Soil & Tillage Research 172 (2017) 59–68

Contents lists available at ScienceDirect

Soil & Tillage Research

journal homepage: www.elsevier.com/locate/still

Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy

Andre Carnieletto Dotto[a,*], Ricardo Simão Diniz Dalmolin[a], Sabine Grunwald[b], Alexandre ten Caten[c], Waterloo Pereira Filho[d]

[a] Department of Soil, Federal University of Santa Maria, Room 3314, Building 42, CCR, Av. Roraima 1000, CEP 97105-900, Santa Maria, RS, Brazil
[b] Soil and Water Sciences Department, University of Florida, 2181 McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA
[c] Department of Agriculture, Biodiversity and Forestry, Federal University of Santa Catarina, Rod. Ulysses Gaboardi, km 3, Caixa Postal 101, CEP 89520-000, Curitibanos, SC, Brazil
[d] Department of Geoscience, National Institute for Space Research, INPE, Room 2048, UFSM, Av. Roraima 1000, CEP 97105-900, Santa Maria, RS, Brazil

ARTICLE INFO

Keywords:
Visible-near infrared spectroscopy
Continuum removal
Detrend
Band ratio

ABSTRACT

Proximal sensing provides an alternative method to physical and chemical laboratory soil analyses. The aim of this study is to predict soil organic carbon (SOC), clay, sand, and silt content using reduced spectral features as covariables selected by two spectral preprocessing. A total of 299 soil samples were collected in Santa Catarina state, Brazil. Two preprocessing techniques, detrend transformation and continuum removal (CR), were applied to isolate particular absorption features in the reflectance spectrum. Two techniques were used to select the spectral features in the spectrum: hand and mathematical selection. Partial least squares regression (PLSR) and Support vector machines (SVM) were applied to predict the soil properties. The reduction of predictor covariables by hand selection technique contributed in developing a high-level prediction model for SOC. PLSR and SVM presented no statistical difference between the RMSE results, except for clay content, where SVM presented superior performance. The preprocessing techniques were statistically identical based on RMSE results. Overall, the prediction of SOC, clay, sand and silt presented suitable results using reduced spectral features as covariables in modeling process.

1. Introduction

Soil is one of the most important components of environmental resources and it has an enormous influence on agricultural productivity (Lal and Moldenhauer, 1987). Soil information is necessary to make decisions concerning management practices, food security (Andrews et al., 2004), and soil security (Koch et al., 2013; McBratney et al., 2014). Soil organic carbon (SOC) and particle size modulate nutrient supply, water holding capacity, soil structure aggregation, and erosion prevention. Moreover, SOC has a significant impact on the global carbon cycle as well as climate change (Janzen, 2004), and is recognized as a key component of well-functioning ecosystems (Stockmann et al., 2015).

To develop a faster and more accurate method for SOC and particle size analysis, proximal sensing has been successfully applied to predict these parameters (Conforti et al., 2015; Knox et al., 2015; Ramirez-Lopez et al., 2013). The visible-near infrared (Vis-NIR) reflectance region (350–2500 nm) stands out for its applicability to measure and predict a wide variety of properties of soil samples (Dalmolin et al., 2005; Viscarra Rossel et al., 2006). Vis-NIR uses spectral reflectance to identify properties without any interaction with objects and has the advantages of extensive soil sample volume analysis, non-intrusiveness, timeliness, and affordability (Viscarra Rossel et al., 2006). In addition, soil sample preprocessing is fast, without the use of environmentally harmful chemical reagents (McBratney et al., 2006; Viscarra Rossel et al., 2010). This new soil analysis approach can be considered an alternative to improve the conventional methods of analysis carried out in the laboratory (Minasny and McBratney, 2008).

In the new concept of digital soil morphometrics (Hartemink and Minasny, 2014), the application of tools, such as proximal soil sensing and techniques for measuring and quantifying soil attributes, help enhance pedological understanding. Consequently, spectral reflectance has been applied in soil survey, mapping, and quantitative soil property characterization. Various research teams have used preprocessing and regression analysis to predict various soil properties, but no single preprocessing method stood out as the best performing one among these studies (Araújo et al., 2014; Knox et al., 2015; Ramirez-Lopez et al., 2013; Stevens et al., 2010; Terra et al., 2015; Vasques et al., 2008;

* Corresponding author.
E-mail address: andrecdot@gmail.com (A.C. Dotto).

http://dx.doi.org/10.1016/j.still.2017.05.008
Received 17 December 2015; Received in revised form 14 May 2017; Accepted 15 May 2017
Available online 23 May 2017

Viscarra Rossel and Behrens, 2010). Despite these advances, research gaps exist regarding new modeling techniques that have the potential to improve the predictive capabilities using proximal sensing.

Relating spectral data to a specific soil property requires a mathematical model. This task is not simple because many factors can influence soil spectroscopy. Soil spectra are complex, and soil attributes interact in complex ways, masking correlations between specific spectral reflectance signatures and a specific soil property. Furthermore, the process is complicated because only overtones of the native chemical structures of soil constituents are found in the Vis-NIR spectrum. According to Wight et al. (2016), impacts from specific soil characteristics on NIR performance are not well understood. These authors created an association of artificial soils based on primary soil characteristics, where a single optimized NIR model's predictive capability was compared by each soil characteristic subset. They concluded that the type of organic matter can affect NIR's predictive ability and, depending upon the accuracy chosen, it may be possible to separate sample populations into categories based on the nature of the organic substrate. In addition, Wight et al. (2016) suggested that texture is the principal characteristic that interferes with the model's accuracy, and it affects the spectral reflectance in the entire region of the Vis-NIR. According to Ben-Dor et al. (1997), soil organic matter influences all of the Vis-NIR spectral region and customizes the shape and the albedo of the spectral curve.

Recently, preprocessing techniques have been utilized to transform soil spectral data, remove noise, accentuate features, and detect patterns, including smoothing, detrending, derivatives, averaging, normalization, scatter correction, non-linear transformations, and absorbance transformation. In Vasques et al. (2008), thirty preprocessing transformations were compared to predict soil carbon, e.g., Savitzky–Golay smoothing, averaging, normalization by the range, Norris Gap Derivative, Savitzky–Golay derivatives, and standard normal variate. To select spectral features of interest and make the spectra suitable for modeling by reducing the spectral covariates, detrend, continuum removal (CR) and band ratio (BR) preprocessing techniques can be applied. These preprocessing can be used to interpret and extract information from spectral reflectance sets and to identify spectral features related to specific soil properties. Detrend is applied for removing baseline of the signals. CR, proposed by Clark and Roush (1984), consists of removing the continuous features of the spectra and is often used to isolate specific absorption features present in the spectrum to minimize the noise partially. The continuum is represented by a mathematical function used to separate and highlight specific absorption bands of the reflectance spectrum (Mutanga et al., 2005). BR is used to emphasize how two wavelengths affect each other. This preprocessing has the advantage of combining information from two prominent features and it is an approach used to reduce the size of spectral data.

For soil property predictions from Vis-NIR spectra, a mathematical analysis is required to quantify each specific soil property. Generally, the most frequently used multivariate methods are partial least square regression (PLSR) (Chacón Iznaga et al., 2014; Conforti et al., 2015; Knox et al., 2015) and support vector machines (SVM) (Ramirez-Lopez et al., 2013; Terra et al., 2015). One obstacle related to soil spectra and soil property characterization is the complexity of soil components shown in the spectra (Ge et al., 2011; Wight et al., 2016). To solve this problem, SVM and PLSR methods were applied in this study. SVM is a non-parametric data mining method, and PLSR is the most common multivariate calibration model. SVM and PLSR have already shown good results in soil properties predictions (Araújo et al., 2014; Conforti et al., 2015; Knox et al., 2015; Kuang et al., 2015; Nawar et al., 2016; Stevens et al., 2010; Terra et al., 2015). Viscarra Rossel and Behrens (2010) compared the predictions using SVM and PLSR for SOC and clay content (n = 1104) using Vis-NIR spectroscopy. The authors presented the best number of wavelet coefficients to use in the regressions, showing that 72 coefficients produced the smallest RMSE when used to predict SOC, and 132 coefficients for clay content. The number of covariables in the model can be reduced based on the preprocessing selection to maintaining the robustness of prediction accuracy.

