Soil and Tillage Research, Volume 155, 2016, Pages 510-522
DOI:10.1016/j.still.2015.07.021

Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy

Said Nawar[1, 2]*, Henning Buddenbaum [3], Joachim Hill [3], Jacek Kozak[1], Abdul M. Mouazen[4]

[1] Institute of Geography and Spatial Management, Jagiellonian University, Krakow, 30-387, Poland

[2] Faculty of Agriculture, Suez Canal University, Ismailia, 41522, Egypt

[3] Environmental Remote Sensing and Geoinformatics, Trier University, 54286 Trier, Germany;

[4] Cranfield Soil and AgriFood Institute, Cranfield University, Bedfordshire MK43 0AL, UK

*correspondence; E-Mail: snawar@gis.geo.uj.edu.pl;
Tel.: +48797941230; Fax: +48126645385.

Abstract

The selection of calibration method is one of the main factors influencing measurement accuracy of soil properties estimation in visible and near infrared reflectance spectroscopy. In this study, the performance of three regression techniques, namely, partial least-squares regression (PLSR), support vector regression (SVR), and multivariate adaptive regression splines (MARS) were compared to identify the best method to assess organic matter (OM) and clay content in the salt-affected soils. One hundred and two soil samples collected from Northern Sinai, Egypt, were used as the data set for the calibration and validation procedures. The dry samples were scanned using a FieldSpec Pro FR Portable Spectroradiometer (Analytical Spectral Devices, ASD) with a measurement range of 350–2500 nm. The spectra were subjected to seven pre-processed techniques, e.g., Savitzky–Golay (SG) smoothing, first derivative with SG smoothing, second derivative with SG smoothing, continuum removed reflectance, standard normal variate and detrending (SNV-DT), multiplicative scatter correction (MSC) and extended MSC. The results of cross-validation showed that in most cases MARS models performed better than PLSR and SVR

models. The best predictions were obtained using MARS calibration methods with CR pre-processing, yielding R^2, root mean squared error, and ratio of performance to deviation values of 0.85, 0.19 %, and 2.63, respectively, for OM; and 0.90, 5.32 %, and 3.15, respectively, for clay content.

Keywords: Organic matter; clay content; reflectance spectroscopy; PLSR; SVR; MARS.

1. Introduction

Diffuse reflectance spectroscopy is considered as a promising method for the estimation of soil properties and a low cost alternative to traditional soil analysis methods because of its efficiency. Furthermore, determination of soil organic matter (OM) and clay content is complex and requires more chemical reagents and caution especially in salt-affected soils (Richards, 1954; Jackson, 1973). Whereas, visible and near infrared reflectance spectroscopy (VNIRRS) requires little sample preparation and provides accurate estimation of several soil properties from one spectral reading (Viscarra Rossel et al., 2006; Vohland et al., 2011). VNIRRS reflects the compositional and structural information of molecules at spectral wavelengths of 350–2500 nm. The analytical ability of VNIRRS depends on the repetitive and broad absorption of VNIRRS light by C–H, O–H and N–H bonds (Li et al., 2008; Xuemei and Jianshe, 2013), determined as overtones and combinations in the NIR range.

Different spectral pre-processing methods have been used to transform soil spectra, remove noise due to light scattering, emphasise features, and extract useful data for quantitative predictive models. Spectral pre-processing methods represent an important phase in multivariate calibration (Buddenbaum and Steffens, 2012) and have enhanced the accuracy of predictive models (Bilgili et al., 2010; 2011; Mashimbye et al., 2012). Typical spectral pre-processing includes smoothing, normalisation, scatter correction, continuum removal, and derivative

algorithms. For instance, first- and second-derivative absorption spectra have been used by Ben-Dor et al. (1997) to enhance spectral information and demonstrate spectral changes during a decomposition process. Stevens et al. (2010) used derivatives and Savitzky–Golay (SG) smoothing to enhance the performance of their models to predict soil organic carbon level. Rinnan et al. (2009) reviewed the most common spectral pre-processing methods for near-infrared spectroscopy. They concluded that, the optimal pre-processing method for a spectral data set in a calibration model is the signal transformation that gives the best regression performance after rigorous and comprehensive validation for all possible variation in the data.

The prediction mechanisms that may be used in a spectral calibration model depend on the corresponding spectral interaction of the predominant soil chromophores (Vohland et al., 2011). Therefore, selection of the calibration method and its performance in modelling reflectance spectra is one of the main factors for calibration success (Mouazen et al., 2010). Several calibration methods based on VNIRRS have been used to predict soil properties, including organic matter (OM) and clay content, among them partial least-squares regression (PLSR) is the standard method used for its simplicity and robustness (Farifteh et al., 2007; Vasques et al., 2008). The PLSR approach has inference capabilities, which are useful to model a probable linear relationship between the measured reflectance spectra and soil attributes (Farifteh et al., 2007; Viscarra Rossel and Behrens, 2010). Further, this approach enables the modelling of several response variables simultaneously, while effectively addressing strongly collinear and noisy predictor variables (Wold et al., 2001). However, the complex and non-linear reflectance behaviour in soils suggests that the use of linear models such as PLSR may be insufficient (Vohland et al., 2011).

Compared to linear and non-parametric regression methods, multivariate adaptive regression splines (MARS), a nonparametric method that estimates complex nonlinear relationships among

72 independent and dependent variables (Friedman, 1991), has been effectively applied in different
73 fields (Luoto and Hjort, 2005; Bilgili et al., 2010; Felicísimo et al., 2012; Samui, 2012) and
74 generally exhibits higher performance results. Bilgili et al. (2010) used MARS to model soil
75 properties, including soil OM and clay content, and reported that MARS provided better soil
76 property estimations than those by PLSR method. The support vector regression (SVR), a
77 nonlinear method based on the statistical learning theory, which was proposed by Vapnik (1998),
78 is capable of using fewer training data for learning in high-dimensional feature space and
79 employs a set of linear equations to obtain the support vectors (Xuemei and Jianshe, 2013). In
80 addition, SVR is capable of modelling linear and nonlinear relationships and solving calibration
81 problems with high performance (Suykens and Vandewalle, 1999). Recently, Viscarra Rossel and
82 Behrens (2010) successfully used SVR for modelling soil OM and clay content based on
83 reflectance spectroscopy. SVR has gained extensive application in soil spectroscopy, because of
84 its advantages and high performance (Viscarra Rossel and Behrens, 2010; Vohland et al., 2011).
85 A robust method for estimating the OM and clay content must provide sufficiently accurate
86 measurements and be cost-effective without excessive complexity, and applicable at different
87 spatial scales and in different environments (Stevens et al, 2008; Reeves, 2010). Numerous
88 studies have shown the potential of VNIRRS for digital soil mapping (Viscarra Rossel and
89 McBratney, 2008; Viscarra Rossel and Chen, 2011; Sidiki et al., 2014). Moreover, remote
90 sensing (RS) and VNIRRS combined together could be more efficient and cost-effective than
91 traditional soil mapping techniques (Lu et al., 2013; Nawar et al., 2014b). In particular, RS and
92 VNIRRS should be suitable for digital large-scale mapping studies where soil data are sparse
93 such as in the case of salt-affected soils. Although several of the studies estimated OM and clay
94 content using VNIRRS, effective and reliable methods for the accurate estimation in salt-affected
95 soils have not been implemented to date. This research aims to fill this gap and to advance the use

