, & Montorio Lloveria et al., 2017).

The presence of outliers in scaled and centred datasets was assessed with Principal Components Analysis (PCA). Three data points lying outside the 95% confidence level Hotelling's T^2 ellipse in the score plot representing the loadings of the two first principal components (Figure 3) were excluded as outliers, leaving 110 soil spectra for analysis. These were randomly divided into calibration (~65%) and validation (~35%) sets, containing 80 and 30 samples, respectively. To ensure robust results, this procedure was repeated three times to obtain three sample sets of calibration and validation data (S1, S2 and S3). Descriptive statistics of the data used in model building are presented in Table 2 and Figure 4. Since each dataset contains more than thirty samples, the Kolmogorov-Smirnov test with Lilliefors significance correction was applied to test the SOM, silt, clay and sand distributions for normality; one-way ANOVA was run to detect significant differences between the distributions of the tested variables in S1, S2 and S3. All the tests were performed using IBM® SPSS® Statistics version 20.0.0 (2011) software (https://www.ibm.com/products/spss-statistics).

Statistical models were developed using (i) a routinely applied (Demattê et al, 2016) full-spectrum partial least squares regression (PLSR-full) (Wold et al., 2001) implemented in The Unscrambler X® software (2016) (CAMO Software AS, Norway, 2016), version 10.4 (https://www.camo.com/unscrambler); (ii) PLSR with predictors selected by Martens Uncertainty Test (Martens & Martens, 2000) available in The Unscrambler X® software (2016) version 10.4 (PLSR-MUT), and (iii) a novel technique of correlated components regression with a step-down variable selection algorithm (CCR-SD) (Magidson, 2010; Magidson, 2013) implemented as an XLSTAT Pearson Edition (2014) (Addinsoft S.A., New York, NY, USA, 2014), version 2014.5.03 (http://www.xlstat.com) complement for the Microsoft Office Excel (2010) software. For each sample, all the soil properties of interest in our study (SOM, clay, silt and sand) were predicted simultaneously.

Both CCR and PLSR are capable of dealing with a large number of highly correlated predictors (in this study, the correlation coefficients R are in the range of 0.639-0.999). Multicollinearity of spectral data is approached by means of regularisation (the enforcement of model sparsity), consisting in dimension reduction.

PLSR proceeds by calculating a set of orthogonal components (latent variables) which explain most of the variance in predictors and responses (Wold, 2001). Determination of the optimum number of components and selection of the final model is performed through the leave-one-out cross-validation: the model is developed leaving out one of the samples, which is later substituted into the model to evaluate the adjustment; the process is repeated for each sample and the final model is that showing the best fit.

Since exclusion of the less important (noisy) predictors may improve model accuracy, in the second tested method PLSR models used only most important variables selected by the Martens uncertainty test (PLSR-MUT), which estimates uncertainty of regression coefficients obtained in leave-one-out cross validation (Martens & Martens, 2000).

On the other hand, in CCR-SD data dimension is reduced through (i) calculation of correlated components and (ii) elimination of less relevant predictors from the model with step-down variable selection algorithm, resulting in sparser models (Magidson, 2013).

CCR utilizes $K < P$ correlated components, with each S_K component being an exact linear combination of g predictors ($g = 1,2, \ldots P$). Predictions for Y in the first (primary) component ($\hat{Y}$) directly affect the outcome and are obtained from the simple ordinary least squares (OLS) regression of Y on S_1. Similarly, the second component S_2 is calculated by the simple OLS regression of Y on S_1 and S_2. The calculation of the remaining components follows the same process. Once the models for all the components are obtained, the final model (eq. 3) is computed using the expression:

$$\hat{Y} = \alpha^{(K)} + \sum_{g=1}^{P} \beta_g X_g \quad (3)$$

where α and β are regression coefficients.

Thus, the components are not orthogonal; the second and subsequent components are correlated to the first component and represent the influence of "suppressor" variables (Magidson & Wassmann, 2010). The inclusion of suppressor variables removes the noise of some irrelevant variables included in the first component, improving the model quality.

At the same time, the method controls overfitting through a reduction in the number of predictors, leaving out the less important predictors. Thus, CCR was combined with a step-down variable selection algorithm, which excludes the least important predictors (Bennett, 2013; Magidson, 2010). This is achieved through M-fold cross-validation. Each round (10 rounds in this study) consists of a series of operations. First, the data are randomly divided into M groups (folds) of equal size (5 groups of 80/5 = 16 samples each in our study). Next, samples from four groups are used to build the model, while the samples from the fifth group are used for model validation. The process is run for each group (M times). In the next round,

the process is repeated with newly randomized M groups. Thus, the quality of the final model is assessed on the out-of-sample fit, ensuring replication of the calibration results on real-life data, which has been a long-time concern related to published models (Nuzzo, 2014). Model assessment based on new out-of-sample cases means that modelling with CCR does not pose requirements to satisfy sampling assumptions, which are the basis of traditional hypothesis testing (Curl, Thompson, & Aspinall, 2015).

2.4. Model performance assessment

Model performance was evaluated using the coefficient of determination R^2 (eq. 4), which measures how successful the calibration fit is in explaining the variation in the data, root-mean-square error of calibration (RMSEC) (eq. 5) and root-mean-square error of cross-validation (RMSECV) (eq. 6), which assess the model accuracy.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad (4)$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{n_c}(Y_{i,c} - \hat{Y}_{i,c})^2}{n_c - (f+1)}} \quad (5)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n_c}(Y_{i,c} - \hat{Y}_{i,c})^2}{n_c - 1}} \quad (6)$$

where n is the number of samples, subscript c and p refer to calibration and validation datasets; Y_i is the measured value for sample i, $\hat{Y}_i$ is the predicted value for sample i, $\bar{Y}$ is the mean value, and f is the number of variables used in the regression equation. The predictive ability of the models was also evaluated with the root-mean-square error of prediction (RMSEP) (eq. 7), with bias of validation ($bias_{val}$) and standard error of prediction (SEP_c) being independent components of RMSEP (Stevens et al., 2013, Rosero-Vlasova et al., 2016); and the ratio of performance to interquartile (RPIQ) range (Bellon-Maurel & McBratney, 2011), which was calculated according to eq. 8.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p}\left(Y_{i,p}-\hat{Y}_{i,p}\right)^2}{n_p}} \quad (7)$$

$$RPIQ = \frac{Q_3-Q_1}{RMSEP} \quad (8)$$

RPIQ is based on inter-quartile distances (IQ = Q3-Q1), where Q1 represents the lowest 25% of the samples and Q3 is the value below which 75% of the samples can be found. RPIQ is the ratio of IQ to the RMSE of prediction (RMSEP) and adequately represents populations with skewed distributions and a large number of low values, such as the soil sample sets in this study. Finally, the Akaike information criterion (AIC), computed following eq. 9, was applied to determine the model with the best accuracy-parsimony relationship (Akaike, 1973; Viscarra Rossel & Behrens, 2010).

$$AIC = n \ln RMSEP + 2f \quad (9)$$

where n is the number of samples and f is the number of predictors. The smaller the AIC criterion is, the better the model.

3. RESULTS AND DISCUSSION

3.1. SOM content and texture fractions

Table 2 and Figure 4 present descriptive statistics for the SOM and soil texture fractions (clay, silt and sand) of the used datasets. ANOVA detected no statistically significant differences among S1, S2 and S3 ($p < 0.05$).

The SOM content ranges between ~1 g 100 g^{-1} and ~20 g 100 g^{-1}, and the mean is ~6 g 100 g^{-1}, demonstrating levels characteristic of the study area (Pérez-Cabello, Echeverría, Ibarra, & Riva, 2009). These values are higher than the average values registered in global (Brown, Shepherd, Walsh, Mays, & Reinsch, 2006) and European (Stevens et al., 2013) soil

databases, which is rather surprising given that approximately one-third of the samples is from the burned areas. However, the unusually high SOM content in samples of burned soils may be due to vigorous vegetation development in the burned areas, which may have contributed to accelerated recovery of organic material destroyed by the fire (Jiménez-González et al., 2016; Vlassova & Pérez-Cabello, 2016).

The soil texture fraction values are highly variable, especially for sand (Table 2 and Figure 5): the values change from approximately 5% to close to 65%, which is not strange considering the landscape heterogeneity in the study area. However, samples from different land covers in the cropland abandonment areas show similar textures dominated by silt, confirming the findings by Laudicina et al. (2012), who observed that land use change did not affect soil texture. Thus, the variability in the soil texture of the analysed set is caused by the contribution of soils from wildfire burns. In general, the clay content in sampled soils is higher (mean 27%) and the sand proportion is lower (mean 33%) than the average for European soils in the LUCAS database (Stevens et al., 2013).

3.2. Soil spectra

Spectral curves of the analysed soils are presented in Figure 6 (a, b) (wildfire burns) and Figure 6 (c, d) (abandoned croplands). Their form is typical for soil reflectance spectra: a gradual increase through the visible wave range, an almost flat segment in NIR, and slightly lower reflectance values in SWIR (Ben-Dor, Irons, & Epema, 1999). The small number of absorbance features can be ascribed to the presence of water (1400 nm and 1900 nm) and clay minerals (2200 nm) (Brown, 2007; Brunet, Barthès, Chotte, & Feller, 2007).

The soil spectra differ mainly in reflectance intensity, confirming the results of previous research (Bellon-Maurel, Fernandez-Ahumada, Palagos, Roger, & McBratney, 2010; Chabrillat, Ben-Dor, Viscarra-Rossel, & Demattê, 2013; Demattê, Campos, Alves, Fiorio, & Nanni, 2004; Stenberg, Viscarra Rossel, Mouazen, & Wetterlind, 2010). Thus, the maximum

reflectance values of soils from wildfire burns range from 0.25 to 0.65, while the maximum reflectance values of soil samples from cropland abandonment areas are considerably smaller (0.38-0.55). This finding can be explained by differences in organic matter and texture: high SOM contents and smaller particle size result in spectral curves with lower reflectance (Ben-Dor et al., 2009; Conforti, Froio, Matteucci, & Buttafuoco, 2015; Viscarra Rossel, Walvoort, McBratney, Janik, & Skjemstad, 2006). The high variability in the soil spectra from wildfire burns is caused by the inclusion of undisturbed forest soil samples with organic matter content higher than that of any of the soils from crop abandonment, as well as burned soils, whose organic matter was completely destroyed by fire (Figure 6b).