The motivation to undertake this study comes from different sources. First, there is a lack of studies applying spectral feature selection in order to reduce spectral covariables and improve soil property prediction. Second, the selection of spectral features facilitates understanding and reduces the multicollinearity of hyperspectral data. Third, there are few soil spectroscopy studies in Brazil. The objective is to predict SOC, clay, sand, and silt content using reduced spectral features as covariables selected by spectral preprocessing.

2. Material and methods

2.1. Study site and sample collection

Soil samples were collected in an area of about 1700 km^2 in the region within the watershed of the Marombas River in the central region of Santa Catarina state, Brazil. A total of 299 soil samples were collected following the GlobalSoilMap (Arrouays et al., 2014) depths specifications of 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm along with additional samples from profile horizons. The study area presented similar soils due to the homogeneity of the parent material, which were predominantly basalt rocks from a landscape dominated by a smooth relief plateau and few areas with sedimentary rock. According to the Köppen climate classification, the study area has a humid subtropical climate (Cfa). These factors have led to an advanced degree of weathering and the development of deep soils, such as Oxisols, which were predominant in the area and showed high concentrations of iron oxides. Low clay content values were measured in sandy soils, which were often characterized by intense water erosion and low SOC content caused by unsustainable agricultural practices. Moreover, soil samples with very low sand content were mostly associated with Oxisols. Furthermore, in some slope areas, it is possible to find shallow soils, such as Entisols and Inceptisols. The prominent land uses in this region were forest, grassland, and agriculture.

2.2. Soil analysis in the laboratory

The soil samples were sieved (2 mm) and dried at 45 °C (for 72 h) adopting the standard Brazilian soil analysis method (Donagemma et al., 2011). The soil particle size was determined according to the Pipette method using NaOH dispersant (Donagemma et al., 2011). The SOC was determined by total organic carbon content using the Mebius method in the digestion block (Yeomans and Bremner, 1988). Using this method, the soil organic matter is oxidized with a mixture of $K_2Cr_2O_7$ 0.167 mol L^{-1} and concentrated H_2SO_4, and the excess of dichromate is titrated with ferrous ammonium sulfate. The reduced dichromate during the reaction with the soil corresponds to organic carbon in the sample.

2.3. Spectral reflectance measurements

The spectral reflectance of soil samples was obtained using a FieldSpec 3 spectroradiometer (Analytical Spectral Devices, Boulder, USA) with a spectral range of 350–2500 nm and a spectral resolution of 1 nm. To carry out the spectral measurements, soil samples were distributed homogeneously in petri dishes. The spectral sensor that was used captured the light through a fiber optic cable allocated 8 cm from the sample surface. The sensor reading area was approximately 2 cm^2 and the lighting was provided by two external halogen lamps of 50 W. The lamps were positioned at a distance of 35 cm from the sample (non-collimated rays and zenithal angle of 30°) and between them at an angle of 90°. A Spectralon standard white plate was scanned every 20 min for calibration. For each sample, two replications (one involving a 180° turn of the petri dish) were obtained. Each spectrum

was averaged from 100 readings over 10 s. Mean values of two replicates were adopted for each subsample.

2.4. Spectral preprocessing

Soil spectral data were smoothed by the Savitzky–Golay first-order polynomial across a moving window of five bands (Savitzky and Golay, 1964) to reduce the noise. The first order detrending transformation was used to remove the baseline of the signals in the spectral data (Barnes et al., 1989) and isolate particular absorption features. The detrend function, which is recommended only when the overall signal is dominated by backgrounds that are generally of the same shape, is recommended to be utilized prior to the multivariate analysis (Barnes et al., 1989). The CR was used to isolate particular absorption features in the reflectance spectra (Clark and Roush, 1984). CR allowed the normalization of the spectra and thereby facilitated the identification of significant absorption features that ranged across the Vis-NIR spectrum. The CR of the particular absorption feature was calculated by subtracting the band depth (BD) value at a particular wavelength (λ) from 1 (i.e. CR = BD (λ) − 1). Detrend and CR were performed in R programing language (R Core Team, 2016) by appling prospectr package (Stevens and Ramirez-Lopez, 2013). BR was determined by the differences between a pair of spectral features (e.g., first spectral feature divided by second, second divided by third and so on). BR was applied after spectral features selection by detrend (Det + BR) and CR preprocessing (CR + BR).

The selection of spectral features or spectral peaks were achieved by two techniques: hand selection (by observing the shapes, peaks, valleys of the preprocessed spectra with pedological knowledge) and mathematical selection (automated; computerized selection in MALDIquant R package (Gibb and Strimmer, 2012)). The criterion used to define the spectral features by hand selection came from the need to consider the entire region of the spectrum and to associate the specific spectral features with the soil characteristics. The hand selection technique elected spectral bands associating the iron oxide features at 412, 448, and 476 nm; the water, hydroxyl, and clay mineral absorption at 1400, 1900, and 2200 nm, respectively; additional features associated with organic matter around 750, 1650, 2200, 2400, 2350 nm were also considered.

The mathematical selection method looked for peaks in spectrometry data. A peak is a local maximum above a user defined noise threshold. The mathematical selection estimated and removed the baseline of spectrum by applying the 'SNIP' method. This baseline estimation is based on the 'Statistics-sensitive Non-linear Iterative Peak clipping' algorithm (SNIP) described in Ryan et al. (1988). This technique was applied by detectPeaks function in MALDIquant R package (Gibb and Strimmer, 2012). The whole spectra (entire region of Vis-NIR) of detrend and CR preprocessing were used as control treatment in the modeling process.

2.5. Statistical analysis

Descriptive statistics were calculated to summarize the data set, and the coefficient of variation (CV) provided the variation of the data. The descriptive statistic was performed in the R programming language. The Levene's test (Levene, 1960) (car R package (Fox and Weisberg, 2011)) was used to verify the assumption that variances are equal across training and validations groups with significance level of 5%. The independent *t*-test was used to determine whether a statistically significant difference exists between the means in the two unrelated groups (training and validation sets). The SVM regression analysis (e1071 R package (Meyer et al., 2017)) applied is a non-parametric statistical data mining method that belongs to the statistical learning theory (Ivanciuc, 2007). In SVM regression analysis, a training model of a sample set (training set) is performed. The procedure is to find a functional model that predicts correctly new cases that are not yet presented with SVM previously. SVM is a group of supervised learning methods that can be applied to classification or regression analysis, with several applications in many scientific areas (Ivanciuc, 2007). PLSR (pls R package (Mevik et al., 2016)) is a method that models linear relationships and is one of the most widely applied methods to predict soil properties from spectral data. PLSR is based on a projection of the predictor *x* and response *y* variables into a set of latent variables and corresponding scores, minimizing the dimensionality of the data while maximizing the covariance between *x* and *y* variables (Wold et al., 2001). To compare the modeling performance of both spectral features selection techniques, the Scott Knott test (5%) was applied. The whole spectra were used as control treatment. RMSE values were considered in order to verify the statistical difference of hand selection, mathematical selection techniques, and whole spectra. Scott Knott test was also applied in order to verify the statistical difference between preprocessing. Scott Knott test was carried out by ScottKnott R package (Jelihovschi et al., 2014).

2.6. Model training and validation

A total of 299 soil samples were randomly split into training set [~70%] (n = 209) and validation set [~30%] (n = 90). The fit and accuracy assessment of the models used the following validation parameters: Coefficient of determination (R^2), root mean square error of prediction (RMSE).