of diffuse reflectance spectroscopy for assessing OM and clay content in the salt-affected soils based on spectral library with limited soil samples. The main goals of this study were (1) to investigate the feasibility of using VNIRRS to predict OM and clay content in salt-affected soils; (2) to select the most effective pre-processing methods for OM and clay content estimation; and (3) to compare between PLSR, SVR, and MARS models for optimal estimation of OM and clay content.

2. Materials and methods

2.1. Study area

The study area is the El-Tina Plain, which is located on the north-western Sinai Peninsula in Egypt between longitudes 32°20′35″ and 32°33′10″ E and latitudes 30°57′25″ and 31°04′28″ N and has an area of approximately 175 km^2 (Fig. 1). The El-Tina Plain is characterised by arid conditions, with annual rainfall ranging from 33.4 mm to 70.3 mm. The mean air temperatures ranged from 7.5°C to 23.3°C in winter and between 16.3°C and 35.6°C in summer. The mean evaporation is high and ranges from 3.6 mm/day to 7.3 mm/day. The land surface is nearly flat and ranges in elevation, from below sea level to 5 m above sea level. Nawar et al. (2014a, b) found that soil texture varied from loamy sand to clay, and the soil salinity (EC_e) ranged from 3.3 to 166 dSm^{-1} and the mean salinity value was 33.03 dSm^{-1}, indicated that the soils in the study area were highly salt-affected. Nawar et al. (2011) classified the soils of the El-Tina Plain into two orders—Entisols and Aridisols, which include eight subgroups: Typic Aquisalids, Typic Haplosalids, Aquic Torriorthents, Typic Torriorthents, Aquic Torripsamments, Typic Torripsamments, Gypsic Aquisalids, and Gypsic Haplosalids.

(Fig. 1)

118 *2.2. Soil sampling and analysis*

119 One hundred and two topsoil samples (depth, 0–20 cm) were collected in the study area
120 between 28 August and 12 September 2012 (Fig. 1). The collected soil samples were air-dried,
121 crushed, and passed through a 2-mm sieve and the resulting fine earth (<2 mm) was retained for
122 analysis. OM content was determined using the modified Walkley and Black method (Page et al.,
123 1982). Particle size distribution was measured using the pipette method (Kilmer and Alexander,
124 1949). Samples (20 g) were dispersed in 400 ml of distilled water containing 10 ml of 10 percent
125 sodium hexametaphosphate by stirring with high-speed mixer for 30-60 m, and then settled for 30
126 seconds. Suspensions wer poured through 325 mesh sieves, and delivered into 1000-mL
127 graduated cylinders. Aliquots of 10 ml were taken at a 10-cm depth following a settling time as
128 calculated by Stokes' equation (Baver, 1965). Water from the aliquots was evaporated, the fines
129 dried at 105º C and the amount of suspended solids weighed. The remaining dispersed sample
130 was passed through a 300-mesh sieve; the retained sands were washed, dried at 105º C, and
131 fractionated using a nest of sieves. For pretreatment, OM, soluble salts and gypsum in the
132 samples were removed prior to particle-size analysis. OM was removed using hydrogen peroxide,
133 H_2O_2, 30%. Gypsum was removed by heating the sample to 105º C and dialysis (Rivers et al.,
134 1982). Soluble salts were removed by dialysis against water.

135 *2.3. Spectral measurements*

136 Diffuse reflectance spectra of soil samples were measured using a portable spectroradiometer
137 (FieldSpec-FR, ASD), with a measurement range of 350–2500 nm and a resolution of
138 approximately 10 nm and sampling interval of 3 nm in the short-wave infrared domain (Hatchell,
139 1999). The measurements were conducted in a dark laboratory environment. Plastic dishes were

used to keep the soil samples, which were levelled off to a thickness of 2.0 cm (Mouazen et al., 2007). A light source (halogen lamp) illuminated the sample surface with a 45° zenith angle from a distance of 30 cm. Spectral measurements were taken from nadir at 1.5 cm height above the sample. . Three spectra were collected from the central area of each sample. The final spectrum measurement was attained by averaging the three representative spectra of each sample. The reflected radiance from a white reference panel with known reflectance was recorded before scanning each sample. To calculate the absolute reflectance of the samples, the radiance from each sample was divided by the radiance from the white reference panel and multiplied by the reflectance of the reference panel.

2.4. Pre-processing transformations

Seven types of spectra pre-processing were used , including (1) raw spectra (R), (2) Savitzky–Golay (SG) smoothing, (3) first-derivative spectra with SG smoothing (FD-SG), (4) second-derivative spectra with SG smoothing (SD-SG), (5) continuum removed reflectance (CR), (6) standard normal variate and detrending (SNV-DT), 7) multiplicative scatter correction (MSC), and (8) extended MSC (EMSC). The details about this spectral transformation can be found in the review by Rinnan et al. (2009) and Buddenbaum and Steffens (2012). The continuum removal was performed using ENVI 5.0 software (Exelis Visual Information Solutions, Inc.). All the remaining pre-processing methods were carried out in MATLAB (Version 8.0; The Mathworks, Natick, MA, USA).

2.5. Multivariate calibration models

Three models, PLSR, SVR and MARS, were constructed based on the measured reflectance spectra (R) with seven spectral pre-processing methods and the measured values of OM and clay

content of the 70 soil samples. In addition to cross-validation, 32 samples (neither used in the calibration of the models nor in the cross-validation) were used for the independent validation.