Shape is another key for the differentiation of soils through visual inspection. Thus, the shape of the bare soil spectrum in Figure 6d is quite different from the rest of the spectra: convex in the 500-600 nm waverange and almost horizontal in the NIR and part of the SWIR spectral regions, which is typical for weathered soils (Demattê, 2002).

3.3. Statistical modelling

The results of simultaneous statistical modelling of SOM, clay, silt and sand from reflectance spectra using the two versions of PLSR (PLSR-full and PLSR-MUT) and CCR-SD methods are presented in Table 3 (model calibration) and Table 4 (model validation).

The values correspond to three datasets (S1, S2 and S3), resulting from different random partitions of available samples in the calibration (80 samples) and validation (30 samples) groups. In each case, the final model was obtained after 100 iterations/rounds.

The optimal number of components is similar for all the models: 9-10 for PLSR-full and 8-10 for other methods (PLSR-MUT and CCR-SD); fewer components used in PLSR-MUT and CCR-SD models for SOM. The number of predictors is more variable. Since standard PLS regression (PLSR-full) does not discard any predictor, these models include the full

range of available bands (2071). In case of PLSR-MUT and CCR-SD models resulting from procedures eliminating less relevant predictors, the number of predictors varies a lot depending on the predicted property and sample dataset. In PLSR-MUT models the number of predictors varies from one dataset to another, but it is the same for all the modelled properties (171, 39 and 197 for S1, S2 and S3, respectively), while in CCR-SD the number of predictors depends not only on the dataset, but also on the modelled property (e.g., there are 49, 28, 19 and 22 predictors in S1 models for SOM, clay, silt and sand, respectively). It is worth to note, that variable selection in PLSR is realized after running the full spectrum model necessary to estimate importance of the variables, while in CCR-SD method estimation of variable relevance and development of the final model are performed simultaneously.

In general, the number of predictors in CCR-SD models is greater for SOM (49, 61, 58 for S1, S2 and S3, respectively), because organic compounds exhibit spectral activity along the whole spectrum, while proportion of specific organic constituents in SOM varies a lot depending, among other factors, on overall SOM concentration and geological heterogeneity of the area (Stenberg et al., 2010).The number of predictors is decreasing in CCR-SD models for silt and sand, although in these models it varies more from one analysed subset to another. The number of variables in clay models developed using the same method varies the least (21, 28 and 29 predictors for S1, S2 and S3, respectively), which is not strange, since it is the only texture fraction directly associated with minerals having detectable spectral features in VIS-NIR-SWIR (Escribano, Schmid, Chabrillat, Rodríguez-Caballero, & García, 2017), often referred to as clay minerals (for example, kaolinite and illite).

Examples of the scatter plots for S2 (modelled versus predicted values) are presented in Figure 7. Among modelled properties, SOM calibrations showed the highest predictive capability accounting on average for 86% (CCR-SD) and 82% (PLSR-full and PLSR-MUT)

of the variance in the calibration, and 87% (PLSR-full and CCR-SD) and 79% (PLSR-MUT) in the validation datasets. The CCR-SD and PLSR-full models for SOM developed in this study are more accurate (R^2 0.05-0.07 higher and the average RPIQ above 3) than the SOM models we built previously for soil samples from wildfire burns from the same area using PLSR with the step-down variable selection algorithm (Rosero-Vlasova, Vlassova, Pérez-Cabello, Montorio, & Nadal-Romero, 2018), which is probably due to the different modelling algorithm and larger calibration dataset used in this study. The superior RPIQ of the SOM models (~3 or above) developed with CCR-SD and PLSR-full is another indicator of their high quality. Similar results were previously reported by researchers working on applications of VIS-NIR-SWIR spectroscopy for soil characterization in other areas of LULC change (Ge, Thomasson, & Sui, 2011; Knadel, Stenberg, Deng, Thomsen, & Greve, 2013).

The CCR-SD models estimating texture fractions also showed good fit, with coefficients of determination in the ranges of 0.84-0.86 (calibration) and 0.68-0.72 (validation) for the silt models and 0.80 (calibration) and 0.70 (validation) for the clay and sand predictions. The coefficients of determination of PLSR silt models were lower and varied considerably among datasets (0.62-0.80). Sand was the only property where CCR-SD models (R^2_C = 0.79; R^2_V = 0.70) were outperformed by those developed with PLSR-full (R^2_C = 0.83; R^2_V = 0.75). The most important difference between methods was observed in clay models, where good fit of CCR-SD models contrasted with considerably lower performance of PLSR calibrations (R^2_C = 0.69; R^2_V = 0.56 for PLSR-full and R^2_C = 0.64; R^2_V = 0.47 for PLSR-MUT versus R^2_C = 0.82; R^2_V = 0.66 for CCR-SD). A better fit of SOM models was previously observed by other researchers (Demattê et al., 2016; Viscarra Rossel et al., 2006). However, it was rather unexpected that CCR-SD models for clay, silt and sand showed similar performance, albeit not as good as that of the SOM models. Usually clay is another successfully modelled

property, but satisfactory fit for silt calibrations is rarely obtained (Pinheiro, Ceddia, Clingensmith, Grunwald, & Vasques, 2017; Stenberg et al., 2010).

On the whole, validation results were similar for all the methods. They compare well to the best achievements in modelling of the same soil characteristics reported in previous research (Conforti et al., 2018; Demattê et al., 2016; Mouazen et al., 2010; Rosero-Vlasova et al., 2017). However, the important difference lies in the structure of the models created by different algorithms evident in the number of predictors in models developed using the three compared methods. The same high-quality of PLSR-full models using as predictors 2071 bands present in the measured reflectance spectra was achieved by CCR-SD using a greatly reduced number of bands/predictors (49, 21, 16, and 22 for SOM, clay, silt and sand, respectively). Compared to the PLSR-MUT models, which also implements variable selection mechanism, the CCR-SD produced better results for all the modelled properties and datasets, except slightly better fit at validation of PLSR sand models for of S2 and S3 datasets (Table 4). The accuracy-parsimony relationship estimated by the AIC shows considerable superiority of all the CCR-SD models in this aspect (average AIC = 122.80, Table 4); the lowest (best) AIC values correspond to the CCR-SD models for clay (AIC = 100.35). Average AIC values for PLSR-full and PLSR-MUT models are 4186.93 and 327.19, respectively.

Figure 8 presents the coefficients for the CCR-SD model predictors and shows the relative importance of specific bands the studied soil properties. Thus, for analysed soil samples spectral regions closely related to SOM in these models (Figure 8a) include the 500-550 nm, 1000-1050 nm, 1500-1550 nm, 1800-1910 nm, 2200-2250 nm and 2310-2350 nm wave ranges and can be attributed to the presence of water and organic molecules with C-O, C=O and N-H bonds (Bellon-Maurel et al., 2010). Although the absorption features characteristic of clay minerals in these soils are masked by the high content of organic matter, the highest

coefficients in the clay model (Figure 8b) correspond to bands related to clay minerals, such as kaolinite (1395 nm, 1414 nm and 2208 nm) and illite (2206 nm, 2300-2340 nm) (Bellon-Maurel et al., 2010; Ben-Dor & Banin, 1995; Brunet et al., 2007). The presence of a considerable quantity of illite in soils from the studied areas of cropland abandonment (Nadal-Romero, Regüés, Martí-Bono, & Serrano-Muela, 2007) supports these findings. In most cases bands from these intervals are also selected as important predictors in PLSR models.

4. CONCLUSIONS

The study confirmed the viability of using CCR-SD algorithm in modelling of organic matter content and texture fractions of soils from VIS-NIR-SWIR spectra for monitoring soil quality in areas recovering from natural (wildfires) and anthropic (agricultural cultivation) disturbances. A novel CCR-SD algorithm created models with good predictive capacities that simultaneously estimated SOM, clay, silt and sand (R^2 in the range of 0.80–0.86 for the calibration dataset and 0.70–0.87 for the validation dataset), with the highest coefficient of determination being achieved by the SOM predictions.

The reliability of the CCR-SD models resulted similar to the PLSR models with full (PLSR-full) and reduced (PLSR-MUT) number of predictors. However, the CCR-SD models achieved good fit using a smaller number of available predictors. One of the advantages of CCR-SD application is the possibility of running calibrations in a familiar interface of EXCEL (Microsoft) software package. Further research is planned to test the methodology on a wider database of soils from erosion-risk environments, such as areas of slash-and-burn agriculture.

Up-to-date information supporting activities protecting soil from degradation will allow the control of short- and long-term consequences of management decisions. The methodological results obtained in this work may provide an interesting operational tool to

analyse soil properties and support sustainable management programmes in forest areas with degradation risks, thematic area explicitly mentioned in SDG 15: “Sustainably manage forests, combat desertification, halt and reverse land degradation, halt biodiversity loss” (2030 Agenda for Sustainable Development).

Acknowledgements

Authors appreciate the financial support provided to the first author of this article by Secretariat for Higher Education, Science, Technology and Innovation (SENESCYT), Ecuador, grant no. 211-2012 in the initial stage of the research, following the funding received in the framework of the “Mobility Program for Latin Americans 2017/2018” of Bank Santander and University of Zaragoza. Soil samples from abandoned croplands were collected during the Marie Curie Intra-European Fellowship project “MED-AFFOREST” (PIEF-GA-2013-624974) and Estela Nadal-Romero was the recipient of a “Ramón y Cajal” postdoctoral contract (RYC-2013-14371, Spanish Ministry of Economy and Competitiveness).