3. Results and discussion

3.1. Exploratory results

Considering the training and validation set, only clay showed a negatively skewed distribution, with means of 59.56% and 57.53%, respectively (Table 1). The minimum and maximum described the variation in the soil data sets. Generally, higher SOC values appeared in Inceptisols and lower values in Oxisols. In addition, the SOC decreased with increasing depth. The combination of high altitude and low temperature frequently promotes accumulation of carbon in these soils due to the low decomposition of organic matter. The clay content showed the lowest CV, which denotes that the variation from the mean indicates low data dispersion. SOC and silt content showed intermediate dispersion, and sand content exhibited extreme data dispersion (i.e., a high CV). The results of the predictive models confirm the same trend due to the CV in the descriptive statistics. The Levene's test indicated the homogeneity of variance between training and validation sets for SOC (*p*-value = 0.357), along with clay (*p*-value = 0.943), sand (*p*-value = 0.847), and silt (*p*-value = 0.452). Since *p*-values are much higher than the significance level of 5%, the variances have no significant difference. This similarity between sets indicates that the

Table 1
Descriptive statistics of soil properties for the training and validation.

	Training set (%)				Validation set (%)			
	SOC	Clay	Sand	Silt	SOC	Clay	Sand	Silt
Observations	209	209	209	209	90	90	90	90
Minimum	0.17	20.94	1.00	16.54	0.38	25.41	1.56	18.39
Maximum	4.83	78.48	35.48	77.99	4.21	75.85	32.15	72.94
1 st quartile	1.06	53.63	2.98	26.61	1.35	51.35	3.58	29.71
3rd quartile	2.46	68.28	9.95	38.06	2.66	66.22	8.43	39.74
Mean	1.84	59.56	7.51	32.94	2.04	57.53	7.70	34.77
Median	1.68	59.57	4.80	31.04	2.20	57.89	5.37	34.47
Std error of mean	0.07	0.77	0.46	0.67	0.10	1.17	0.76	0.92
Skewness	0.44	−0.65	1.80	1.47	0.00	−0.39	2.02	0.84
Kurtosis	−0.39	0.44	3.11	3.96	−0.74	−0.35	3.50	2.61
CV (%)	55	19	89	30	46	19	93	25

Table 2
Predictive performance of soil properties for the validation set.

Soil Property	Method	Preprocessing	Technique of spectral features selection	R^2_{val}	$RMSE_{val}$(%)*
SOC	PLSR	CR	W.S.	0.90	0.32
	SVM	CR	H.S.	0.87	0.35
	PLSR	Det	W.S.	0.86	0.36
	SVM	Det	W.S.	0.86	0.36
	SVM	CR	W.S.	0.86	0.36
	PLSR	CR	H.S.	0.83	0.41
	SVM	CR + BR	H.S.	0.81	0.42
	PLSR	Det + BR	H.S.	0.81	0.42
	SVM	Det	H.S.	0.81	0.42
	PLSR	Det	H.S.	0.81	0.42
	SVM	Det + BR	H.S.	0.81	0.43
	PLSR	CR + BR	H.S.	0.79	0.46
	SVM	Det + BR	M.	0.76	0.48
	SVM	Det	M.	0.73	0.50
	PLSR	Det + BR	M.	0.72	0.50
	PLSR	Det	M.	0.72	0.51
	SVM	CR + BR	M.	0.69	0.53
	PLSR	CR + BR	M.	0.69	0.54
	PLSR	CR	M.	0.68	0.54
	SVM	CR	M.	0.68	0.56
Clay	SVM	Det	W.S.	0.62	6.84
	SVM	CR	W.S.	0.58	7.18
	SVM	CR	H.S.	0.56	7.21
	SVM	CR + BR	H.S.	0.56	7.30
	PLSR	CR	H.S.	0.52	7.46
	SVM	CR	M.	0.52	7.70
	SVM	CR + BR	M.	0.52	8.03
	SVM	Det	H.S.	0.47	8.04
	SVM	Det + BR	H.S.	0.48	8.08
	SVM	Det	M.	0.47	8.08
	SVM	Det + BR	M.	0.44	8.31
	PLSR	CR + BR	H.S.	0.45	8.33
	PLSR	CR	M.	0.42	8.45
	PLSR	CR + BR	M.	0.42	8.47
	PLSR	Det + BR	H.S.	0.40	8.72
	PLSR	Det	W.S.	0.41	8.74
	PLSR	Det	H.S.	0.40	8.75
	PLSR	Det	M.	0.35	8.93
	PLSR	Det + BR	M.	0.35	8.94
	PLSR	CR	W.S.	0.42	8.96
Sand	PLSR	CR	W.S.	0.33	6.00
	PLSR	Det	W.S.	0.26	6.15
	SVM	CR + BR	H.S.	0.25	6.26
	SVM	CR	W.S.	0.25	6.28
	PLSR	Det + BR	H.S.	0.22	6.36
	SVM	Det	W.S.	0.25	6.41
	PLSR	Det	H.S.	0.19	6.45
	PLSR	CR + BR	H.S.	0.18	6.46
	PLSR	CR	H.S.	0.17	6.50
	SVM	CR + BR	M.	0.20	6.52
	PLSR	Det + BR	M.	0.16	6.57
	PLSR	CR + BR	M.	0.14	6.62
	PLSR	CR	M.	0.13	6.66
	PLSR	Det	M.	0.13	6.67
	SVM	CR	H.S.	0.17	6.68
	SVM	CR	M.	0.14	6.70
	SVM	Det	H.S.	0.16	6.79
	SVM	Det + BR	H.S.	0.16	6.81
	SVM	Det	M.	0.13	6.93
	SVM	Det + BR	M.	0.13	6.97
Silt	PLSR	CR	H.S.	0.56	5.26
	SVM	CR	H.S.	0.57	5.35
	SVM	CR + BR	H.S.	0.54	6.06
	SVM	Det	W.S.	0.50	6.17
	SVM	CR	W.S.	0.50	6.20
	PLSR	Det + BR	H.S.	0.46	6.51
	SVM	Det + BR	H.S.	0.45	6.54
	SVM	Det	H.S.	0.44	6.54
	PLSR	CR + BR	H.S.	0.44	6.67
	PLSR	Det	H.S.	0.44	6.71
	SVM	CR + BR	M.	0.40	6.82
	SVM	CR	M.	0.39	6.92
	PLSR	CR	M.	0.34	7.05
	PLSR	CR + BR	M.	0.32	7.16
	PLSR	Det	W.S.	0.41	7.23
	PLSR	Det + BR	M.	0.31	7.23
	SVM	Det + BR	M.	0.32	7.28
	SVM	Det	M.	0.31	7.41
	PLSR	Det	M.	0.28	7.46
	PLSR	CR	W.S.	0.40	7.67

*Sorted by ascending order of $RMSE_{val}$ (root mean square error of prediction for validation set). M: mathematical selection, H.S.: hand selection, W.S: whole spectra, CR: continuum removal, Det: detrend, BR: band ratio, PLSR: Partial least square regression, SVM: Support vector machine.

random split represents the study population.

3.2. Predictive performance of PLSR and SVM

The predictive statistics of all models for the soil properties are shown in Table 2. In this table, the models results are placed in ascending order of RMSE. SOC content showed high accuracy, indicating a strong linear relationship between the measured and predicted variables. The models of SOC prediction showed a R^2_{val} and $RMSE_{val}$ ranging from 0.68 and 0.56% to 0.90 and 0.32%, respectively. The greater predictive performance was achieved by PLSR with CR preprocessing using the whole spectra. Among the 20 SOC predictive models, 11 presented an R^2_{val} higher than 0.81. The statistical difference between the prediction results of PLSR and SVM are revealed in Table 3. The Scott Knott test (5%) presented the mean comparison test of RMSE values for both methods. This test showed that there was no statistical difference between the RMSE values of PLSR and SVM models for SOC prediction. The mean values of RMSE are practically identical: 0.45% and 0.44%, for PLSR and SVM, respectively. This result demonstrated that both multivariate methods are suitable for SOC prediction. On the other hand, there is no right number of spectral features to estimate soil properties because each soil has a particular spectral reflectance signature and thereby distinct spectral features will be selected in model building.