2.5.1. Partial least-squares regression

PLSR is a popular regression method that is often applied in chemometrics, and it was introduced and statistically described by Geladi and Kowalski (1986) and Wold et al. (2001). This method is frequently used to conduct quantitative spectral analyses (Viscarra Rossel and Behrens, 2010), and the algorithm follows a linear multivariate model to relate the predictor (X) and response (Y) variables and select successive orthogonal (latent) factors, thereby maximising the X and Y covariance or the covariance between the spectra (X) and a measured soil property (Y). Compared to multiple linear regression, PLSR is an appropriate method for managing data with severe collinearity in the independent variables, particularly in cases where the sample size is small. To determine the number of latent factors when using PLSR, leave-one-out cross-validation (LOOCV) was used (Efron and Tibshirani, 1994) to prevent over- or under-fitting the data, which may produce models with poor performance. The root mean squared error (RMSE) of the predictions and the coefficient of determination (R^2) were ascertained to identify the optimal cross-validated calibration model. Generally, the model with the highest cross-validated R^2 value and lowest RMSE value is selected. The PLSR process was performed using MATLAB (Version 8.0; The Mathworks, Natick, MA, USA).

2.5.2. Support vector regression (SVR)

SVR is a powerful calibration model based on the kernel learning methods (Vapnik, 1999). This method utilises an implicit mapping of the input data into a high dimensional feature space defined by a kernel function (Karatzoglou and Feinerer, 2010). The epsilon-SV regression (ε-

SVR) uses training data to obtain a calibration model as the so-called ε-insensitive loss function, which is able to map independent data with deviation smaller or equal to the ε deviation from dependent training data. Errors within the predetermined distance ε from the true value are ignored, whereas errors greater than ε are penalised using parameter C (Vohland et al., 2011). Consequently, SVR reduces the complexity of the training data to an optimal number of so-called support vectors.

In this study, the models were developed using the LIB-SVR algorithm with type epsilon-SVR and radial basis function (RBF) kernel. The optimal parameters of SVR, namely, cost, epsilon, gam (γ), and sig^2 (σ^2) were adopted to regulate the models. For each combination of cost, γ, and σ^2 parameters, the RMSE of cross-validation was calculated, and the optimum parameters were selected, which resulted in the smallest RMSE. Further, we optimised cost in the range of 1–10 and epsilon and γ in the range of 0.01–0.1. X-block compression using PLS with 4 components were selected to process the data. The LOOCV was used for the validation process. The LIB-SVR model was built by utilising PLS Toolbox (Version 7.9; Eigenvector, Wenatchee, Washington) and MATLAB (Version 8.0; The Mathworks, Natick, MA, USA).

2.5.3. Multivariate adaptive regression splines

MARS is an explanatory data analysis (data mining) technique developed by Friedman (1991). Recently, MARS was applied as a regression method in several disciplines (Shepherd and Walsh, 2002; Yang et al., 2003; Bilgili et al., 2010; 2011; Nawar et al., 2014) and was generally reported to show better performance than that by conventional statistical methods. The MARS analysis uses basis functions to model the predictor and response variables (Hastie et al., 2009). To construct the basis functions, MARS splits the data into sub-regions (splines) with different interval ending knots where the regression coefficients change and fits the data in each sub-

region by using a set of adaptive piecewise linear regressions. These basis functions are subsequently used as new predictor variables for modelling purposes. Each basis function may contain nonlinear and variable interaction factors of the second and third order as well as linear combinations. The basis functions take two forms, one for the values on the left of the knot (negative) and one for the values on the right of the knot (positive):

$$Y = \max(0, Xf - cf)$$

$$Y = \max(0, cf - Xf)$$

where *cf* is the threshold value for a predictor or knot and *Xf* denotes a predictor variable (Bilgili et al., 2010). The number of knots and basis functions are determined using a forward stepwise procedure in which a deliberately over-fitted model is first constructed. Next, backward stepwise elimination is applied during which the basis functions that do not contribute sufficiently to the accuracy of the fit are discarded to obtain a final model of the proper size (Friedman, 1991).

To measure lack of fit, avoid over-fitting, and improve prediction, the redundant basis functions are removed one at a time using a backward stepwise procedure. To determine which basis functions should be included in the model, MARS uses a modified form of the generalised cross-validation method (GCV, Vidoli, 2011). The GCV is the mean squared residual error divided by a penalty dependent on the model complexity and is expressed as follows:

$$GCV = \frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i - f(x_i)\right]^2}{\left[1 - \frac{C(m)}{n}\right]^2}$$

where *n* is the number of observations, $f(x_i) = \hat{y}_i$ and *C(m)* is the cost-complexity measure of a model containing m basis functions used to penalise the model complexity to avoid over-fitting

228 by introducing a cost for the added basis functions in the model. Additional basis functions in the
229 model provide greater flexibility but also add complexity, and MARS attempts to minimise
230 model complexity by selecting the optimal model based on the lowest generalised cross-
231 validation value. The MARS analysis was performed using the ARESLab toolbox (Jekabsons,
232 2011) with selected adaptations based on MATLAB 8.0 software.

233 *2.6. Prediction accuracy*

234 The R^2, RMSE, and ratio of performance to deviation (RPD), which is the standard deviation
235 divided by RMSE were used to assess the performance of the OM and clay content prediction
236 models. The RPD was classified into three classes by Chang et al. (2001): category A (RPD > 2)
237 includes models that accurately predict a given property; category B (1.4 < RPD < 2) has limited
238 predictive ability and category C (RPD < 1.4) has no predictive ability. These classes were
239 considered in the current work to compare the performance of different modelling approaches to
240 predict OM and clay content.

241 **3. Results**

242 *3.1. Soil data*

243 The results of the descriptive statistical analyses for 70 soil samples were shown in Table 1.
244 The OM content of the samples was low with the mean and maximum values of 0.85% and 2.3%,
245 respectively (Table 1). The OM content >1.0% accounted for 66% of all samples. Clay content
246 ranged between 0.02% and 54.3%, with a mean value of 27.22%, and samples with clay content
247 >40% comprised 71% of all soil samples, indicating that clay texture is dominant among the soil
248 samples. Further, a high correlation was noted between soil OM and clay content (r = 0.63).