References

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika 60*(2), 255-265. https://doi.org/10.2307/2334537
- Alkerwi, A., Vernier, C., Sauvageot, N., Crichton, G. E., & Elias, M. F. (2015). Demographic and socioeconomic disparity in nutrition: application of a novel Correlated Component Regression approach. *BMJ Open, 5*(5), e006814. http://dx.doi.org/10.1136/bmjopen-2014-006814
- Assessment, M. E. (2005). Ecosystems and Human Well-Being: Biodiversity: Synthesis. In J. Sarukhán, & A. Whyte (Eds.), *A Report of the Millennium Ecosystem Assessment* Washington, DC, USA: Island Press. Retrieved from https://www.millenniumassessment.org/documents/document.356.aspx.pdf
- Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411-421. Retrieved from https://journals.lww.com/psychosomaticmedicine/Fulltext/2004/05000/What_You_See_May_Not_Be_What_You_Get__A_Brief,.21.aspx
- Badía-Villas, D., & del Moral, F. (2016). Soils of the Arid Areas *The Soils of Spain* (pp. 145-161): Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-20541-0_4

- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J. M., & McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry 29*, 1073-1081. https://doi.org/10.1016/j.trac.2010.05.006
- Bellon-Maurel, V., & McBratney, A. (2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils–Critical review and research perspectives. *Soil Biology and Biochemistry, 43*(7), 1398-1410. http://dx.doi.org/10.1016/j.soilbio.2011.02.019
- Ben-Dor, E., & Banin, A. (1995). Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. *Soil Science Society of America Journal, 59*(2), 364-372. https://doi.org/10.2136/sssaj1995.03615995005900020014x
- Ben-Dor, E., Chabrillat, S., Demattê, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L., & Sommer, S. (2009). Using Imaging Spectroscopy to study soil properties. *Remote Sensing of Environment, 113, Supplement 1*(0), S38-S55. https://doi.org/10.1016/j.rse.2008.09.019
- Ben-Dor, E., & Demattê, J. (2016). Chapter 25-Remote sensing of soil in the optical domains. In P. S. Thenkabail (Ed.), *Remote Sensing Handbook Vol. II: Land Resources Monitoring, Modeling, and Mapping with Remote Sensing* (pp. 733-787). Boca Raton, FL, USA: CRC Press (Taylor & Francis Group).
- Ben-Dor, E., Heller, D., & Chudnovsky, A. (2008). A novel method of classifying soil profiles in the field using optical means. *Soil Science Society of America Journal, 72*(4), 1113-1123. https://doi.org/10.2136/sssaj2006.0059
- Ben-Dor, E., Irons, J. R., & Epema, G. F. (1999). Chapter 3-Soil reflectance. In A. N. Rencz, & R. A. Ryerson (Eds.), *Manual of remote sensing*, *Remote Sensing for the Earth Sciences* (3rd ed., Vol. 3) (pp. 111-188). New York, NY, USA: John Wiley & Sons.
- Bennett, G. (2013). Regression modelling with many correlated predictors and few cases. *Presentation in ASSESS Event*. York, United Kingdom. Retrieved from http://www.spssusers.co.uk/Events/2013/Bennett2013.pdf
- Brown, D. J. (2007). Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma, 140*(4), 444-453. https://doi.org/10.1016/j.geoderma.2007.04.021
- Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., & Reinsch, T. G. (2006). Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma, 132*(3), 273-290. https://doi.org/10.1016/j.geoderma.2005.04.025
- Brunet, D., Barthès, B. G., Chotte, J.-L., & Feller, C. (2007). Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma, 139*(1–2), 106-117. https://doi.org/10.1016/j.geoderma.2007.01.007
- Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., . . . Mäder, P. (2018). Soil quality–A critical review. *Soil Biology and Biochemistry, 120*, 105-125. https://doi.org/10.1016/j.soilbio.2018.01.030
- Castro-Esau, K., Sanchez-Azofeifa, G., & Rivard, B. (2006). Comparison of spectral indices obtained using multiple spectroradiometers. *Remote Sensing of Environment, 103*(3), 276-288. https://doi.org/10.1016/j.rse.2005.01.019
- Cerdà, A. (Ed.), & Robichaud, P. (Ed.). (2009). Fire effects on soil infiltration. *Fire Effects on Soils and Restoration Strategies* (pp. 81-103). Boca Raton, FL, USA: CRC Press (Taylor & Francis Group). https://doi.org/10.1201/9781439843338

- Chabrillat, S., Ben-Dor, E., Viscarra-Rossel, R. A. V., & Demattê, J. A. M. (2013). Quantitative soil spectroscopy. *Applied and Environmental Soil Science*. 2013(ID 616578), 1-3. http://dx.doi.org/10.1155/2013/616578
- Chapman, G., Gray, J., Murphy, B., Atkinson, G., Leys, J., Muller, R., . . . McInnes-Clarke, S. (2011). Assessing the condition of soils in NSW. *State of the Catchments 2010: Soil Condition–Technical Report Series*. Office of Environment and Heritage, Sydney. Retrieved from https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Land-and-soil/assessing-condition-soils-nsw.pdf
- Conforti, M., Froio, R., Matteucci, G., & Buttafuoco, G. (2015). Visible and near infrared spectroscopy for predicting texture in forest soil: an application in southern Italy. [Visible and near infrared spectroscopy for predicting texture in forest soil: an application in southern Italy]. *iForest - Biogeosciences and Forestry, 8*(3), 339-347. https://doi.org/10.3832/ifor1221-007
- Conforti, M., Matteucci, G., & Buttafuoco, G. (2018). Using laboratory Vis-NIR spectroscopy for monitoring some forest soil properties. *Journal of Soils and Sediments, 18*(3), 1009-1019. https://doi.org/10.1007/s11368-017-1766-5
- Cuadrat, J. M. (Ed.), & Martín-Vide, J. M. (Ed.). (2007). *Spanish Climatology: Past, Present and Future*. Zaragoza, Ar, Spain: University of Zaragoza Press.
- Curl, A., Thompson, C. W., & Aspinall, P. (2015). The effectiveness of 'shared space'residential street interventions on self-reported activity levels and quality of life for older people. *Landscape and Urban Planning, 139*, 117-125. https://doi.org/10.1016/j.landurbplan.2015.02.019
- Danner, M., Locherer, M., Hank, T., & Richter, K. (2015). Spectral Sampling with the ASD FIELDSPEC 4. In EnMap (Ed.), *EnMAP Field Guides Technical Report* (pp. 20). Potsdam. http://doi.org/10.2312/enmap.2015.008
- Demattê, J. A., Morgan, C. L.S., Chabrillat, S., Rizzo, R., Franceschini, M., Terra, F. D., . . . Wetterlind, J. (2016). Chapter 24-Spectral sensing from ground to space in soil science: State of the art, applications, potential and perspectives. In P. S. Thenkabail (Ed.), *Remote Sensing Handbook Vol. II: Land Resources Monitoring, Modeling, and Mapping with Remote Sensing* (pp. 661-732). Boca Raton, FL, USA: CRC press (Taylor & Francis Group).
- Demattê, J. A. M. (2002). Characterization and discrimination of soils by their reflected electromagnetic energy. *Pesquisa Agropecuária Brasileira, 37*(10), 1445-1458. http://dx.doi.org/10.1590/S0100-204X2002001000013
- Demattê, J. A. M., Campos, R. C., Alves, M. C., Fiorio, P. R., & Nanni, M. R. (2004). Visible–NIR reflectance: a new approach on soil evaluation. *Geoderma, 121*(1–2), 95-112. https://doi.org/10.1016/j.geoderma.2003.09.012
- Demattê, J. A. M., & da Silva Terra, F. (2014). Spectral pedology: A new perspective on evaluation of soils along pedogenetic alterations. *Geoderma, 217–218*(0), 190-200. https://doi.org/10.1016/j.geoderma.2013.11.012
- Esbensen, K. H. (Ed.), Guyot, D. (Ed.), Westad, F. (Ed.), & Houmoller, L. P. (Ed.). (2002). Multivariate data analysis-in practice: An introduction to multivariate data analysis and experimental design. *Multivariate Data Analysis*, (5th ed.) (pp 600). Oslo, Norway: CAMO Process AS.
- Escribano, P., Schmid, T., Chabrillat, S., Rodríguez-Caballero, E., & García, M. (2017). Optical remote sensing for soil mapping and monitoring. In P. Pereira, E. Brevik, M. Muñoz-Rojas, & B.A. Miller (Eds.), *Soil Mapping and Process Modeling for Sustainable Land Use Management* (pp. 87-125): Elsevier. https://doi.org/10.1016/B978-0-12-805200-6.00004-9

- FAO. (2015). World Reference Base for soil resources. Retrieved from Rome, Italy: http://www.fao.org/3/i3794en/I3794en.pdf
- García-Ruiz, J. M. (2010). The effects of land uses on soil erosion in Spain: a review. *Catena, 81*(1), 1-11. http://dx.doi.org/10.1016/j.catena.2010.01.001
- Garver, M. S., & Williams, Z. (2018). Improving the validity of theory testing in logistics research using correlated components regression. *International Journal of Logistics Research and Applications, 21*(4), 363-377. https://doi.org/10.1080/13675567.2017.1401054
- Ge, Y., Thomasson, J. A., & Sui, R. (2011). Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science, 5*(3), 229-238. https://doi.org/10.1007/s11707-011-0175-0
- Gholizadeh, A., Borůvka, L., Saberioon, M., & Vašát, R. (2016). A Memory-Based Learning Approach as Compared to Other Data Mining Algorithms for the Prediction of Soil Texture Using Diffuse Reflectance Spectra. *Remote Sensing, 8*(4), 341. https://doi.org/10.3390/rs8040341
- Gholizadeh, A., Saberioon, M., Carmon, N., Boruvka, L., & Ben-Dor, E. (2018). Examining the Performance of PARACUDA-II Data-Mining Engine versus Selected Techniques to Model Soil Carbon from Reflectance Spectra. *Remote Sensing, 10*(8), 1172. https://doi.org/10.3390/rs10081172
- Goldman, S. J., Bursztynsky, T. A., & Jackson, K. (1986). *Erosion and Sediment Control Handbook*. New York [etc.]: McGraw-Hill.
- Grunwald, S., Thompson, J., & Boettinger, J. (2011). Digital soil mapping and modeling at continental scales: Finding solutions for global issues. *Soil Science Society of America Journal, 75*(4), 1201-1213. https://doi.org/10.2136/sssaj2011.0025
- Jiménez-González, M. A., De la Rosa, J. M., Jiménez-Morillo, N. T., Almendros, G., González-Pérez, J. A., & Knicker, H. (2016). Post-fire recovery of soil organic matter in a Cambisol from typical Mediterranean forest in Southwestern Spain. *Science of The Total Environment, 572*, 1414-1421. https://doi.org/10.1016/j.scitotenv.2016.02.134
- Knadel, M., Stenberg, B., Deng, F., Thomsen, A., & Greve, M. H. (2013). Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination. *Journal of Near Infrared Spectroscopy, 21*(1), 67-80. https://doi.org/10.1255/jnirs.1035
- Kobza, J. (2015). Permanent soil monitoring system as a basic tool for protection of soils and sustainable land use in Slovakia. *IOP Conference Series: Earth and Environmental Science, 25*(1), 1-10. https://doi.org/10.1088/1755-1315/25/1/012011
- Lal, R. (Ed.), & Stewart, B. A. (Ed.). (2010). *Food Security and Soil Quality.* Boca Raton, FL, USA: CRC Press (Taylor & Francis Group).
- Laudicina, V. A., De Pasquale, C., *Conte, P., Badalucco, L., Alo*nzo, G., & Palazzolo, E. (2012). Effects of afforestation with four unmixed plant species on the soil–water interactions in a semiarid Mediterranean region (Sicily, Italy). *Journal of Soils and Sediments, 12*(8), 1222-1230. https://doi.org/10.1007/s11*368-012-0522-0*
- Lu, D., Li, G., Valladares, G. S., & Batistella, M. (2004). Mapping soil erosion risk in Rondônia, Brazilian Amazonia: using RUSLE, remote sensing and GIS. *Land Degradation & Development, 15*(5), 499-512. https://doi.org/10.1002/ldr.634
- Lugassi, R., Ben-Dor, E., & Eshel, G. (2014). Reflectance spectroscopy of soils post-heating—Assessing thermal alterations in soil minerals. *Geoderma, 213*(0), 268-279. https://doi.org/10.1016/j.geoderma.2013.08.014