These results are comparable to studies in the literature. Stevens et al. (2010) applied SVM to predict SOC in Luxembourg using different soil types (clay, silty–clay, silt, sandy–loam, and sand), and their validation results were slightly higher (R^2 = 0.84), but with an identical RMSE (0.43%). In Australia, a study presented by Viscarra Rossel and Behrens (2010) showed that the SVM produced the highest fitted model (R^2 = 0.84) and lowest error (RMSE = 0.92%) for SOC estimation. The similar performance of the SVM model may be

Table 3
Statistical difference between the prediction results of PLSR and SVM methods for each soil property.

	Method	Mean of $RMSE_{val}$(%)*	Scott Knott test (5%)
SOC	SVM*	0.44	a
	PLSR*	0.45	a
Clay	SVM	7.68	a
	PLSR	8.58	b
Sand	PLSR	6.44	a
	SVM	6.64	a
Silt	SVM	6.53	a
	PLSR	6.90	a

*PLSR: Partial Least Square Regression, SVM: Support Vector Machine. $RMSE_{val}$: root mean square error of prediction for validation set.

attributed to the similarity of the sample observations used in their study with a total of 302 (201 for training and 101 for validation). Chacón Iznaga et al. (2014) used SVM to predict organic matter within a field in the central region of Cuba and found high R^2 = 0.92 and RMSE = 0.14%. The performance in the current study showed that SOC can be properly estimated by using supervised learning models. In Ramirez-Lopez et al. (2013), the SVM prediction results for modeling organic carbon using Vis-NIR spectra (not continuum removed reflectance) were moderate (R^2_{val} = 0.54) for a regional soil spectral library with a low $RMSE_{val}$ = 0.27% when compared to the results in this study. Large datasets have typically larger variances; therefore, well-performing models are more difficult to develop. According to Guerrero et al. (2015), small, rather than large, spectral libraries for local scale SOC assessment provide accurate predictions for effective model performance. Steffens and Buddenbaum (2013) presented SVM models that produced results for a concentration of SOC with R^2 = 0.97 and RMSE = 1.13% to provide laboratory imaging spectroscopy of soil profiles from Munich, Germany.

The predictive performance of clay content presented a R^2_{val} and $RMSE_{val}$ ranging from 0.42 and 8.96% to 0.62 and 6.84%, respectively (Table 2). The best model was achieved by SVM with detrend preprocessing using the whole spectra. Scott Knott test showed that there was statistical difference between the RMSE values of PLSR and SVM models for clay prediction (Table 3). Clay content was the only soil property where the performances of PLSR and SVM presented statistical difference. SVM presented higher predictive performance for clay compared to PLSR. The mean value of RMSE for SVM and PLSR models was 7.68% and 8.58%, respectively. Higher performances to predict clay content using SVM were achieved by Viscarra Rossel and Behrens (2010) (R^2 = 0.84, RMSE = 7.63%). This achievement was attributed to the substantially larger soil sample sets located in different regions in Australia, including a diverse number of soil classes (n = 1104), which was three times larger compared to the present study. In addition, Kovačević et al. (2010) achieved high-quality results by applying SVM to predict clay content in eastern Serbia (R^2 = 0.76 and normalized root mean squared deviation = 0.11%), although with a small data set (n = 151). Terra et al. (2015) used Vis-NIR reflectance and SVM to predict various soil properties in the Midwest and Southeast regions of Brazil, such as particle size, chemical properties that include macro and micronutrients, and iron oxides. The authors achieved high-quality predictions for clay (R^2_{val} = 0.86, $RMSE_{val}$ = 95.34 g kg^{-1}) and sand contents (R^2_{val} = 0.86, $RMSE_{val}$ = 22.16 g kg^{-1}). These high-quality results are associated with the correlation among clay activity and other soil properties. According to Stevens et al. (2013), variations in clay content induce large differences in the spectral shape with non-variation of SOC content.

The lowest predictive performance was achieved for sand content. The inferior model result showed a R^2_{val} of 0.13 and $RMSE_{val}$ of 6.97% while the superior showed a R^2_{val} of 0.33 and $RMSE_{val}$ of 6.00%, which can be considered a low prediction (Table 2). Scott Knott test showed that there was no statistical difference between the RMSE values of PLSR and SVM models for sand prediction (Table 3). The mean values of RMSE were 6.44% and 6.64%, for PLSR and SVM, respectively. The R^2 had the lowest value among all four modeled soil properties. This result may have occurred due to the soil classes being mostly composed of Oxisols, which has a relatively low sand fraction (Table 1). Kovačević et al. (2010) applied the SVM to estimate soil properties in eastern Serbia and the performance for sand content was greater compared to this study (R^2 = 0.59), with a normalized root mean squared deviation of 0.14%. The high CV value for sand may also explain the relatively large uncertainty in the prediction of sand content.

The predictive performance of silt content was considered moderate with a R^2_{val} and $RMSE_{val}$ ranging from 0.40 and 7.67% to 0.56 and 5.26%, respectively (Table 2). The higher model was found applying PLSR with CR preprocessing using spectral features by hand selection. In the Scott Knott test (Table 3) there was no statistical difference between the RMSE values of PLSR and SVM models for silt prediction. The mean values of RMSE were 6.53% and 6.90%, for SVM and PLSR, respectively. There are some caveats to silt content, which is not directly measured by the pipette method, and occasionally, the silt value adds up the clay and sand error measurement.

The distinctive parent material found in the area of study (sedimentary and basalt rocks) may have affected the performance for clay, sand, and silt. Because of the increased iron oxide concentration at sites characterized as Oxisols, the depth of absorption from 390 to 550 nm also increased (Ben-Dor, 2002; Summers et al., 2011). The influence of iron oxide on the reflectance spectra in the visible spectral region may have masked or decreased the inference of some soil properties, such as particle size content.

The SVM acceptance in soil properties estimation has increased in recent years (Araújo et al., 2014; Ramirez-Lopez et al., 2013; Terra et al., 2015) and has generated more accurate calibration results than PLSR in some studies (Thissen et al., 2004; Viscarra Rossel et al., 2010). In Nawar et al., 2016, the results for PLSR with different preprocessing transformation showed a low R^2 between $0.33 \leq 0.52$ (RMSE $0.42\% \geq 0.36\%$) for organic matter. In this same study, for clay content the R^2 fluctuated between 0.14 to 0.82. On the other hand, PLSR can also provide satisfactory results. Kuang et al. (2015), compared the performance of PLSR prediction models for SOC and clay content and found: $R^2 \leq 0.81$ and RMSE $\geq$ 1.46% for SOC, and $R^2 \leq 0.81$ and RMSE $\geq$ 1.04% for clay.

For quite a long time, the most widely used regression method applied to predict soil properties from spectral data was PLSR. Wold et al. (2001) drew our attention to PLSR in handling numerous and collinear variables and to investigate more compounded problems. However, PLSR models are not designed for the complexity of chemical and biological systems. They are also not often used to screen out latent variables that are not useful in explaining the response. In Gomez et al. (2008), PLSR showed better performance when there was no well-identified spectral feature for the property of interest (clay and calcium carbonate).

3.3. Performance of spectral feature selection techniques

The two techniques of spectral feature selection, represented by hand selection and mathematical selection, were analyzed by its potential to reduce the covariables for modeling procedure. In detrend preprocessing, 13 spectral features were selected by hand selection and only 8 by mathematical selection (Fig. 1). On the other hand, in CR preprocessing, 11 spectral features were selected by hand selection and by mathematical selection (Fig. 2). For detrend preprocessing the mathematical selection reduced 5 spectral features and for CR preprocessing the number of spectral selected were identical.