249 **(Table 1)**

3.2. Effects of pre-processing methods on modelling

Results showed that the different pre-processing methods had considerable effects on the performance of OM and clay content regression models (Table 2). Compared with the PLSR model without pre-processing (raw [R] spectra), CR, FD-SG, and SD-SG improved the model performance. Models based on CR showed the best accuracy (R^2 = 0.52 and 0.82, RMSE = 0.36% and 7.10%, and RPD = 1.45 and 2.34 for OM and clay content, respectively), followed by FD-SG (R^2 = 0.45 and 0.67, RMSE = 0.38% and 9.6%, and RPD = 1.35 and 1.46 for OM and clay, respectively) and SD-SG (R^2 = 0.42 and 0.53, RMSE = 0.39% and 11.44%, and RPD = 1.31 and 1.46 for OM and clay, respectively). In contrast, PLSR models based on SNV-DT, MSC, and EMSC spectra resulted in lower estimation accuracies. For OM, the worst performance was with EMSC (R^2 = 0.33, RMSE = 0.42% and RPD = 1.23), and SNV-DT produced the worst performance (R^2 = 0.18, RMSE = 15.42% and RPD = 1.08) for clay content. Even though, the performance for OM with SG (R^2 = 0.51) was better than raw spectra, the results with for clay (R^2 = 0.50) were lower than FD-SG and SD-SG (R^2 = 0.67 and 0.53). Furthermore, both FD-SG and SD-SG included smoothing step. Thus, the calibrations of the three spectral types that achieved better results with PLSR (CR, FD-SG, and SD-SG) besides raw spectra were used for calibrating the SVR and MARS models.

Figure 2 presents PLS regression coefficients for both OM and clay content PLSR models, and shows the correlation between both OM and clay content and reflectance spectra of the selected four pre-processing methods. The regression coefficients showed both positive and negative correlations at various wavelengths across spectra. The highest correlation for OM was found at 600 nm for raw spectra and at 1900 nm for CR, FD-SG, and SD-SG spectra. For clay content, the highest correlation observed is at 600 nm for raw spectra. The spectrum for CR showed positive

correlation at 700 and 1400 nm, and negative correlation at 1900 and 2200 nm. Positive correlations were obtained for FD-SG and clay content at 1400 nm, but at 1900 and 2270 nm, the correlation was negative. The highest reflectance magnitude (albedo) of samples was low (maximum 35%), except for five sandy samples (40–60%).

(Table 2)

(Fig. 2)

3.3. Performance of PLSR, SVR, and MARS for estimating OM and clay content

The calibration methods considered in this study provided different prediction accuracies of OM (Table 3,Figures 3 and 4). For the calibration data set, the best results were obtained using MARS based on CR (R^2_{Cal} = 0.85, $RMSE_{Cal}$ = 0.19%, and RPD_{Cal} = 2.63; Table 3; Fig. 3), whereas the least acceptable results for OM were shown by PLSR and SD-SG spectra (R^2_{Cal} = 0.42, $RMSE_{Cal}$ = 0.39%, and RPD_{Cal} = 1.31; Table 3; Fig. 3). Thus, the MARS model outperformed the SVR and PLSR models with respect to the R, CR, and SD-SG, but for FD-SG, the SVR model showed better result (R^2_{Cal} = 0.74, $RMSE_{Cal}$ = 0.26%, and RPD_{Cal} = 1.96). In addition, SVR showed good performance for modelling OM, and all the calibration models had RPD above 1.4. The worst results for SVR was with R and SD-SG (RPD_{Cal} = 1.52 and 1.44, respectively).

(Table 3)

(Fig. 3)

Using the validation dataset, the prediction of OM content was close to the calibration results for MARS and SVR with improvement for PLSR results. The results of MARS and PLSR with CR were the better accuracy (R^2 = 0.81, $RMSE_{Val}$ = 0.22% and RPD = 2.27 for MARS, and R^2 = 0.79, $RMSE_{Val}$ = 0.28% and RPD = 2.16 for PLSR). The validation for SVR was close to the

296 calibration results yet not as good with R^2 = 0.65, $RMSE_{Val}$ = 0.29% and RPD = 1.69. The better
297 results for SVR were received with FD-SG R^2 = 0.75, $RMSE_{Val}$ = 0.26% and RPD = 2.00.
298 Considering a relatively large number of samples in the calibration step, MARS, PLSR and SVR
299 were found to be suitable methods for determining OM content of salt-affected soils. The slopes
300 for the MARS and PLSR models with CR as well as SVR with FD-SG using the validation data
301 set were well distributed along the 1:1 line, indicating good validations, whereas the slopes for
302 MARS, SVR and PLSR with R, FD-SG (except SVR) and SD-SG spectra with the validation
303 dataset were under the 1:1 line, indicating under-estimation of OM (Fig. 4).

304 **(Fig. 4)**

305

306 Compared to OM, clay content was determined with higher accuracy. CR transformed spectra
307 produced the best calibration models for clay estimation with PLSR, SVR, and MARS models.
308 The MARS results outperformed PLSR and SVR models, with all spectra pre-processing types,
309 and the best results were recorded with CR (R^2 = 0.9, $RMSE_{Cal}$ = 5.32%, RPD = 3.15; Table 4;
310 Fig. 5). The clay content results of MARS with raw spectra showed good performance (RPD =
311 2.41), and those of the SVR were good with all spectra types and RPD ranged from 1.44 to 2.13.
312 The best result for SVR was obtained with CR (R^2=0.78, $RMSE_{Cal}$ = 7.7% and RPD = 2.13;
313 Table 4; Fig. 5). PLSR generated good results and the RPD ranged between 1.44 and 2.34. The
314 best result for PLSR was with CR and achieved R^2 of 0.82 , $RMSE_{Cal}$ of 7.1% and RPD of 2.34.

315 **(Table 4)**

316 **(Fig. 5)**

For clay content, validation results with CR were the best among all models with R^2 = 0.79, $RMSE_{Val}$ = 7.60% and RPD = 2.24, R^2 = 0.72, $RMSE_{Val}$ = 7.75% and RPD = 1.90, and R^2 = 0.64, $RMSE_{Val}$ = 8.84% and RPD = 1.67 for MARS, PLSR and SVR, respectively (Figure 6 and Table 4). Validation accuracy of SVR was the worst and slightly lower than the calibration results, however still reasonable for a relatively small validation data set (R^2 between 0.52 and 0.64, $RMSE_{Val}$ between 8.84 and 11.35%, RPD between 1.44 and 1.67). The slopes for MARS, PLSR and SVR models with CR using the validation dataset were well distributed along the 1:1 line of the calibration models, indicating good fit, whereas the slopes for MARS, SVR and PLSR with R, FD-SG and SD-SG spectra using the validation dataset were over the 1:1 line, indicating over-estimation of clay content (Fig. 6).