- Magidson, J. (2010). Correlated component regression: a prediction/classification methodology for possibly many features. *Proceedings of the American Statistical Association*. pp.4372-4386.
- Magidson, J. (2013). Correlated component regression: Re-thinking regression in the presence of near collinearity. In H. Abdi, W. W. Chin, V. E. Vinzi, G. Russolillo, & L. Trinchera (Eds.), *New Perspectives in Partial Least Squares and Related Methods* (pp. 65-78): New York, NY, USA: Springer International Publishing. https://doi.org/10.1007/978-1-4614-8283-3_3
- Magidson, J., & Wassmann, K. (2010). The role of proxy genes in predictive models: an application to early detection of prostate cancer. *Proceedings of the American Statistical Association, Biometrics section*. pp.2739-2753.
- Martens, H., & Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference, 11*(1-2), 5-16. https://doi.org/10.1016/S0950-3293(99)00039-7
- Melendez-Pastor, I., Navarro-Pedreño, J., Gómez, I., & Koch, M. (2008). Identifying optimal spectral bands to assess soil properties with VNIR radiometry in semi-arid soils. *Geoderma, 147*(3), 126-132. https://doi.org/10.1016/j.geoderma.2008.08.004
- Merino, A., Moreno, G., Navarro, F. B., & Gallardo, J. F. (2016). Future issues. In J. F. Gallardo (Ed.), *The Soils of Spain* (pp. 189-195). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-20541-0_6
- Montanarella, L., & Panagos, P. (2018). Soil Data Needs for Sustainable Agriculture. In: H. Ginzky, E. Dooley , I. Heuser, E. Kasimbazi, T. Markus, & T. Qin (Eds). *International Yearbook of Soil Law and Policy 2017*. (IYSLP, Vol 2017). (pp. 151-166). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-68885-5_9
- Mouazen, A. M., Karoui, R., De Baerdemaeker, J., & Ramon, H. (2005). Classification of Soil Texture Classes by Using Soil Visual near Infrared Spectroscopy and Factorial Discriminant Analysis Techniques. *Journal of Near Infrared Spectroscopy, 13*(4), 231-240. https://doi.org/10.1255/jnirs.541
- Mouazen, A. M., Kuang, B., De Baerdemaeker, J., & Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma, 158*(1), 23-31. https://doi.org/10.1016/j.geoderma.2010.03.001
- Nadal-Romero, E., Cammeraat, E., Pérez-Cardiel, E., & Lasanta, T. (2016). Effects of secondary succession and afforestation practices on soil properties after cropland abandonment in humid Mediterranean mountain areas. *Agriculture, Ecosystems & Environment, 228*, 91-100. https://doi.org/10.1016/j.agee.2016.05.003
- Nadal-Romero, E., Regüés, D., Martí-Bono, C., & Serrano-Muela, P. (2007). Badland dynamics in the Central Pyrenees: temporal and spatial patterns of weathering processes. *Earth Surface Processes and Landforms, 32*(6), 888-904. https://doi.org/10.1002/esp.1458
- Novara, A., La Mantia, T., Barbera, V., & Gristina, L. (2012). Paired-site approach for studying soil organic carbon dynamics in a Mediterranean semiarid environment. *Catena, 89*(1), 1-7. https://doi.org/10.1016/j.catena.2011.09.008
- Nuzzo, R. (2014). Scientific method: statistical errors. *Nature, 506*(7487), 150. https://doi.org/10.1038/506150a

- Ogen, Y., Neumann, C., Chabrillat, S., Goldshleger, N., & Ben-Dor, E. (2018). Evaluating the detection limit of organic matter using point and imaging spectroscopy. *Geoderma, 321*, 100-109. https://doi.org/10.1016/j.geoderma.2018.02.011
- Ostovari, Y., Ghorbani-Dashtaki, S., Bahrami, H.-A., Abbasi, M., Demattê, J. A. M., Arthur, E., & Panagos, P. (2018). Towards prediction of soil erodibility, SOM and CaCO3 using laboratory Vis-NIR spectra: A case study in a semi-arid region of Iran. *Geoderma, 314*, 102-112. https://doi.org/10.1016/j.geoderma.2017.11.014
- Pausas, J. G., Llovet, J., Rodrigo, A., & Vallejo, R. (2009). Are wildfires a disaster in the Mediterranean basin?–A review. *International Journal of Wildland Fire, 17*(6), 713-723. https://doi.org/10.1071/WF07151
- Pérez-Cabello, F., Echeverría, M., Ibarra, P., & Riva, J. (2009). Effects of Fire on Vegetation, Soil and Hydrogeomorphological Behavior in Mediterranean Ecosystems. In E. Chuvieco (Ed.), *Earth Observation of Wildland Fires in Mediterranean Ecosystems* (pp. 111-128). Berlin, Heidelberg: Springer International Publishing. https://doi.org/10.1007/978-3-642-01754-4_9
- Pimentel, D. (2006). Soil erosion: a food and environmental threat. *Environment, Development and Sustainability, 8*(1), 119-137. https://doi.org/10.1007/s10668-005-1262-8
- Pinheiro, É. F., Ceddia, M. B., Clingensmith, C. M., Grunwald, S., & Vasques, G. M. (2017). Prediction of Soil Physical and Chemical Properties by Visible and Near-Infrared Diffuse Reflectance Spectroscopy in the Central Amazon. *Remote Sensing, 9*(4), 293. https://doi.org/10.3390/rs9040293
- Rosero-Vlasova, O., Borini Alves, D., Vlassova, L., Perez-Cabello, F., & Montorio Lloveria, R. (2017). *Modeling soil organic matter (SOM) from satellite data using VISNIR-SWIR spectroscopy and PLS regression with step-down variable selection algorithm: case study of Campos Amazonicos National Park savanna enclave, Brazil.* Paper presented at the SPIE Remote Sensing Conference. USA: The International Society for Optics and Photonics. https://doi.org/10.1117/12.2278701
- Rosero-Vlasova, O., Pérez-Cabello, F., Montorio Llovería, R., & Vlassova, L. (2016). Assessment of laboratory VIS-NIR-SWIR setups with different spectroscopy accessories for characterisation of soils from wildfire burns. *Biosystems Engineering, 152,* 51-67. https://doi.org/10.1016/j.biosystemseng.2016.06.011
- Rosero-Vlasova, O. A., Vlassova, L., Pérez-Cabello, F., Montorio, R., & Nadal-Romero, E. (2018). Modeling soil organic matter and texture from satellite data in areas affected by wildfires and cropland abandonment in Aragón, Northern Spain. *Journal of Applied Remote Sensing, 12*(4), 042803. https://doi.org/10.1117/1.JRS.12.042803
- Ruiz-Rodado, V., Marcos Luque-Baena, R., te Vruchte, D., Probert, F., H. Lachmann, R., J. Hendriksz, C., . . . Grootveld, M. (2014). 1H NMR-Linked Urinary Metabolic Profiling of Niemann-Pick Class C1 (NPC1) Disease: Identification of Potential New Biomarkers using Correlated Component Regression (CCR) and Genetic Algorithm (GA) Analysis Strategies. *Current Metabolomics, 2*(2), 88-121. https://doi.org/10.2174/2213235X02666141112215616
- Ruiz de la Torre, J. (1990). Distribución y características de las masas forestales españolas. *Ecología* (Nº. 1), (pp. 11-13). Retrieved from https://dialnet.unirioja.es
- Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Chapter five-visible and near infrared spectroscopy in soil science. *Advances in Agronomy, 107*, 163-215. https://doi.org/10.1016/S0065-2113(10)07005-7
- Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of soil organic carbon at the European scale by visible and near infrared

reflectance spectroscopy. *PloS One, 8*(6), e66409. https://doi.org/10.1371/journal.pone.0066409