For SOC prediction, the results of RMSE values showed statistical difference between spectral feature selection techniques and whole spectra (Fig. 3). The models applying the whole spectra achieved the best performance in SOC prediction with a mean RMSE value of 0.35%. Among hand and mathematical, the first selection presented a mean of RMSE value of 0.42%, and the second a value of 0.52%. The results of RMSE values for clay content were statistically identical regardless the spectral feature selection or whole spectra were applied. The prediction models using whole spectra presented a mean RMSE value of 7.93% (Fig. 3). Hand selection showed best performance compared to mathematical selection for clay prediction with a mean RMSE value of 7.99% and 8.36%, respectively. The results of sand content showed that the RMSE value for whole spectra was statistically different from hand and mathematical selection techniques. The mean RMSE values were 6.21%, 6.54% and 6.70% for whole spectra, hand and mathematical selection, respectively (Fig. 3). For silt content, hand selection presented statistical difference in RMSE value from whole spectra and mathematical selection technique. The silt content was the only soil property where the hand selection presented superior RMSE results.

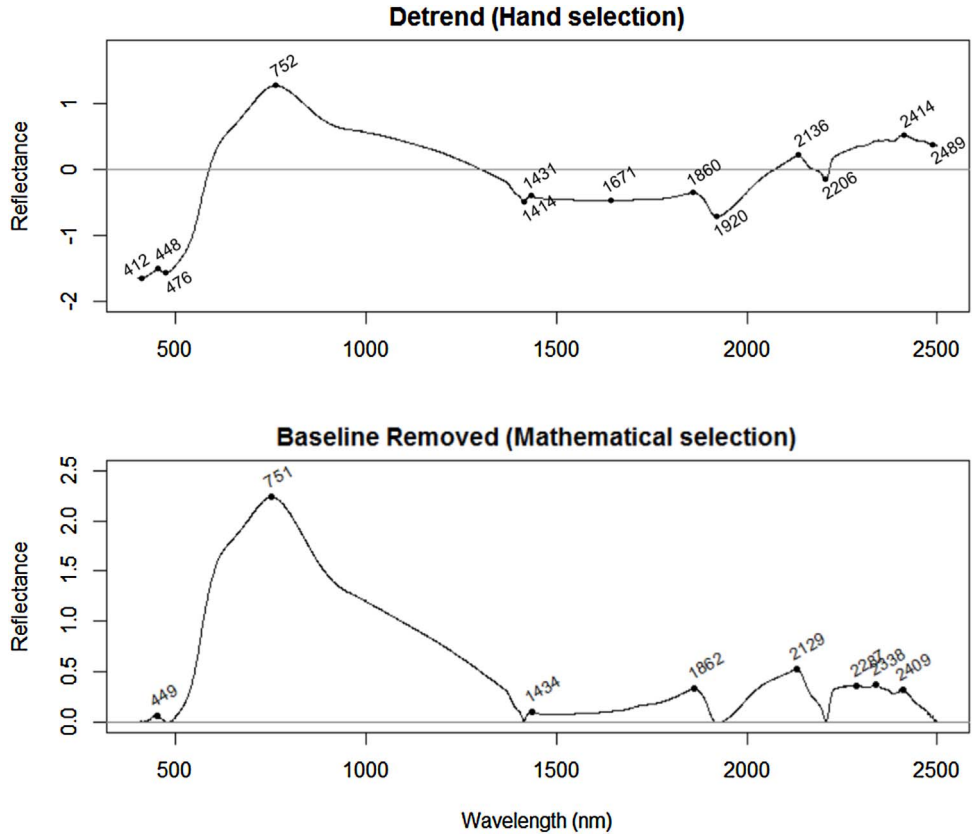

Fig. 1. Spectral curves of detrending transformation and its baseline removed in the visible-near infrared spectrum (average of 299 soil samples). Hand selection has 13 spectral features and mathematical selection has 8 spectral features.

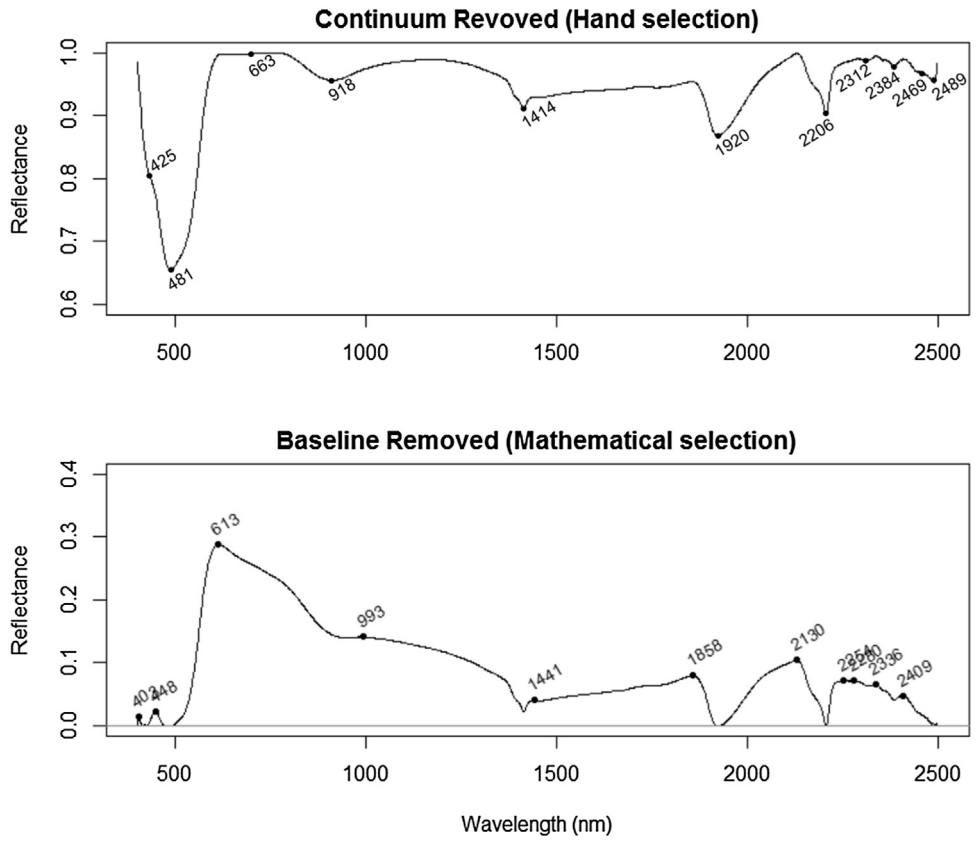

Fig. 2. Spectral curves of continuum removed preprocessing and its baseline removed in the visible-near infrared spectrum (average of 299 soil samples). Hand selection and mathematical selection have 11 spectral features.