(Fig. 6)

4. Discussion

4.1 Soil spectra

Significant absorbance features around 1400, 1900, 2200, and 2400 nm were observed. These are strongly associated with free water OH features at 1400 and 1900 nm and clay lattice OH features at 1400 and 2200 nm (Vicente and Souza, 2011). The strong absorptions near 1416 nm and 1911 nm in the spectra indicate the presence of water bound in the interlayer lattices (Bishop et al., 2008). The absorption feature near 2200 nm may is likely associated with the absorption of Al–OH, and the small absorption feature near 2280 nm may appear because of Fe–OH, as Fe is exchanged in the octahedral sheet, e.g. in montmorillonite. In the spectra, the slight shoulder at 2340 nm may represent illite or mixtures of muscovite minerals (Post and Noble, 1993).

Many studies have concentrated on sensitive spectral wavebands and their relationship with OM content (Viscarra Rossel et al., 2006; Bartholomeus et al., 2008; Wang et al., 2010; Tekin et

al., 2012; Nocita et al., 2014). For instance, Viscarra Rossel et al. (2006) reported that the wavelengths 410, 570, and 660 nm in the Vis range showed good correlations with soil OM. Furthermore, Wang et al. (2010) reported 5 wavelength bands (440, 560, 625, 740, and 1336 nm) as the optimal spectral bands for the estimation of soil OM under laboratory conditions. Bartholomeus et al. (2008) found the highest correlation between absorption and soil organic carbon at wavelengths between 640 nm and 690 nm. In addition, Nacita et al. (2014) reported an absorption peak centred at 600 nm that was associated with the high estimation of soil organic carbon. The results of the present work also showed that the highest correlation for OM was found at a region between 550 and 680 nm of the raw spectrum, which may be associated with soil colour (Mouazen et al., 2007). This is consistent with the results of Bartholomeus et al. (2008) and Nacita et al. (2014), suggesting that the raw spectral reflectance in the 550–680 nm is sufficiently sensitive for evaluating the OM content in soil.

Viscarra Rossel and McBratney (1998) found that the significant wavelengths for estimating clay content in the NIR range were 1600, 1800, 2000, and 2100 nm. Lagacherie et al. (2008) reported that the wavelength 2206 nm of CR was the optimal band for quantifying clay content. A similar result has been confirmed by Viscarra Rossel et al. (2009), who reported that the absorption features at 2206 and 2207 nm of the CR spectrum are the optimal wavelengths for estimating defined clay minerals. Our study results are in line with the results achieved in these previous studies, showing that the features at wavelengths 1900, 2000, and 2200 of the CR had strong correlation with clay content (Figs. 2b).

4.2. Influence of pre-processing methods on estimation models

In this study, the pre-processing methods significantly affected the results of calibration models (Table 2 and Fig. 2). For OM, CR pre-processing achieved the best performance followed

by SG and FD-SG methods. Similarly, CR pre-processing achieved the best performance for clay content models followed by FD-SG and SD-SG methods. Although PLSR results with SG for OM (R^2=0.50 and RMSE=0.37%) were better than models with the raw spectra (R^2=0.44 and RMSE=0.39%), the results of clay content modelling, contrary to OM models, were better for the raw spectra (R^2=0.56 and RMSE=11.08%) than for the spectra processed with SG technique (R^2=0.50 and RMSE=12.42%). Furthermore, smoothing using SG technique was included in FD-SG, and SD-SG pr-processing methods. Using SG, SNV-DT, MSC, and EMSC led to the deterioration of the models accuracy as compared to the PLSR model based on raw spectra. SNV-DT is sensitive to the noise in the spectrum (Rinnan et al., 2009).

CR was the best pre-processing method for the PLSR, SVR, and MARS estimation models in this study. CR has also been successfully used in some other studies for predicting soil properties (Lagacherie et al., 2008; Weng et al., 2008; Nawar et al., 2014a). For example, Lagacherie et al. (2008) recorded improvements in the estimation of the clay content by using reflectance spectroscopy after implementing CR pre-processing on the soil spectra. On the other hand, FD-SG was used to improve the estimation of OM (Vasques et al. 2008). The reason of better performance of CR, as well as FD-SG and SD-SG pre-processing methods was effective noise removal and enhanced signal-to-noise ratio, beneficial for OM and clay content estimation.

4.3. Comparison of PLSR, SVR, and MARS estimation models

In our study, the PLSR, SVR, and MARS methods were used to compare the estimation of soil OM and clay content. Generally, the accuracy of the PLSR, SVR, and MARS models is affected by variations in the soil texture and moisture content (Farifteh et al., 2007), and a successful pre-processing method for the spectral data may improve the performance of such models (Vasques et al., 2008; Rinnan et al., 2009). Amongst the three investigated multivariate techniques, MARS

performed best. Compared to MARS, the two other multivariate techniques (SVR and PLSR) showed a noticeable decrease in accuracy with PLSR performing somewhat better than SVR in the validation data set for OM and clay content estimation.

For OM estimation, MARS provided very good calibration accuracies depending on the pre-treatment method used with R^2 between 0.70 and 0.85, $RMSE_{Cal}$ ranged between 19% and 28%, and all RPD higher than 2.0, except with FD-SG (RPD =1.86). The best model performance (R^2 =0.85 and $RMSE_{Cal}$ = 0.19%) was obtained using MARS with CR. The accuracy of the test set validation is good and in the same range as the calibration (RPD close to 2.0). These models particularly showed good $RMSE_{Val}$ around 0.35%. The validation accuracy was slightly lower but reasonable (R^2 between 0.49 and 0.62, $RMSE_{Val}$ between 0.43% and 0.54%, RPD between 1.26 and 1.57).

As shown in Table 3, the PLSR model with CR pre-processing performed relatively well compared to MARS and better than SVR with CR as well as with SD-SG and R for the validation data set. The good accuracy of PLSR with validation data set may be attributed to the fine texture of the soil samples (75% of samples had clay texture) The best results among SVR models were with FD-SG (R^2 = 0.76 and RMSE = 0.25%). CR generally emphasises the spectral absorption features (Clark and Roush, 1984) and may improve the estimations. Weng et al. (2008) recorded improvements in the estimation of the soil salinity content by using reflectance spectroscopy after implementing CR pre-processing on the soil spectra. Our study results showed that CR generally improved these estimations compared to estimations from the raw spectra. SVR produced good results with FD-SG (R^2 = 0.76 and RMSE = 0.25%), which was better than that by PLSR, but not as good as the results obtained by using the MARS model.