- Terra, F. S., Demattê, J. A. M., & Viscarra Rossel, R. A. (2018). Proximal spectral sensing in pedological assessments: vis–NIR spectra for soil classification based on weathering and pedogenesis. *Geoderma, 318*, 123-136. https://doi.org/10.1016/j.geoderma.2017.10.053
- Tóth, G., Hermann, T., da Silva, M. R., & Montanarella, L. (2018). Monitoring soil for sustainable development and land degradation neutrality. *Environmental Monitoring and Assessment, 190*(2), 57. https://doi.org/10.1007/s10661-017-6415-3
- Troeh, F. R., & Thompson, L. M. (2005). Chapter 20-Soil Erosion and Its Control. *Soils and Soil Fertility*, (6th ed.). Ames, Iowa, USA: Blackwell (John Wiley & Sons).
- United Nations. (2015). Transforming our world: the 2030 Agenda for Sustainable Development. *Resolution Adopted by the General Assembly,* (A/RES/70/1). United Nations Headquarters, New York, NY, USA: Press release. Retrieved from https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf
- USDA. (1996). *Soil Survey Manual*. Retieved from Washington, USA: https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcseprd1335011.pdf
- Vasques, G. M., Demattê, J. A. M., Viscarra Rossel, R. A., Ramírez-López, L., & Terra, F. S. (2014). Soil classification using visible/near-infrared diffuse reflectance spectra from multiple depths. *Geoderma, 223–225*(0), 73-78. https://doi.org/10.1016/j.geoderma.2014.01.019
- Viscarra Rossel, R., & Behrens, T. (2010). Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma, 158*(1), 46-54. https://doi.org/10.1016/j.geoderma.2009.12.025
- Viscarra Rossel, R., & Brus, D. J. (2018). The cost-efficiency and reliability of two methods for soil organic C accounting. *Land Degradation & Development*, *29*(3), 506-52. https://doi.org/10.1002/ldr.2887
- Viscarra Rossel, R., Walvoort, D., McBratney, A., Janik, L. J., & Skjemstad, J. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma, 131*(1-2), 59-75. https://doi.org/10.1016/j.geoderma.2005.03.007
- Viscarra Rossel, R. A., McGlynn, R. N., & McBratney, A. B. (2006). Determining the composition of mineral-organic mixes using UV–vis–NIR diffuse reflectance spectroscopy. *Geoderma, 137*(1–2), 70-82. https://doi.org/10.1016/j.geoderma.2006.07.004
- Vlassova, L., & Pérez-Cabello, F. (2016). Effects of post-fire wood management strategies on vegetation recovery and land surface temperature (LST) estimated from Landsat images. *International Journal of Applied Earth Observation and Geoinformation, 44*, 171-183. https://doi.org/10.1016/j.jag.2015.08.011
- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109-130. https://doi.org/10.1016/S0169-7439(01)00155-1

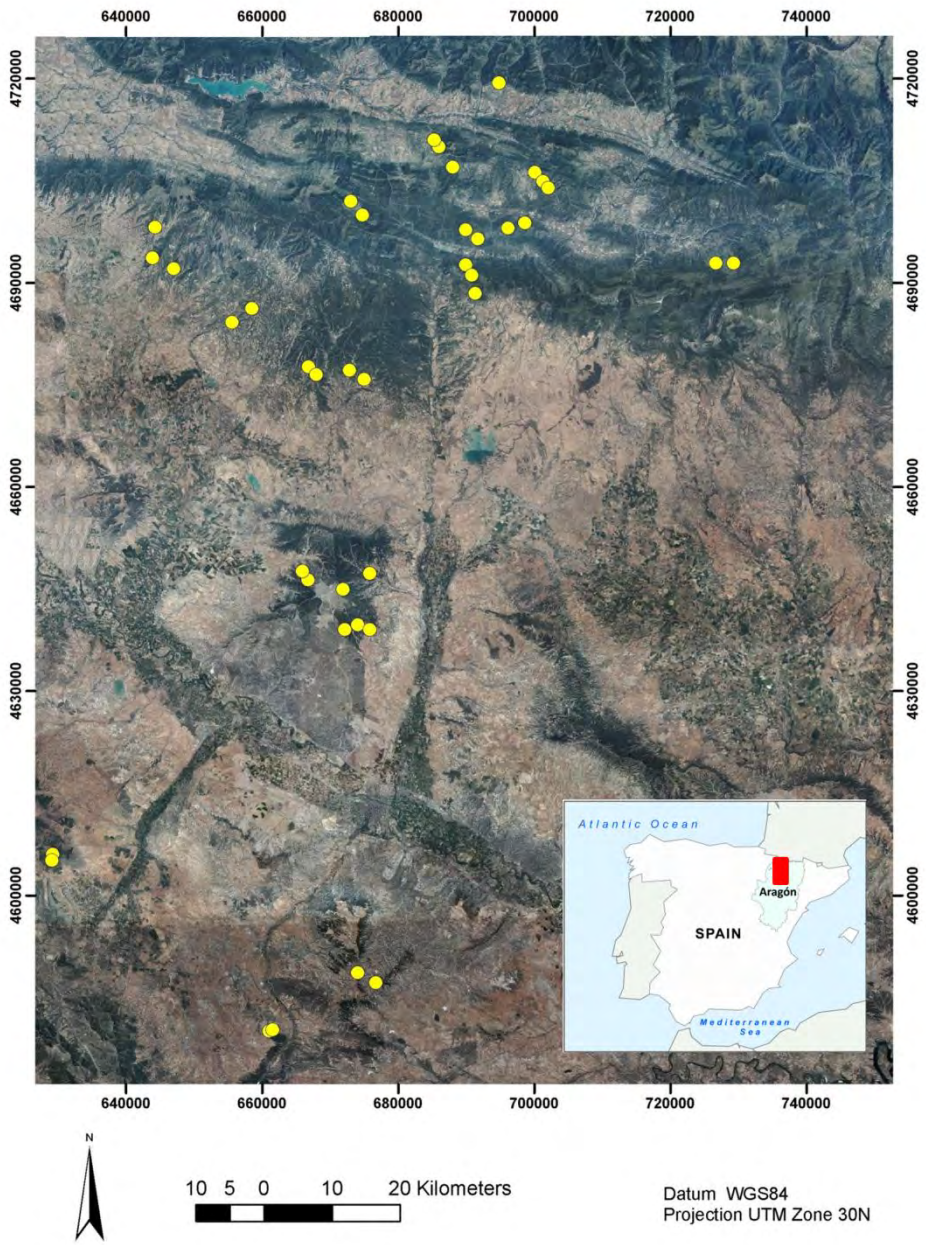


Figure 1. Location of the study area and sampling sites.

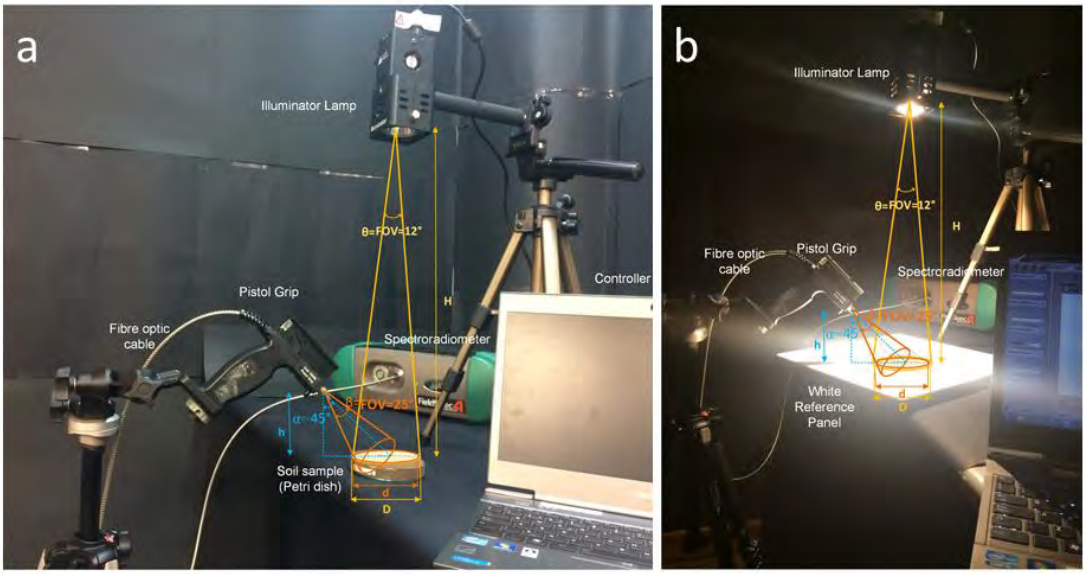


Figure 2. Experimental setup: (a) general view and (b) view during spectral measurements (optimization process).

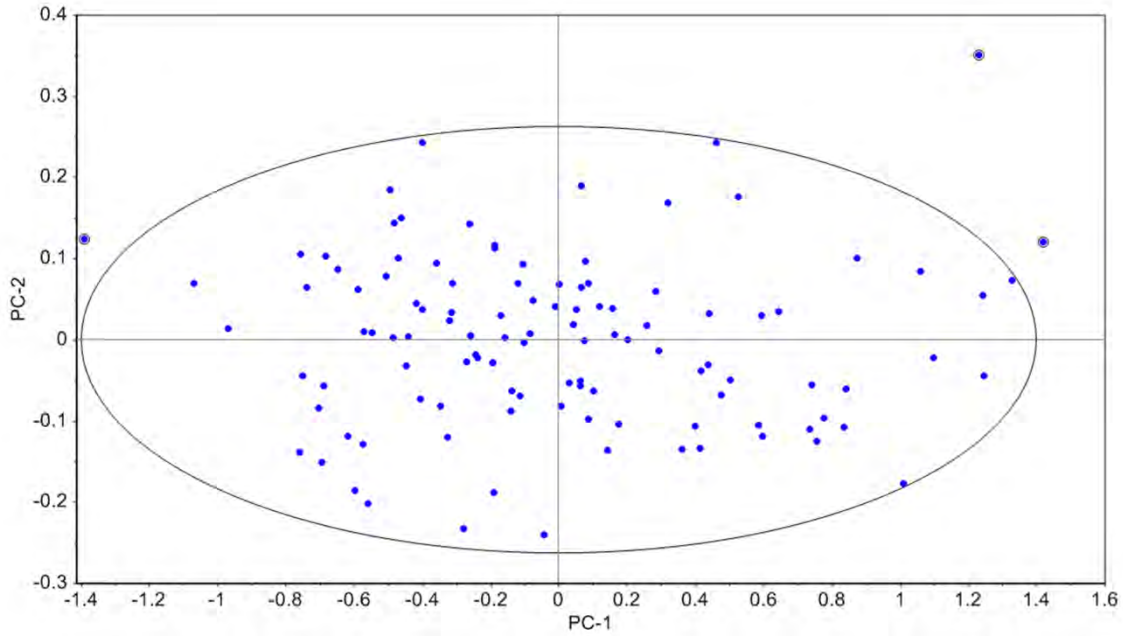


Figure 3. Score plot from principal components analysis (PCA) showing PC1 vs. PC2: Hotelling's T^2 ellipse (95% confidence level) for outlier detection.