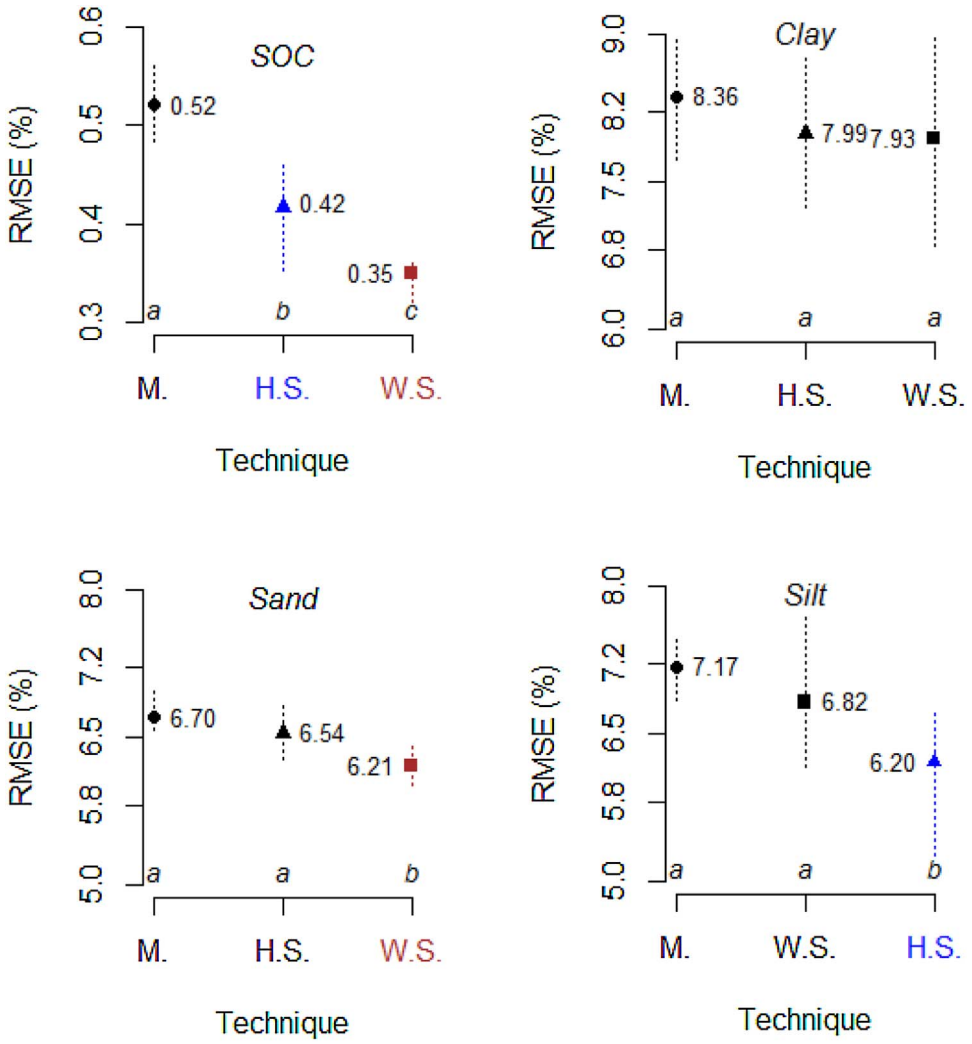

Fig. 3. Statistical difference between spectral feature selection techniques. In the graphics are the mean, maximum and minimum values of RMSE. Letters represent the results of Scott Knott test (significance level of 5%). M: mathematical selection, H.S.: hand selected, W.S: whole spectra.

The mean RMSE value of hand selection, whole spectra and mathematical selection was 6.20%, 6.82% and 7.17%, respectively (Fig. 3).

The models using all Vis-NIR spectral region (whole spectra) showed superior performance for SOC, clay and silt content. The reduction of spectral features revealed that the predictive performances of all soil properties were greater for hand selection technique. The mathematical selection, in which the features were selected by automated approach presented poor prediction for all soil properties. This is because mathematical selection does not take into consideration the preeminent spectral features to predict soil properties.

Selecting the spectral features by observing the shapes of the preprocessed spectra with pedological knowledge led to better prediction results. The reduction of spectral features in the Vis-NIR spectrum by hand selection technique increased the predictive performance of models by choosing spectral regions that are associated with specific soil characteristics.

Important spectral features chosen by hand selection were located in the near infrared region. Generally, the spectral features are linked with important spectral active soil components, for example, mineralogy, texture, and iron content (Stevens et al., 2013). Furthermore, the two spectral selection techniques shared several wavelengths, particularly near 1400 nm and 1900 to 2400 nm, which confirmed that these wavelengths in the near infrared spectral region provide valuable contribution for soil property estimations.

The spectral features selected by hand selection technique had a considerably higher estimation performance of SOC content compared to textural properties. Many of the spectral features were likewise selected in the present study in accordance with the features of specific soil constituents documented in the literature. The spectral features at 1414 nm and 1920 nm were related to the vibration activity of the hydroxyl group in water molecules (Ben-Dor, 2002). These features may be indicative of the insufficient air-drying in green houses. According to Ben-Dor (2002), the spectral regions of 1300–1450 nm, 1850–1950 nm, and 2200–2400 nm are linked to clay minerals. According to Chang et al. (2001), these are the most predominant spectral bands to predict clay content. However, the equivalent spectral features selected for both clay and SOC could have under-fitted the model estimation for clay since the SOC could have masked or diminished the clay content. Xie et al. (2012) presented five wavelength ranges that had major contributions to predict organic matter in the NIR region: 1386–1401 nm, 2133–2138 nm, 2175–2194 nm, 2229–2273 nm, and 2315–2327 nm. In addition, soil organic matter also showed correlation bands in the visible region (400–750 nm) (Stenberg et al., 2010), where a total of seven spectral features were selected. Some auxiliary spectral features at near infrared were also selected by hand selection.

In hyperspectral data, reduction techniques have promulgated to filter out the most important features. However, the least number of spectral features have affected model performance. This can be credited to great capacity of multivariate methods, such as PLSR and SVM, in estimating attributes based on the spectral behavior. In the study of

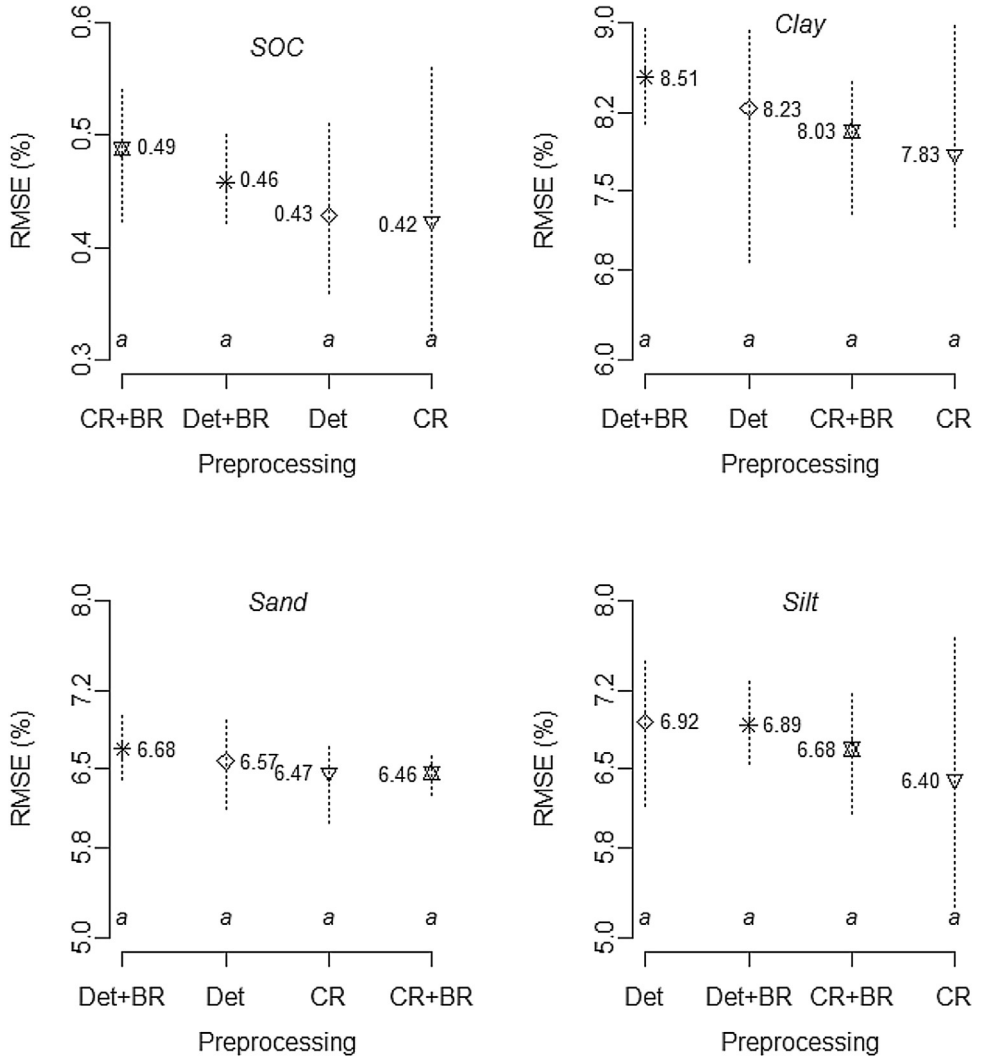

Fig. 4. Statistical difference between preprocessing techniques. In the graphics are the mean, maximum and minimum values of RMSE. Letters represent the results of Scott Knott test (significance level of 5%). CR: continuum removal, Det: detrend, BR: band ratio.