Compared to published results obtained using MARS (R^2_{Cal} = 0.79–0.91) (e.g. Shepherd et al., 2002; Bilgili et al., 2010; Viscarra Rossel and Behrens, 2010), SVR (R^2_{Cal} = 0.73–0.89 (e.g. Viscarra Rossel and Behrens, 2010; Vohland et al., 2011; Peng et al., 2014), and PLSR (R^2_{Cal} = 0.79–0.85) (e.g. Vasques et al., 2008; Bilgili et al., 2010; Viscarra Rossel and Behrens, 2010), the OM models used in the present study may be considered as good according to the standard by Chang et al. (2001). The estimation quality of the OM was dependent on the different pre-treatment and calibration methods. For instance, the MARS model developed with raw spectra, CR and SD-SG produced better results than the PLSR and SVR (Table 3). However, the results of SVR with FD-SG were better than the MARS and PLSR. Similarly, averaging the spectra consistently optimised the calibrations with PLSR. Averaging and smoothing of the spectra removes noise and may eliminate factors that lead to nonlinearities (Bilgili et al., 2010). The MARS method typically yields better results when a nonlinear relationship between absorption and concentration exists, whereas the PLSR model fits only linear relationships (Brown et al., 2006; Mouazen et al., 2010; Viscarra Rossel and Behrens, 2010).

Very good predictions for clay content were obtained using MARS (R^2_{Cal} = 0.90) and good performances were obtained with PLSR (R^2_{Cal} = 0.82) and SVR (R^2_{Cal} = 0.78). Table 4 shows that the calibration models with CR performed well and provided the best results for MARS, PLSR, and SVR, respectively. These results were consistent with previous studies for estimating clay content based on reflectance spectroscopy (R^2_{Cal} = 0.82–0.91) (Shepherd and Walsh, 2002; Lagacherie et al., 2008; Bilgili et al., 2010; Viscarra Rossel and Behrens, 2010). For example, Bilgili et al. (2010) obtained good results for estimating clay content of 512 soil samples collected from northern Turkey, by using MARS and PLSR and reported R^2_{Cal} of 0.89 and 0.82,

430 respectively. According to Chang et al. (2001), the predictive performance of the MARS model
431 in this study was high.

432 In line with findings of Kuang and Mouazen (2011), it is believed that a fundamental factor
433 that improved the performance of the models in this study was a wide range of clay content (0.0–
434 54.3). Furthermore, the model performance depends to a large extent on variability encountered
435 in the data set, including soil types (Stenberg, 2010; Wang et al., 2010).that due to variety of
436 parent material can have considerably different physical and chemical characteristics influencing
437 soil spectra. Thus, if soil heterogeneity increases regression can be more successful and this may
438 influence the model accuracy. Kuang and Mouazen (2011) confirmed that although R^2 and RPD
439 increase with variability measured as SD and range, RMSE also increased. The high performance
440 for estimating clay content in the current research obtained with MARS and SVR may be
441 attributed to the wide range of variation of clay content as shown in Table 1.

442 The performance of the MARS, SVR, and PLSR models to predict soil properties is evaluated by
443 R^2, RMSE, and RPD. Of all models, MARS showed higher performance than PLSR and SVR for
444 estimating clay content and OM, except SVR with FD-SG in calibration and PLSR with CR in
445 validation step. According to Chang et al. (2001), estimation models of soil properties with RPD
446 >2 are considered as accurate calibration models. The performance of OM and clay content
447 predictive models were good in terms of RPD. MARS provided very good models for clay
448 content estimation (RPD_{Cal} = 3.15) with CR. Good models were also received for PLSR (RPD_{Cal}
449 = 2.34) and SVR (RPD_{Cal} = 2.13) with CR. For OM, a good model was achieved by MARS with
450 CR (RPD_{Cal} = 2.63), and a satisfactory model by SVR with FD-SG spectra (RPD_{Cal} = 1.96).

451 The predictive models used in this study have not been tested on digital hyperspectral airborne
452 or satellite images. They will be tested using airborne or satellite imagery at a selected study site

where further work is currently being conducted. Future research should focus on possible integration between spectra obtained in the field or in the laboratory, and spectra from airborne or satellite imagery. This integration could be useful for accurate OM and clay content mapping that employs predictive models (Lu et al., 2013; Nawar et al., 2015). These estimation models should be subjected to further examination and optimisation before broad application in soil mapping, as the estimation accuracy of OM and clay contents in the field could be negatively affected by several factors that are well controlled in laboratory conditions, such as spectral mixture, atmospheric conditions, and spatial variability of soil moisture content (Gomez et al., 2008; Ben-Dor et al., 2009).

5. Conclusions

In this study, different pre-processing methods were utilised for the estimation of OM and clay content in salt-affected soils based on VNIRRS, and the accuracies were compared between PLSR, SVR, and MARS models. The CR and FD-SG pre-processing methods had strong positive influence on the performance of most models. Overall, MARS was the most reliable method for estimating the OM and clay content under laboratory conditions and provided better estimations than the SVR and PLSR at both the calibration validation steps. The main strengths of MARS method, besides accuracy of modeling OM and clay content in salt-affected soils, are the lower costs compared to the conventional field-based approaches and simplicity to transfer models to image pixels for digital soil mapping. It is proposed that these estimation models undergo further testing and optimization before broad application in soil mapping. For instance, appropriate methods of atmospheric correction and elimination of spectral noise (e.g., vegetation influences) are needed to advance laboratory-based models to operational soil mapping of large areas.

References

Bartholomeus, H.M., Schaepman, M.E., Kooistra, L., Stevens, A., Hoogmoed, W.B., Spaargaren, O.S.P., 2008. Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145, 28–36. doi:10.1016/j.geoderma.2008.01.010

Baver, L. D. 1965. Soil physics. John Wiley and Sons, New York.

Ben-Dor, E., Chabrillat, S., Dematte, J. A. M., Taylor, G. R., Hill, J., Whiting, M. L. and Sommer, S. 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. 113: 538–555. doi:10.1016/j.rse.2008.09.019

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. 15, 1–15.