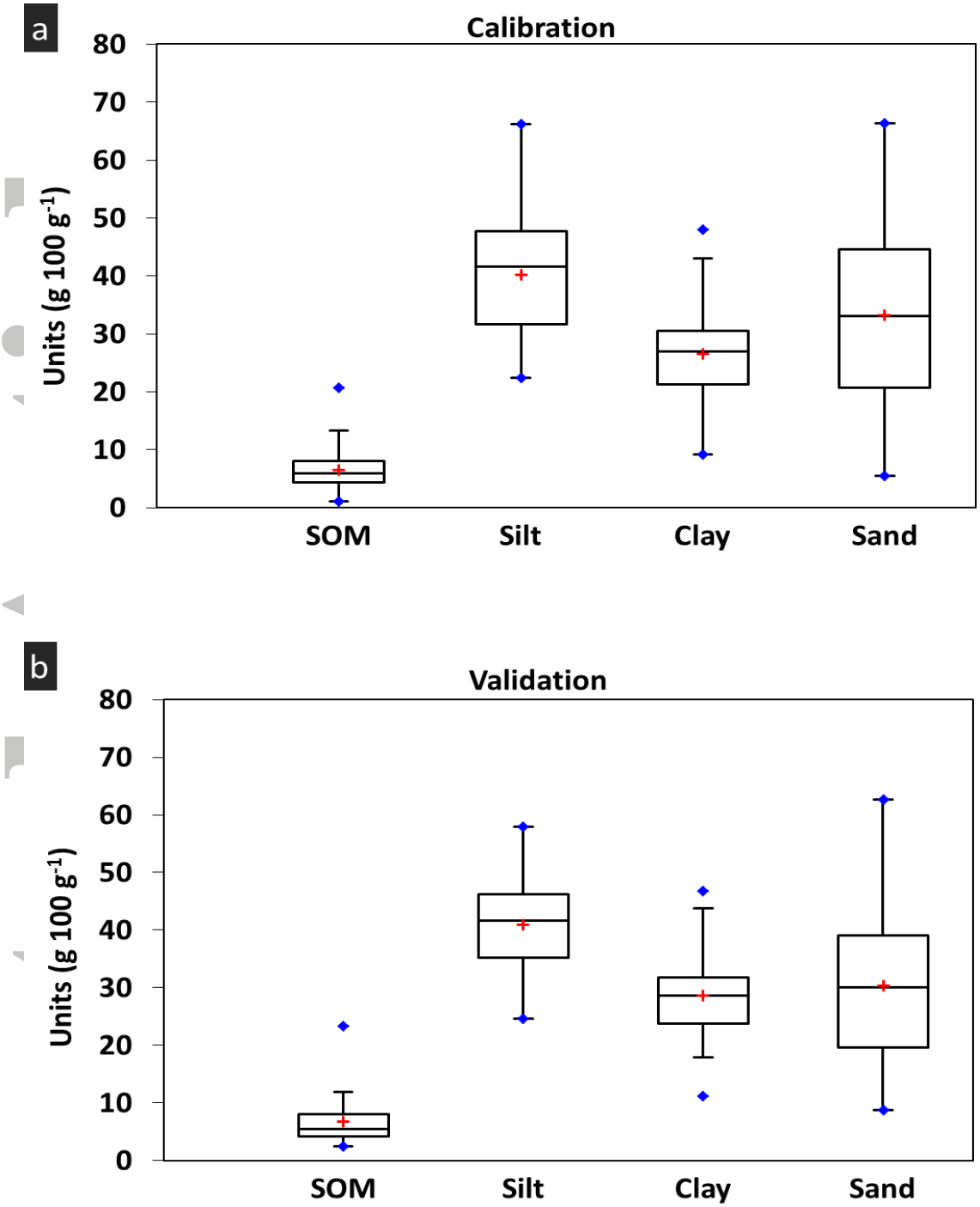


Figure 4. Box-whisker plots showing the SOM, clay, silt and sand distribution in S2 for the (a) calibration set and (b) validation set. The bottom and top of the box represent the 25th and 75th percentiles. The cross inside the box indicates the mean value. The band near the middle of the box is the median. The whiskers represent the 5th and 95th percentiles. The solid dots correspond to maximum and minimum values.

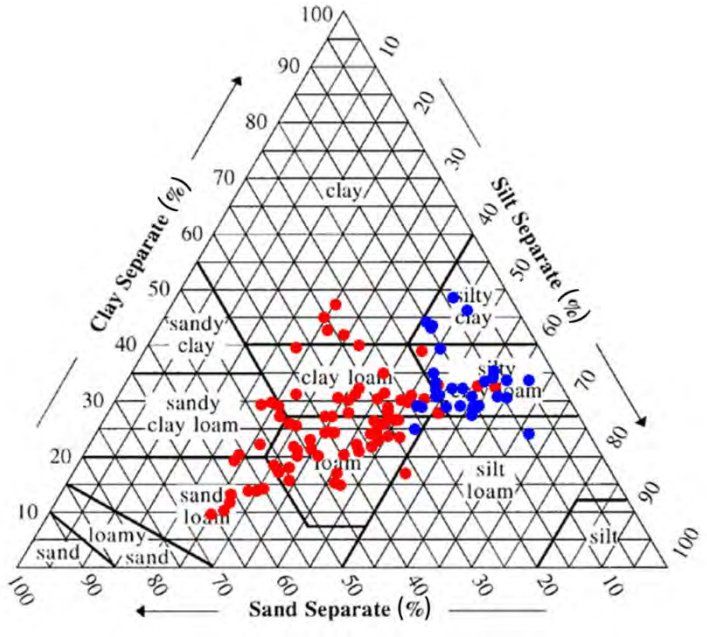


Figure 5. A soil texture triangle (USDA, 2010) showing the soil textures as determined by the proportion of sand, silt and clay. The red points represent soil samples from areas affected by wildfires and the blue points represent soil samples from areas of crop abandonment.

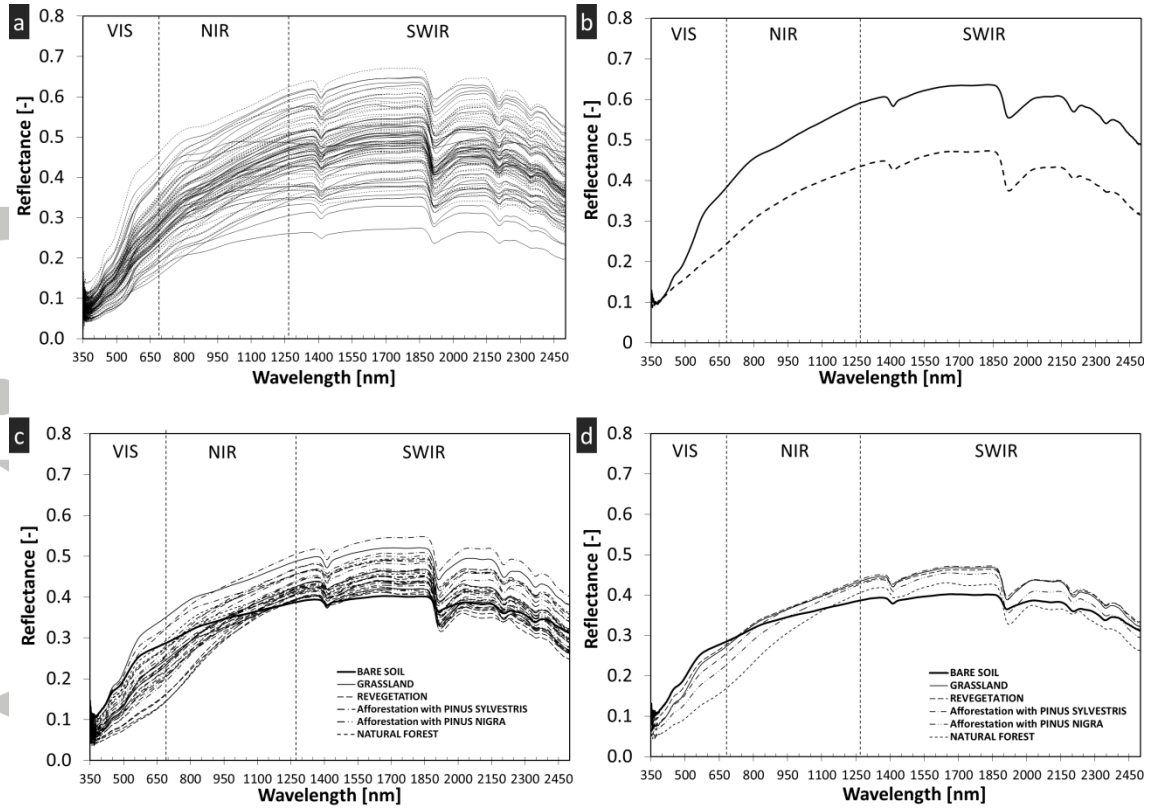


Figure 6. Spectra of soils from wildfire burns: (a) all the spectra and (b) spectra of the burned (dotted line) and unburned (dashed line) soil samples. Spectra of soils from agricultural abandonment areas: (c) all the spectra and (d) average spectra of soils from areas with different land use types after crop abandonment.