Üstün (2003), SVM outperformed PLSR if there is no wavelength selection applied. For Üstün (2003), SVM has some advantages in comparison with PLSR: i) it finds a general solution and thus avoids overtraining; ii) it gives a solution which is sparse and; iii) it is able to model non-linear relations. However, SVM also has a disadvantage such as high computation time in case of a large data set, which leads to a time-consuming optimization.

3.4. Performance of preprocessing techniques

The RMSE result for each soil property revealed that there was no statistical difference between the four preprocessing techniques applied (Fig. 4). However, CR preprocessing yielded the lowest RMSE results for SOC, clay and silt content. CR was the most reliable preprocessing method for estimating the soil properties, and overall, provided better estimations than detrend preprocessing.

Recent worldwide publications are targeting the CR as a preprocessing technique to estimate soil properties, especially for SOC. The content of SOC had a huge impact on CR absorption feature since soils with high SOC indicate a decrease in albedo across the entire Vis-NIR spectrum (Ben-Dor, 2002). Stenberg (2010) used CR to examine the effect of soil moisture content on Vis-NIR spectra. The results revealed that the CR technique was effective in distinguishing wet and dry soils. In addition, dry soils resulted in deeper absorption features along with high amounts of clay. The CR approach presents the advantage of addressing specific absorptions features as covariables derived from reflectance measurements. Furthermore, preprocessing contributed in the reduction of multicollinearity; otherwise, the variance of the coefficients may be very large and the model might apply unnecessary information. The results in the present study confirmed that CR preprocessing contributed in selecting the most significant features to estimate the soil properties. Nawar et al. (2016) revealed that for SOC and clay the best predictive results were found by applying continuum removal preprocessing transformation. The appropriate selection of explanatory variables (spectral features) in the CR preprocessing was essential to improve the modeling performance and reduce the complexity of the models.

4. Conclusions

Overall, the prediction of SOC, clay, sand and silt presented suitable results using reduced spectral features as covariables in modeling process. SOC presented a high–level prediction model. The results for clay and silt content showed moderate performances, as opposed to sand content, which showed inferior performance. The hand selection technique showed superior performance in predicting soil properties due to the pedological knowledge, which can associate the spectral features with specific soil characteristics. The predictive performances of PLSR and SVM multivariate methods showed that there was no statistical difference between the RMSE results, except for clay content,

where SVM presented superior performance. These was no statistical difference between preprocessing techniques in predicting SOC, clay, sand, and silt. However, CR preprocessing presented the lowest RMSE results compared to detrend, CR + BR and detrend + BR.

The main strength of spectral feature selection techniques was their effectiveness in reducing predictor covariables, which enhances interpretability and transparency of models. Both techniques contributed by highlighting the bands and features produced by optically active soil components. The selection of spectral features from entire spectra region accomplished reliable outcomes. This study confirmed the high potential of using spectral preprocessing techniques to estimate soil properties and examining the metrological quality of soil properties from Vis-NIR spectral data. The predictive model performances are influenced by the multivariate method, spectral preprocessing, homogeneity of soil samples, and type of estimated soil properties. The authors suggest that, selecting spectral features is an imminent choice for developing prediction models in upcoming studies.

Further studies have to consider that there is no optimal or 'best' amount of spectral features to estimate soil properties because each soil has distinct spectral reflectance signatures. These alternatives for spectral features selection accentuated soil features and detected patterns of individual soil spectrum. Modeling strategies that differ in their capabilities to extract pedological characteristics from the Vis-NIR spectra need to be carefully considered in future studies.

Acknowledgements

The first author would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), and the Santa Catarina Research Foundation (FAPESC), (Project no. 2012000094) for providing scholarship and funding to carry out this research. The authors would also like to thank the Geomatics Laboratory at the Federal University of Santa Catarina for collecting the soil samples, the Laboratory of Pedology at Federal University of Santa Maria for their support in the laboratory analysis, the Geotechnologies in Soil Science Group at the Soil Science Department, ESALQ − University of Sao Paulo for their support in the spectroscopy analysis.

References

Üstün, B., 2003. A Comparison of Support Vector Machines and Partial Least Squares Regression on Spectral Data. Department Anal. Chem Master Thesis.

Andrews, S.S., Karlen, D.L., Cambardella, C.A., 2004. The soil management assessment framework: a quantitative soil quality evaluation method. Soil Sci Soc Am J. Soil Sci. Soc. Am. J. 68, 1945–1962.

Araújo, S.R., Wetterlind, J., Demattê, J. a. M., Stenberg, B., 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. Eur. J. Soil Sci. 65, 718–729. http://dx.doi.org/10.1111/ejss.12165.

Arrouays, D., McKenzie, N., Hempel, J., de Forges, A.C.R., McBratney, A.B. (Eds.), 2014. GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRC Press/Balkema.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-Infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777.

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500 nm) during a controlled decomposition process. Remote Sens. Environ. 61, 1–15. http://dx.doi.org/10.1016/S0034-4257(96)00120-4.

Ben-Dor, E., 2002. Quantitative remote sensing of soil properties. Advances in Agronomy 75, 173–243. http://dx.doi.org/10.1016/S0065-2113(02)75005-0.

Chacón Iznaga, A., Rodríguez Orozco, M., Aguila Alcantara, E., Carral Pairol, M., Díaz Sicilia, Y.E., de Baerdemaeker, J., Saeys, W., 2014. Vis/NIR spectroscopic measurement of selected soil fertility parameters of Cuban agricultural Cambisols. Biosyst. Eng. 125, 105–121. http://dx.doi.org/10.1016/j.biosystemseng.2014.06.018.

Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-Infrared reflectance Spectroscopy?Principal components regression analyses of soil properties. Soil Sci. Soc. Am. J 65, 480–490. http://dx.doi.org/10.2136/sssaj2001.652480x.

Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. J. Geophys. Res. Solid Earth 89, 6329–6340. http://dx.doi.org/10.1029/JB089iB07p06329.

Conforti, M., Castrignanò, A., Robustelli, G., Scarciglia, F., Stelluti, M., Buttafuoco, G., 2015. Laboratory-based Vis?NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content. CATENA 124, 60–67. http://dx.doi.org/10.1016/j.catena.2014.09.004.

Dalmolin, R.S.D., Gonçalves, C.N., Klamt, E., Dick, D.P., 2005. Relationship between the soil constituents and its spectral behavior. Ciênc. Rural 35, 481–489. http://dx.doi.org/10.1590/S0103-84782005000200042.

Donagemma, G.K., Campos, D.V.B., de, Calderano, S.B., Teixeira, W.G., Viana, J.H.M., 2011. Manual De Métodos De Análise De Solo, edition 2 rev. pp. 230.

Fox, J., Weisberg, S., 2011. An R Companion to Applied Regression, Second ed. SAGE Publications, Thousand Oaks, CA.

