International Journal of Geography, Geology and Environment

P-ISSN: 2706-7483
E-ISSN: 2706-7491
IJGGE 2020; 2(1): 04-09
https://www.geojournal.net
Received: 04-11-2019
Accepted: 06-12-2019

Abd El-Rahman A Mustafa
Division of Soil and Water Sciences, Faculty of Agriculture, Sohag University, Egypt

Ali RA Moursy
Soil Science and Agricultural Chemistry, Indian Agricultural Research Institute, New Delhi, India

Corresponding Author:
Ali RA Moursy
Soil Science and Agricultural Chemistry, Indian Agricultural Research Institute, New Delhi, India

Using a multivariate regression model and hyperspectral reflectance data to predict soil parameters of Agra, India

Abd El-Rahman A Mustafa and Ali RA Moursy

DOI: https://doi.org/10.22271/27067483.2020.v2.i1a.12

Abstract
The diffuse reflectance spectroscopy was found to be a promising technique for soil parameters' estimation. Moreover, the integration of multivariate regression models and Vis-NIR hyperspectral reflectance data proved high efficiency for soil characterization. Thus, this study aimed to estimate pH, ECe and $CaCO_3$ soil parameters using partial least square regression (PLSR) and soil hyperspectral signature. Surface soil samples were collected from Agra, Uttar Pradesh, India. Soil samples were prepared and analyzed for examined parameters. In hyperspectral remote sensing laboratory conditions, soil hyperspectral signatures were collected using an analytical spectroradiometer devise in the spectral range from 350 to 2500 nm. The PLSR model was applied to soil spectra and soil parameters' data to develop the calibration and validation models. The obtained results showed that pH and $CaCO_3$ parameters were having high predictability whereas R^2 values of prediction were 0.69 and 0.83 with RPD values were 1.70 and 2.06, respectively. The PLSR prediction model did not perform well for predicting ECe parameter whereas R^2 and RPD values were 0.31 and 1.20, respectively. These techniques can be applied in both laboratory and field conditions by using spectroradiometers. It is rapid, time and cost-effective, and friendly to the environment. Furthermore, it can estimate many soil parameters at the same time with minimum or without samples preparation.

Keywords: Spectroradiometer, Vis-NIR, PLSR, hyperspectral

1. Introduction

Soil is a heterogeneous system which difficult to be comprehended and fully understood. The most reliable way to characterize soil is the conventional methods of soil analysis (Rossel *et al.*, 2006) [1]. Unfortunately, these methods are expensive, time and chemicals consuming, laborious, and require a lot of preparation stages (Disla *et al.*, 2014) [2].

For that, the Diffuse Reflectance Spectroscopy (DRS) has proven high efficiency for estimating soil properties (Dammate *et al.*, 2015) [3]. This technique can be applied in both laboratory and field conditions by using spectroradiometers. It can estimate many soil properties at the same time with minimum or without samples preparation (Kadupitiya *et al.*, 2010) [5]. The Vis-NIR-MIR spectral range (0.35 to 25 µm) is suitable for estimating the majority of soil properties (Ogen *et al*, 2019) [8]. Nowadays, multivariate statistics and chemometrics are used in the prediction of soil parameters by quantitative soil spectroscopy, and these techniques still growing (Chabrillat *et al.*, 2013) [9].

Data analysis techniques are dependent on the number of spectral variables of the soil spectral data. The spectral data obtained from field or laboratory conditions by ground spectrometers are noisy and hard to be evaluated. Here the role of spectral transformation appears to clean noises, correct non-linearity measurement, sample variations and develop fit soil spectral curves (Stenberg, 2010) [10]. Partial Least Square Regression (PLSR) is the most popular and widely used technique in chemometrics for quantitative analysis of reflectance spectra (Wold *et al.*, 2001). The ability of the hyperspectral RS technique to predict a soil property could be evaluated using statistical parameters such as the correlation coefficient (R^2), the Root Mean Square Error (RMSE) and Ratio of Performance Deviation (RPD) which are commonly used for the DRS technique (Woodcock, 2006) [12].

Many researchers reported good results with regression analysis for soil properties characterization. Rossel *et al.* (2006) [1] applied DRS in across the Vis-NIR-MIR spectra to analyze the soil properties using the PLSR algorithm for prediction.