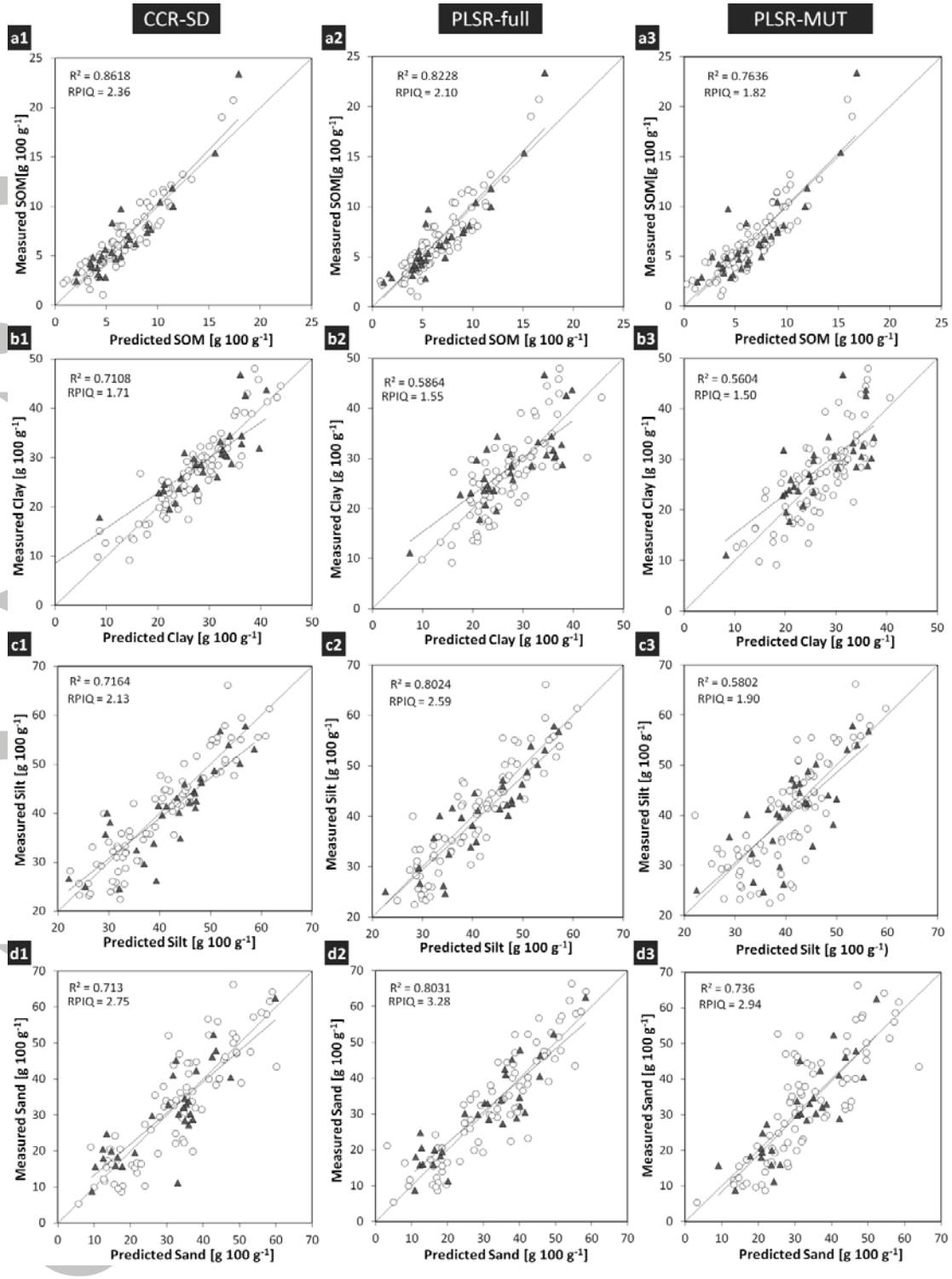
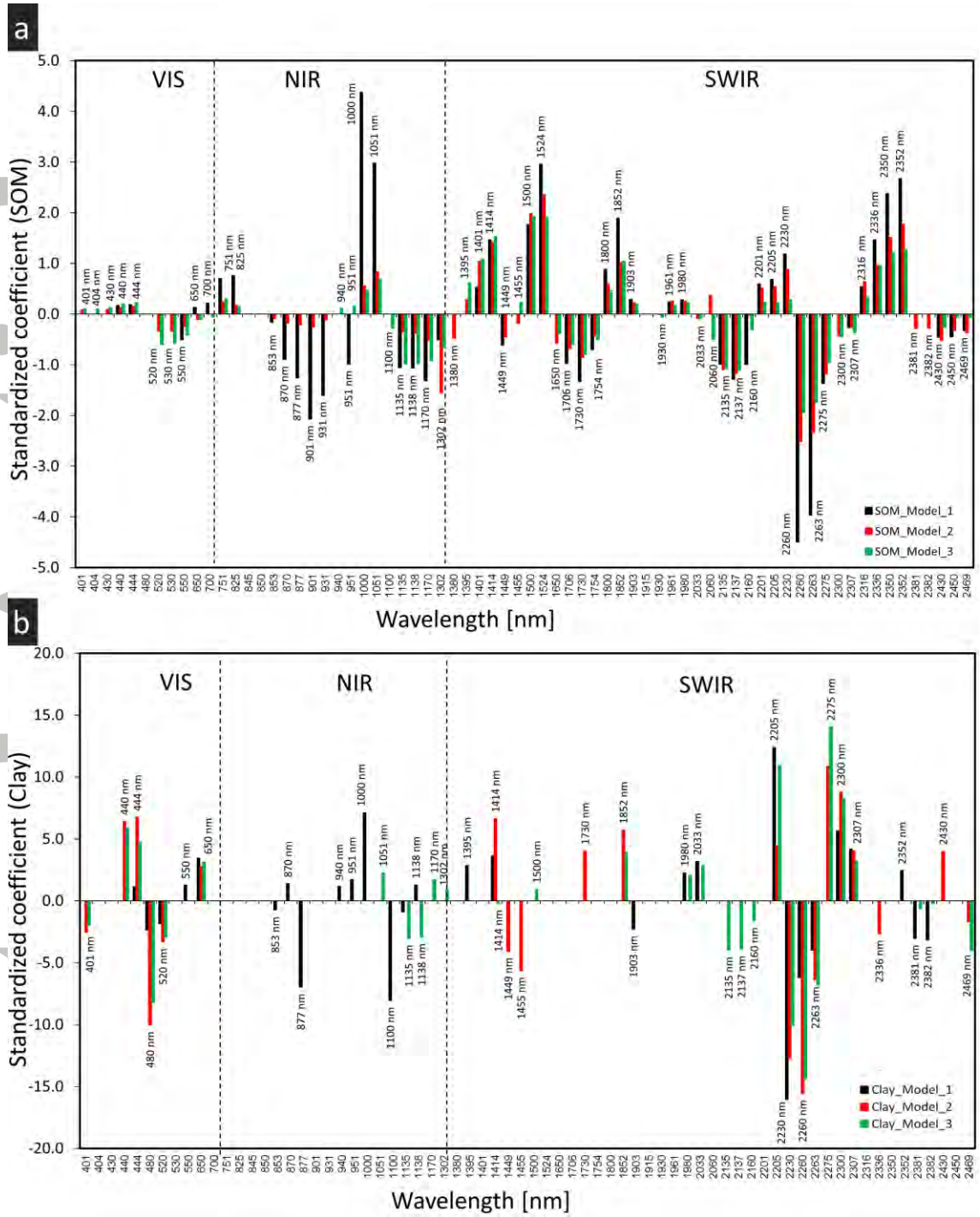


Figure 7. Scatter plot examples of predicted versus observed values of SOM (a), clay (b), silt (c) and sand (d) for S2 for CCR-SD, PLSR-full and PLSR-MUT models. In each plot, the data points corresponding to calibration are shown as circles, and the data points corresponding to validation are shown as triangles.



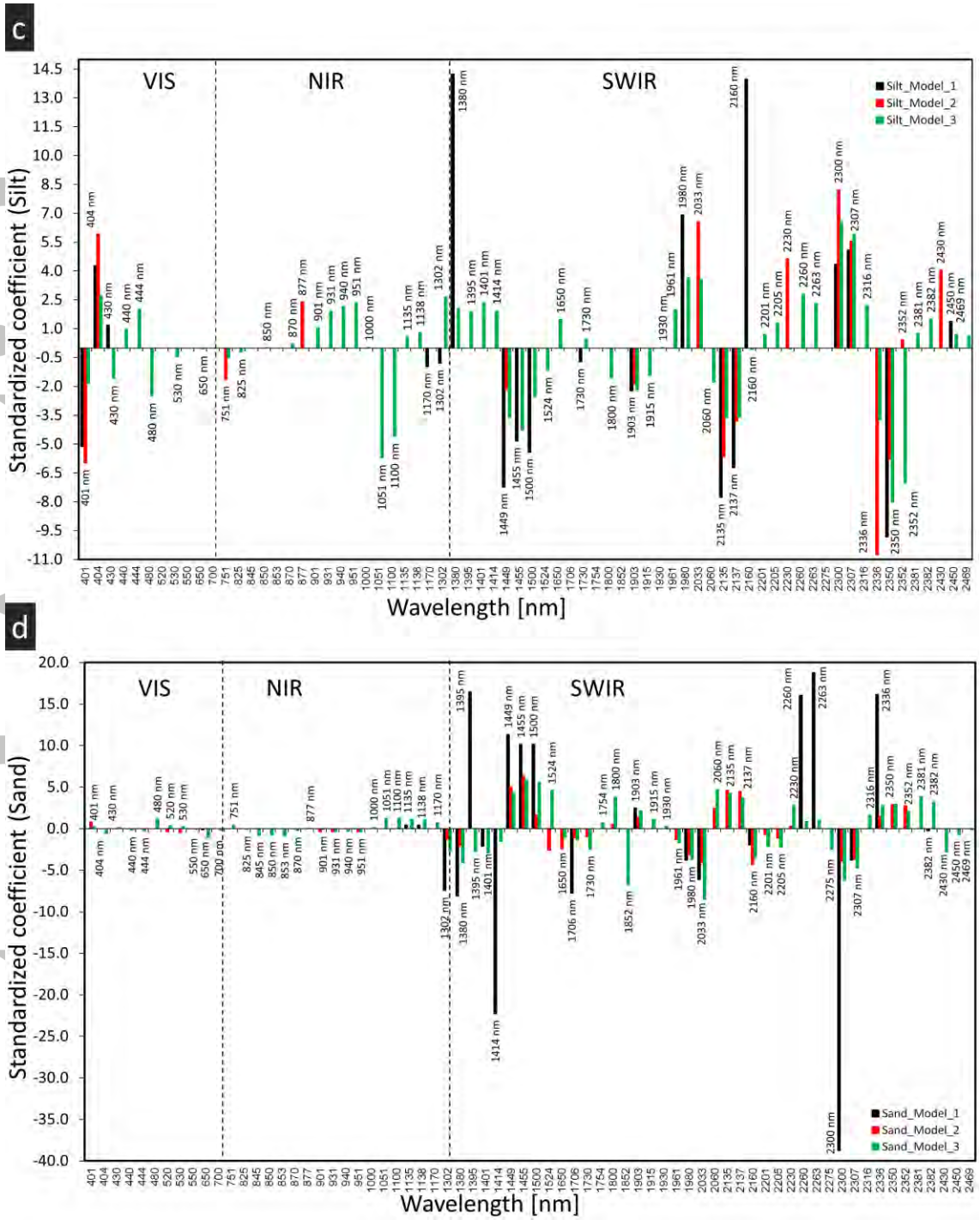


Figure 8. Band coefficients in the predictive models for S1, S2 and S3 for (a) SOM and (b) clay (c) silt and (d) sand.