Bilgili, A.V., Cullu, M.A., van Es, H., Aydemir, A., Aydemir, S., 2011. The Use of Hyperspectral Visible and Near Infrared Reflectance Spectroscopy for the Characterization of Salt-Affected Soils in the Harran Plain, Turkey. Arid L. Res. Manag. 25, 19–37. doi:10.1080/15324982.2010.528153

Bilgili, A.V., van Es, H. M., Akbas, F., Durak, A., Hively, W. D. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. J. Arid Environ. 2010, 74, 229–238.doi:10.1016/j.jaridenv.2009.08.011

Bishop, J.L., Lane, M.D., Dyar, M.D., Brown, A. J., 2008. Reflectance and emission spectroscopy study of four groups of phyllosilicates: smectites, kaolinite-serpentines, chlorites and micas. Clay Miner. 43, 35–54. doi:10.1180/claymin.2008.043.1.03

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132, 273–290. doi:10.1016/j.geoderma.2005.04.025

Buddenbaum, H., Steffens, M., 2012. The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy. Appl. Environ. Soil Sci. 2012, 1–12. doi:10.1155/2012/274903

Chang, C., Laird, D., Mausbach, M.J., 2001. Near-Infrared Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties Analyses of Soil Properties.

Clark, R. N., Roush, T. L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. J Geophys. Res., 89, 6329-6340.

Hatchell., D. C., 1999. ASD technical guide (3rd Ed.). Analytical Spectral Devices, Inc. Boulder, CO 80301, USA.

Efron, B; Tibshiran, R. (Ed.), 1993. An introduction to the bottstrap. Chapman &Hall, Inc., London.

Farifteh, J., Van der Meer, F., Atzberger, C., Carranza, E.J.M., 2007. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (PLSR and ANN). Remote Sens. Environ. 110, 59–78. doi:10.1016/j.rse.2007.02.005

Felicísimo, Á.M., Cuartero, A., Remondo, J., Quirós, E., 2012. Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. Landslides 10, 175–189. doi:10.1007/s10346-012-0320-1

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. Ann. Stat. 19, 1–67. 19, 1–67.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17. doi:10.1016/0003-2670(86)80028-9

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field VIS–NIR spectroscopy: an Australian case study. Geoderma 146, 403–411.

Hastie T, Tibshirani R, Friedman JH. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer-Verlag. 744 pp.

Jackson, M.L., 1973. Soil Chemical Analysis; Prentice-Hall of India private limited: New Delhi, India; p. 498.

Jekabsons, G., 2011. ARESLab Adaptive Regression Splines toolbox for Matlab / Octave 1–19. available at http://www.cs.rtu.lv/jekabsons/

Karatzoglou, A., Feinerer, I., 2010. Kernel-based machines learning for fast text mining in R. Comput. Stat. Data Anal. 54, 290–297. doi:10.1016/j.csda.2009.09.023

Kilmer, V. J.; Alexander L.T. Methods of making mechanical analysis of soils. Soil Science, 68, 15-24.

Kuang, B., Mouazen, A.M., 2011. Calibration of a visible and near infrared spectroscopy for soil analysis at field scales across three European farms. Eur. J. Soil Sci. 62, 629–636.

Lagacherie, P., Baret, F., Feret, J., Madeira, J., Robbez-masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory , field and airborne hyperspectral measurements 112, 825–835. doi:10.1016/j.rse.2007.06.014

Li, X., He, Y., Wu, C., 2008. Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. J. Stored Prod. Res. 44, 264–268. doi:10.1016/j.jspr.2008.01.004

Lu, P., Wang, L., Niu, Z., Li, L., Zhang, W., 2013. Prediction of soil properties using laboratory VIS – NIR spectroscopy and Hyperion imagery. J. Geochemical Explor. 132, 26–33.

Luoto.,M., Hjort, J., 2005. Evaluation of current statistical approaches for predictive geomorphological mapping. Geomorphology 67, 299–315. doi:10.1016/j.geomorph.2004.10.006

Mashimbye, Z.E., Cho, M.A., Nell, J.P., DE Clercq, W.P., van Niekerk, A., Turner, D.P., 2012. Model-Based Integrated Methods for Quantitative Estimation of Soil Salinity from Hyperspectral Remote Sensing Data: A Case Study of Selected South African Soils. Pedosphere 22, 640–649. doi:10.1016/S1002-0160(12)60049-6

Masoud, A., 2014. Predicting salt abundance in slightly saline soils from Landsat ETM+ imagery using spectral mixture analysis and soil spectrometry. Geoderma 217–218, 45–56.

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, 23–31. doi:10.1016/j.geoderma.2010.03.001

Mouazen, A.M., Maleki, M.R., Baerdemaeker, J. De, Ramon, H., 2007. On-line measurement of some selected soil properties using a VIS – NIR sensor. Soil Till. Res. 93, 13–27. doi:10.1016/j.still.2006.03.009

Nawar, S., Buddenbaum, H., Hill, J., 2015. Digital Mapping of Soil Properties Using Multivariate Statistical Analysis and ASTER Data in an Arid Region. Remote Sens. 1181–1205. doi:10.3390/rs70

Nawar, S., Buddenbaum, H., Hill, J., 2014a, Estimation of soil salinity using three quantitative methods based on visible and near-infrared reflectance spectroscopy: a case study from Egypt. Arab. J. Geosci. doi:10.1007/s12517-014-1580-y