Ge, Y., Thomasson, J.A., Sui, R., 2011. Remote sensing of soil properties in precision agriculture: a review. Front. Earth Sci. 5, 229–238. http://dx.doi.org/10.1007/s11707-011-0175-0.

Gibb, S., Strimmer, K., 2012. MALDIquant: a versatile R package for the analysis of mass spectrometry data. Bioinf. 28, 2270–2271. http://dx.doi.org/10.1093/bioinformatics/bts447.

Gomez, C., Lagacherie, P., Coulouma, G., 2008. Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. Geoderma 148, 141–148. http://dx.doi.org/10.1016/j.geoderma.2008.09.016.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., Viscarra Rossel, R.A., 2015. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? Soil Tillage Res. http://dx.doi.org/10.1016/j.still.2015.07.008.

Hartemink, A.E., Minasny, B., 2014. Towards digital soil morphometrics. Geoderma 230 (-231), 305–317. http://dx.doi.org/10.1016/j.geoderma.2014.03.008.

Ivanciuc, O., 2007. Applications of support vector machines in chemistry. In: Lipkowitz, K.B., Cundari, T.R. (Eds.), Reviews in Computational Chemistry. John Wiley & Sons Inc., pp. 291–400.

Janzen, H.H., 2004. Carbon cycling in earth systems—a soil science perspective. Agric. Ecosyst. Environ. 104, 399–417. http://dx.doi.org/10.1016/j.agee.2004.01.040.

Jelihovschi, E.G., Faria, J.C., Allaman, I.B., 2014. ScottKnott: A Package for Performing the Scott-Knott Clustering Algorithm in R. Trends Appl. Comput. Math. 15, 3–17. http://dx.doi.org/10.5540/tema.2014.015.01.0003.

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B., Harris, W.G., 2015. Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. Geoderma 239 (-240), 229–239. http://dx.doi.org/10.1016/j.geoderma.2014.10.019.

Koch, A., McBratney, A., Adams, M., Field, D., Hill, R., Crawford, J., Minasny, B., Lal, R., Abbott, L., O'Donnell, A., Angers, D., Baldock, J., Barbier, E., Binkley, D., Parton, W., Wall, D.H., Bird, M., Bouma, J., Chenu, C., Flora, C.B., Goulding, K., Grunwald, S., Hempel, J., Jastrow, J., Lehmann, J., Lorenz, K., Morgan, C.L., Rice, C.W., Whitehead, D., Young, I., Zimmermann, M., 2013. Soil security: solving the global soil crisis. Glob. Policy 4, 434–441. http://dx.doi.org/10.1111/1758-5899.12096.

Kovačević, M., Bajat, B., Gajić, B., 2010. Soil type classification and estimation of soil properties using support vector machines. Geoderma 154, 340–347. http://dx.doi.org/10.1016/j.geoderma.2009.11.005.

Kuang, B., Tekin, Y., Mouazen, A.M., 2015. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. Soil Tillage Res. 146, 243–252. http://dx.doi.org/10.1016/j.still.2014.11.002.

Lal, R., Moldenhauer, W.C., 1987. Effects of soil erosion on crop productivity. Crit. Rev. Plant Sci. 5, 303–367. http://dx.doi.org/10.1080/07352688709382244.

Levene, H., 1960. Robust tests for equality of variances. Robust Tests Equal. Var. 278–292.

McBratney, A.B., Minasny, B., Viscarra Rossel, R., 2006. Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. Geoderma 136, 272–278. http://dx.doi.org/10.1016/j.geoderma.2006.03.051.

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. http://dx.doi.org/10.1016/j.geoderma.2013.08.013.

Mevik, B.-H., Wehrens, R., Liland, K.H., 2016. pls: Partial Least Squares and Principal Component Regression. R Package Version 2 .6-0.

Meyer, D., Dimitriadou, E., Hornik, K., Leisch, F., Weingessel, A., 2017. E1071: Misc Functions of the Department of Statistics (E1071). TU Wien. R Package Version 1 .6-8.

Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemom. Intell. Lab. Syst. 94, 72–79. http://dx.doi.org/10.1016/j.chemolab.2008.06.003.

Mutanga, O.M.C., Skidmore, A.K., Kumar, L., Ferwerda, J., 2005. Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. Int. J. Remote Sens. 26, 1093–1108. http://dx.doi.org/10.1080/01431160512331326738.

Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. Soil Tillage Res. 155, 510–522. http://dx.doi.org/10.1016/j.still.2015.07.021.

R Core Team, 2016. R: A Language and Environment for Statistical Computing.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: a new local approach for modeling soil vis?NIR spectra of complex datasets. Geoderma 195-196, 268–279. http://dx.doi.org/10.1016/j.geoderma.2012.12.014.

Ryan, C., Clayton, E., Griffin, W.L., Cousens, D.R., 1988. SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At. 34,

396–402.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639. http://dx.doi.org/10.1021/ac60214a047.
Steffens, M., Buddenbaum, H., 2013. Laboratory imaging spectroscopy of a stagnic Luvisol profile — High resolution soil characterisation, classification and mapping of elemental concentrations. Geoderma 195 (-196), 122–132. http://dx.doi.org/10.1016/j.geoderma.2012.11.011.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Chapter five – visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 163–215.
Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma 158, 15–22. http://dx.doi.org/10.1016/j.geoderma.2010.04.008.
Stevens, A., Ramirez-Lopez, L., 2013. An introduction to the prospectr package. R Package Vignette.
Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma 158, 32–45. http://dx.doi.org/10.1016/j.geoderma.2009.11.032.
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, e66409. http://dx.doi.org/10.1371/journal.pone.0066409.
Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S.Y., Rawlins, B.G., Field, D.J., 2015. Global soil organic carbon assessment. Glob. Food Secur. 6, 9–16. http://dx.doi.org/10.1016/j.gfs.2015.07.001.
Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. Ecol. Indic. 11, 123–131. http://dx.doi.org/10.1016/j.ecolind.2009.05.001.
Terra, F.S., Demattê, J.A.M., Viscarra Rossel, R.A., 2015. Spectral libraries for quantitative analyses of tropical Brazilian soils: comparing vis?NIR and mid-IR reflectance data. Geoderma 255-256, 81–93. http://dx.doi.org/10.1016/j.geoderma.2015.04.017.
Thissen, U., Pepers, M., Üstün, B., Melssen, W.J., Buydens, L.M.C., 2004. Comparing support vector machines to PLS for spectral regression applications. Chemom. Intell. Lab. Syst. 73, 169–179. http://dx.doi.org/10.1016/j.chemolab.2004.01.002.
Vasques, G.M., Grunwald, S., Sickman, J.O., 2008. Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146, 14–25. http://dx.doi.org/10.1016/j.geoderma.2008.04.007.
Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. http://dx.doi.org/10.1016/j.geoderma.2009.12.025.
Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75. http://dx.doi.org/10.1016/j.geoderma.2005.03.007.
Wight, J.P., Ashworth, A.J., Allen, F.L., 2016. Organic substrate, clay type, texture, and water influence on NIR carbon measurements. Geoderma 261, 36–43. http://dx.doi.org/10.1016/j.geoderma.2015.06.021.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. chemom. intell. lab. syst. PLS Methods 58, 109–130. http://dx.doi.org/10.1016/S0169-7439(01)00155-1.
Xie, X.-L., Pan, X.-Z., Sun, B., 2012. Visible and near-Infrared diffuse reflectance spectroscopy for prediction of soil properties near a copper smelter. Pedosphere 22, 351–366. http://dx.doi.org/10.1016/S1002-0160(12)60022-8.
Yeomans, J.C., Bremner, J.M., 1988. A rapid and precise method for routine determination of organic carbon in soil. Commun. Soil Sci. Plant Anal. 19, 1467–1476. http://dx.doi.org/10.1080/00103628809368027.