Table 1. Descriptive statistics for soil organic matter (SOM) content (g $100g^{-1}$) and clay, silt and sand (%) in collected soil samples.

Soil properties	n	Min	Max	Median	Mean	SD
SOM (g 100 g^{-1})	113	1.04	23.40	5.80	6.59	3.68
Clay (%)	113	9.21	48.04	27.28	27.14	8.09
Silt (%)	113	22.18	66.19	41.48	40.23	10.26
Sand (%)	113	5.41	66.37	32.16	32.64	14.57

NOTE. n: number of samples; Min: minimum; Max: maximum; SD: standard deviation.

Table 2. Descriptive statistics for soil organic matter (SOM) content (g $100g^{-1}$), clay, silt and sand (%) in three calibration and prediction sample sets (S1, S2 and S3).

Soil properties	Sample set	CALIBRATION								VALIDATION							
		n	Min	Max	Mean	SD	Q25	Q75	Skewness (Pearson)	n	Min	Max	Mean	SD	Q25	Q75	Skewness (Pearson)
SOM (g $100\ g^{-1}$)	**S1**	80	1.04	20.74	6.28	3.45	4.08	6.69	1.79	30	2.59	23.40	7.46	4.30	4.86	9.72	1.78
	S2	80	1.04	20.74	6.53	3.49	4.34	8.00	1.55	30	2.46	23.40	6.81	4.33	4.20	7.99	2.17
	S3	80	1.04	23.40	6.35	3.52	4.25	7.73	2.35	30	1.77	19.02	7.27	4.19	4.52	10.32	0.93
Clay (%)	**S1**	80	9.82	48.04	27.33	8.32	29.63	42.28	0.22	30	9.21	43.77	26.72	7.86	32.45	47.57	-0.08
	S2	80	9.21	48.04	26.59	8.38	31.70	47.77	0.21	30	11.18	46.78	28.69	7.49	35.15	46.27	-0.12
	S3	80	9.21	48.04	27.34	8.39	32.36	47.21	0.15	30	11.18	43.00	26.70	7.66	32.45	47.47	0.10
Silt (%)	**S1**	80	22.46	66.19	40.44	10.30	19.86	29.77	0.40	30	23.66	55.89	40.21	10.30	22.90	30.53	-0.16
	S2	80	22.46	66.19	40.15	10.70	21.25	30.48	0.31	30	24.68	57.91	40.99	9.08	23.80	31.83	0.34
	S3	80	23.28	66.19	40.36	10.20	21.75	31.74	0.31	30	22.46	61.39	40.43	10.57	22.87	30.71	0.10
Sand (%)	**S1**	80	5.41	64.22	32.23	14.66	31.43	45.37	0.20	30	10.80	66.37	33.11	14.85	20.75	44.25	0.34
	S2	80	5.41	66.37	33.27	15.25	20.64	44.58	0.14	30	8.80	62.58	30.33	12.92	19.68	39.08	0.46
	S3	80	8.69	66.37	32.32	14.68	20.45	42.11	0.30	30	5.41	62.58	32.87	14.81	20.12	44.25	0.07

NOTE. S1: sample set 1; S2: sample set 2; S3: sample set 3; n: number of samples; Min: minimum; Max: maximum; SD: standard deviation; Q25: lower quartile; Q75: upper quartile.

Table 3. Calibration and cross-validation results of soil organic matter (SOM), clay, silt and sand modelling (80 samples) obtained with CCR-SD, PLSR-full and PLSR-MUT.

Soil Properties	Sample set	Statistic	Factors	*f*	R^2_C	RMSEC (g 100 g^{-1})	R^2_{CV}	RMSECV (g 100 g^{-1})
SOM	S1	CCR-SD	8	49	0.86	2.11	0.77	1.30
		PLSR-full	10	2071	0.81	1.49	0.73	1.82
		PLSR-MUT	9	171	0.82	1.46	0.76	1.70
	S2	CCR-SD	8	61	0.86	2.75	0.78	1.31
		PLSR-full	9	2071	0.81	1.51	0.72	1.86
		PLSR-MUT	8	39	0.79	1.60	0.73	1.82
	S3	CCR-SD	8	58	0.85	2.62	0.77	1.35
		PLSR-full	10	2071	0.85	1.37	0.78	1.65
		PLSR-MUT	9	197	0.82	1.46	0.75	1.77
CLAY	S1	CCR-SD	10	28	0.80	4.68	0.67	3.76
		PLSR-full	10	2071	0.73	4.32	0.58	5.41
		PLSR-MUT	9	171	0.64	4.96	0.51	5.86
	S2	CCR-SD	10	21	0.83	4.02	0.60	3.45
		PLSR-full	9	2071	0.64	5.01	0.46	6.22
		PLSR-MUT	8	39	0.61	5.18	0.49	6.04
	S3	CCR-SD	10	29	0.83	4.32	0.67	3.44
		PLSR-full	10	2071	0.69	4.63	0.54	5.76
		PLSR-MUT	9	197	0.67	4.81	0.53	5.79
SILT	S1	CCR-SD	10	19	0.86	4.39	0.75	3.83
		PLSR-full	10	2071	0.83	4.23	0.73	5.42
		PLSR-MUT	9	171	0.80	4.52	0.73	5.36
	S2	CCR-SD	10	16	0.84	4.78	0.70	4.27
		PLSR-full	9	2071	0.81	4.67	0.64	6.43
		PLSR-MUT	8	39	0.62	6.58	0.40	8.32
	S3	CCR-SD	10	58	0.86	7.40	0.72	3.82
		PLSR-full	10	2071	0.78	4.66	0.62	6.32
		PLSR-MUT	9	197	0.75	5.10	0.52	6.32
SAND	S1	CCR-SD	10	22	0.84	6.90	0.71	5.86
		PLSR-full	10	2071	0.86	5.46	0.76	7.23
		PLSR-MUT	9	171	0.78	6.87	0.70	8.11
	S2	CCR-SD	8	32	0.75	9.89	0.61	7.63
		PLSR-full	9	2071	0.81	6.68	0.63	9.32
		PLSR-MUT	8	39	0.65	9.00	0.47	11.20
	S3	CCR-SD	10	70	0.79	19.81	0.61	6.69
		PLSR-full	10	2071	0.81	6.40	0.65	8.78
		PLSR-MUT	9	197	0.79	6.64	0.67	8.52

NOTE. S1: sample set 1; S2: sample set 2; S3: sample set 3; *f*: number of variables used in the regression equation; R^2_C: coefficient of determination for calibration; R^2_{CV}: coefficient of determination for cross-validation; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross-validation. The presented values are averages for 100 rounds.

Table 4. Validation results for soil organic matter (SOM), clay, silt and sand modelling (30 samples) obtained with CCR-SD, PLSR-full and PLSR-MUT.

Soil Properties	Sample set	Statistic	R^2_V	RMSEP (g 100 g^{-1})	SEP_c (g 100 g^{-1})	$bias_{val}$ (g 100 g^{-1})	RPIQ	AIC
SOM	S1	CCR-SD	0.87	1.55	1.53	0.36	**3.14**	**111.13**
		PLSR-full	0.88	1.59	1.58	-0.33	3.06	4155.91
		PLSR-MUT	0.77	2.04	2.07	0.08	2.38	363.39
	S2	CCR-SD	0.86	1.61	1.64	0.07	**2.36**	136.30
		PLSR-full	0.82	1.81	1.83	-0.10	2.10	4159.80
		PLSR-MUT	0.76	2.08	2.11	0.03	1.82	99.97
	S3	CCR-SD	0.89	1.89	1.52	1.14	**3.08**	**135.03**
		PLSR-full	0.89	1.96	1.55	-1.24	2.96	4162.19
		PLSR-MUT	0.86	2.05	1.75	-1.12	2.93	415.54
CLAY	S1	CCR-SD	0.54	5.98	6.01	-0.95	**1.28**	**109.66**
		PLSR-full	0.53	6.91	6.68	2.14	1.10	4199.99
		PLSR-MUT	0.28	7.05	7.15	0.44	1.08	400.59
	S2	CCR-SD	0.71	4.71	4.76	0.52	**1.71**	**88.47**
		PLSR-full	0.59	5.18	5.21	-0.75	1.55	4191.34
		PLSR-MUT	0.56	5.34	5.21	-1.49	1.50	128.26
	S3	CCR-SD	0.73	4.47	4.47	0.83	**1.75**	**102.94**
		PLSR-full	0.55	5.80	5.69	-1.54	1.35	4194.74
		PLSR-MUT	0.57	6.14	5.82	-2.21	1.28	448.44
SILT	S1	CCR-SD	0.70	7.08	6.68	2.65	2.14	**96.73**
		PLSR-full	0.74	6.30	5.63	-3.02	2.40	4197.22
		PLSR-MUT	0.65	8.48	7.81	-3.60	1.78	406.13
	S2	CCR-SD	0.72	5.23	5.10	-1.48	2.13	**81.63**
		PLSR-full	0.80	4.29	4.09	1.51	2.59	4185.69
		PLSR-MUT	0.58	5.87	5.93	0.65	1.90	131.10
	S3	CCR-SD	0.68	6.80	6.60	2.01	2.21	**173.49**
		PLSR-full	0.80	5.12	5.21	0.09	2.93	4190.99
		PLSR-MUT	0.64	7.09	7.02	-1.62	2.12	452.76
SAND	S1	CCR-SD	0.69	8.76	8.61	-2.25	**2.68**	**109.10**
		PLSR-full	0.66	9.37	9.49	0.84	2.51	4209.13
		PLSR-MUT	0.59	11.62	11.38	3.12	2.02	415.58
	S2	CCR-SD	0.71	7.04	7.13	0.66	2.75	**122.55**
		PLSR-full	0.80	5.92	5.97	-0.75	3.28	4195.35
		PLSR-MUT	0.74	6.59	6.65	0.85	2.94	134.57
	S3	CCR-SD	0.69	9.18	8.75	-3.20	2.63	**206.50**
		PLSR-full	0.79	7.09	7.06	1.45	3.40	4200.76
		PLSR-MUT	0.71	9.31	8.62	3.84	2.59	460.93

NOTE. S1: sample set 1; S2: sample set 2; S3: sample set 3; R^2_V: coefficient of determination for validation; RMSEP: root mean square error of prediction; SEP_C: standard error of prediction; $bias_{val}$: bias of validation; RPIQ: ratio of performance to interquartile range; AIC: Akaike information criterio.