They found that R^2 values for pH and EC were 0.73 and 0.29, respectively. The integration between Vis-NIR and PLSR model has an advantage for determining soil EC with R^2=0.90 (Fikrat *et al.*, 2016) [14]. Srivastava *et al.*, (2004) [13] applied linear regression-NIR modeling to predict soil pH in a part of central India, the R^2 value was 0.77. By using the PCR model, Kadupitiya *et al.* (2010) [5] were able to predict pH and EC in Punjab soils with R^2=0.82 and 0.85, respectively. Ostovari *et al.* (2018) [15] used the PLSR model and Vis-NIR range of soil spectra to predict soil CaCO3 which R^2=0.56 and RPD=1.50. Miloš and Bensa (2017) [16] predicted CaCO3 content in some soils in Croatia with R^2=0.86 and RPD=2.42.

Thus, the current study aimed to use the hyperspectral remote sensing technique integrated with PLSR for characterizing and predicting soil parameters, and also to assess the performance of the applied prediction model.

2. Materials and Methods

2.1 Study area description

The study was conducted in Kheragarah tehsil of Agra district of Uttar Pradesh state, India. It is located between geo-coordinates 26° 44' 31.43" to 27° 4' 7.80" N and 77° 27' 21.27" to 78° 7' 22.42" E covering about 80,000 ha. The agro-climate of the study area is characterized by hot dry sub-humid to semi-arid transition with intense hot summer, cold winter, and general dryness throughout the year except during July and September. The mean air temperature varies from 27° to 36 °C in summer and the maximum temperature goes up to 50 °C in June. The winter average temperature ranges from 6.5° to 13 °C and dropping to a minimum of 4 °C during January. The area receives mean annual rainfall ranging between 600 to 1000 mm. The mean rainfall in winter is considered as insufficient for growing rabi crops.

2.2. Soil sampling

Forty-seven representative surface soil samples (0–25 cm) were collected from the study area. The soil samples were air-dried, ground and 2mm sieved to be scanned using the spectroradiometer and also analyzed for their properties. Figure (1) showed the location map of the study area and also the soil sampling locations.