566 Nawar S., Buddenbaum H., Hill J., Kozak J., 2014b, Modeling and Mapping of Soil Salinity with
567 Reflectance Spectroscopy and Landsat Data Using Two Quantitative Methods (PLSR and
568 MARS). Remote sesn. 6, 10813-10834. doi:10.3390/rs61110813
569 Nawar, S., Reda, M., Farag, F., El-Nahry, A., 2011. Sustainability assessment using geostatistical
570 analysis and spatial modeling in El-Tina plain in Egypt. Proc. 25th Int. Conf. EnviroInfo,
571 Ispera, Italy, 5–7 Oct. 2011.
572 Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., Montanarella, L., 2014.
573 Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local
574 partial least square regression approach. Soil Biol. Biochem. 68, 337–347.
575 doi:10.1016/j.soilbio.2013.10.022
576 Page, R. H. Miller, D. R. Keeney, D.E. Baker, R. Ellis, Jr., Rhoades, J.D. (eds)., 1982. Methods
577 of Soil Analysis. Part 2. 2nd edition. Chemical and microbiological properties. Agronomy
578 No. 9. American Society of Agronomy, Madison, Wisconsin; pp. 539-580.
579 Peng, X., Shi, T., Song, A., Chen, Y., Gao, W., 2014. Estimating Soil Organic Carbon Using
580 Vis/NIR Spectroscopy with SVMR and SPA Methods. Remote Sens. 6, 2699–2717.
581 doi:10.3390/rs6042699
582 Post, J.L., Noble, P.N., 1993. The near-infrared combination band frequencies of dioctahedral
583 smectites, Micas, and Illites. Clays and Clay Minerals 41, 639–644.
584 Reeves, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis
585 emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs
586 to be done?. Geoderma 158, 3–14. doi:10.1016/j.geoderma.2009.04.005
587 Richards, L.A., 1954. Diagnosis and Improvement of Saline and Alkaline Soils. United States
588 Salinity Laboratory Staff. Agricultural Handbook No. 60. USDA, p. 160.
589 Rinnan, Å., Berg, F. Van Den, Engelsen, S.B., 2009. Review of the most common pre-processing
590 techniques for near-infrared spectra. TrAC Trends Anal. Chem. 28, 1201–1222.
591 doi:10.1016/j.trac.2009.07.007
592 Rivers, E.D., Hallmark, C.T., West, L.T., Drees, L.R., 1982. A technique for rapid removal of
593 gypsum from soil samples. Soil Sci. Soc. Am. J. 46,1338–1340.
594 Samui, P., 2012. Multivariate Adaptive Regression Spline (Mars) for Prediction of Elastic
595 Modulus of Jointed Rock Mass. Geotech. Geol. Eng. 31, 249–253. doi:10.1007/s10706-
596 012-9584-4
597 Shepherd, K.D., Walsh, M.G., 2002. Development of Reflectance Spectral Libraries for
598 Characterization of Soil Properties. Soil Sci. Soc. Am. J. 66, 988.
599 doi:10.2136/sssaj2002.0988
600 Sidike, A., Zhao, S., Wen, Y., 2014. Estimating soil salinity in Pingluo County of China using
601 QuickBird data and soil reflectance spectra. Int. J. Appl. Earth Obs. Geoinf. 26, 156–175.
602 Stenberg, B., 2010. Visible and Near Infrared Spectroscopy in Soil Science 107, 163–215.
603 doi:10.1016/S0065-2113(10)07005-7
604 Stevens, A., Udelhoven, T., Denis, A., Tychon, B., Lioy, R., Hoffmann, L., van Wesemael, B.,
605 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging
606 spectroscopy. Geoderma 158, 32–45. doi:10.1016/j.geoderma.2009.11.032
607 Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008.
608 Laboratory, field and airborne spectroscopy for monitoring organic carbon content in
609 agricultural soils. Geoderma 144, 395–404. doi:10.1016/j.geoderma.2007.12.009
610 Suykens, J., Vandewalle, J., 1999. Least squares support vector machines classifiers. Neural
611 Process. Lett. 293–300.

Tekin, Y., Tumsavas, Z., Mouazen, A. M., 2011. Effect of moisture content on prediction of organic carbon and pH using visible and near infrared spectroscopy. Soil Sci. Soc. Am. J. 76, 188-198. doi:10.2136/sssaj2011.0021.

Obade, V. D. P.; Lal, R., 2013. Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). Catena 104, 77–92.

Vapnik, V.N., 1998. Statistical Learning Theory. John Wiley and Sons, Inc., New York.

Vapnik, V.N., 1999. An overview of statistical learning theory. IEEE Trans. Neural Netw. 10, 988–99. doi:10.1109/72.788640

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. doi:10.1016/j.geoderma.2008.04.007

Vicente, L.E., de Souza Filho, C.R., 2011. Identification of mineral components in tropical soils using reflectance spectroscopy and advanced spaceborne thermal emission and reflection radiometer (ASTER) data. Remote Sens. Environ. 115, 1824–1836. doi:10.1016/j.rse.2011.02.023

Vidoli, F., 2011. Evaluating the water sector in Italy through a two stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines. Eur. J. Oper. Res. 212, 583–595. doi:10.1016/j.ejor.2011.02.003

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. doi:10.1016/j.geoderma.2009.12.025

Viscarra Rossel, R.A., Chen, C., 2011. Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. Remote Sens. Environ. 115, 1443–1455.

Viscarra Rossel, R. A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150, 253–266. doi:10.1016/j.geoderma.2009.01.025

Viscarra Rossel, R.A., McBratney, A.B., 2008. Diffuse reflectance spectroscopy as a tool for digital soil mapping. In: McBratney, A.B., Hartemink, A.E., Mendonc¸a-Santos, L., (Eds.), Digital Soil Mapping with Limited Data. Developments in Soil Science Series, Elsevier Science, Amsterdam.

Viscarra Rossel, R.A., McBratney, A.B., 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geoderma 85, 19–39. doi:10.1016/S0016-7061(98)00023-8

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. doi:10.1016/j.geoderma.2005.03.007

Vohland, M., Besold, J., Hill, J., Fründ, H.-C., 2011. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. Geoderma 166, 198–205. doi:10.1016/j.geoderma.2011.08.001

Wang, J., He, T., Lv, C., Chen, Y., Jian, W., 2010. Mapping soil organic matter based on land degradation spectral response units using Hyperion images. Int. J. Appl. Earth Obs. Geoinf. 12, S171–S180. doi:10.1016/j.jag.2010.01.002

Weng, Y., Gong, P., Zhu, Z., 2008. Soil salt content estimation in the Yellow River delta with satellite hyperspectral data 34, 259–270.

Wold, S., Trygg, J., Berglund, A., Antti, H., 2001. Some recent developments in PLS modeling. Chemom. Intell. Lab. Syst. 58, 131–150. doi:10.1016/S0169-7439(01)00156-3

Xuemei, L., Jianshe, L., 2013. Measurement of soil properties using visible and short wave-near infrared spectroscopy and multivariate calibration. Measurement 46, 3808–3814. doi:10.1016/j.measurement.2013.07.007

Yang, C.-C., Prasher, S.O., Lacroix, R., Kim, S.H., 2003. A Multivariate Adaptive Regression Splines Model for Simulation of Pesticide Transport in Soils. Biosyst. Eng. 86, 9–15. doi:10.1016/S1537-5110(03)00099-0

Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy

Nawar, Said

2015-08-14

Said Nawar, Henning Buddenbaum, Joachim Hill, Jacek Kozak, Abdul M. Mouazen, Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy, In Soil and Tillage Research, Volume 155, 2016, Pages 510-522
http://dx.doi.org/10.1016/j.still.2015.07.021