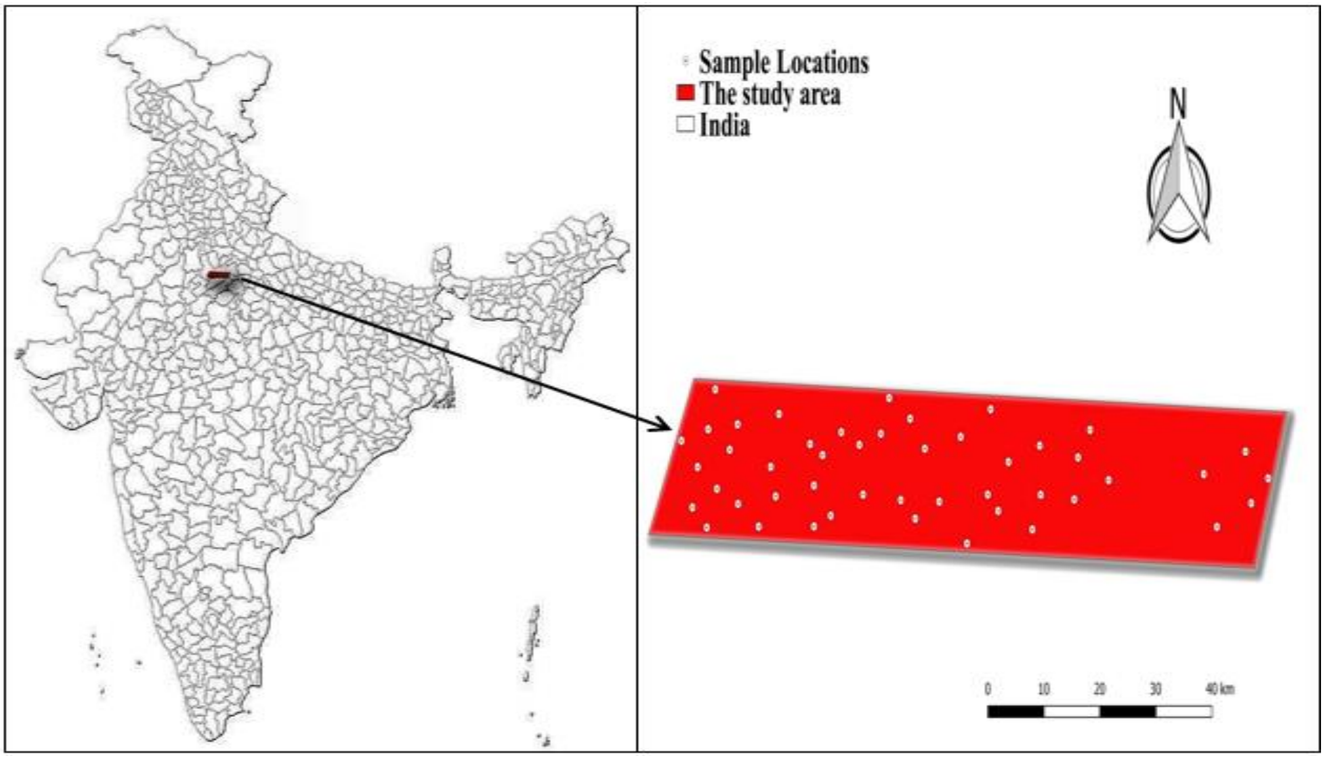

Fig 1: Location map of the study area and soil sampling locations.

2.3 Soil analysis

Conventional methods of soil analysis were followed. The soil pH was determined in 1:2 soil: water suspension by a potentiometric method using a glass electrode (Jackson, 1967) [6]. Electrical conductivity was determined in 1:2 soil-water extract using Conductivity Bridge and expressed as $dS.m^{-1}$ (Jackson, 1967) [6]. Soil content of the calcium carbonate was estimated using a rapid manometric method using Calcimeter (Williams, 1949) [17].

2.4 Hyperspectral data collection

The 2 mm ground soil samples with 2 cm thickness were scanned using Field-Spec3 Analytical Spectral Device (ASD; Boulder, CO, USA) covering wavelength ranging from 350 to 2500nm in the hyperspectral remote sensing laboratory's condition (Liu *et al.*, 2002) [18]. Reflectance spectra were measured under two calibrated halogen lamps (1000 W) situated at 0.70 m with a zenith angle of 30° in a dark room after calibration of the sensor using a white spectral panel. The ASD software has been set to process reflectance at a 1 nm interval. Spectral reflectance was derived as the ratio of reflected radiance to incident radiance estimated by a calibrated white reference. All the recorded soil spectral signatures were converted into Tab-delimited text file format using the View Spec Pro (Version 4.05) software to facilitate data sharing with other software. Figure (2) showed the Vis-NIR soil hyperspectral signatures of the study area.

2.5 Multivariate regression model application

The PLSR is a commonly used technique for quantitative spectral analysis. It is used to develop prediction models when many predictor variables are highly collinear. The PLSR algorithm selects the best orthogonal factors that maximize the covariance between predictor (X spectra) and response variables (y laboratory data). By fitting a PLSR model, a few PLSR factors are selected to explain most of the variation in both predictors and responses. The PLSR decomposes X and y into factor scores (T) and factor loadings (P and q) according to the following equations (1 and 2).

$$X = TP + E \quad (1)$$

$$y = Tq + f \quad (2)$$

whereas, X and y are mean-centered before decomposition. The decomposition is performed simultaneously and in such a way that the first few factors explain most of the variation in X and y. The remaining factors relate to noise and can be ignored, hence the addition of residuals E and f. Generally, the resulting matrices and vectors have a much lower dimension than X and y. Therefore, given a new spectrum x, the soil property y can be estimated as a bilinear combination of the factor scores and factor loadings of x (Martens and Næs 1989) [7].

The PLS package in R studio software was used for developing the calibration and validation models of the different studied soil parameters. Soil spectral data and the laboratory soil data were combined in (.csv) files to be used in R software. Moreover, the spectral data, as well as the soil parameters' data, were processed through different stages.

For enhancing the modeling performance and the predictability of different soil parameters, data processing was done through the following stages: (i) data normalization (giving values between 0 and 1 for the soil spectral data); (ii) data dividing (the whole spectral and soil data were divided into two data sets; 70% of the data (*n*=33) was separated to be as a calibration data set and 30% of the data (*n*=14) for a validation); (iii) data sorting (data arrangement for the randomized distributing the values depending on their weights among the calibration and validation data sets); and (iv) removing the outliers (the much higher or lower values in the whole soil parameters' data set were removed as outliers).

2.6 Models quality evaluation (Validation of the developed prediction models)

Two statistical indices were used for validation of developed prediction models and were the R^2, Randomized Mean Square Error (RMSE), Ratio Prediction Deviation (RPD) as described by (Islam *et al.*, 2003) [4] and shown in equations (3, 4 and 5).

2.6.1 The correlation coefficient (R^2)

$$R^2 = 1 - \left(\frac{\Sigma(Y_{pred} - Ymeas)^2}{\Sigma(Y_i - Ymeas)^2}\right) \quad (3)$$

Where, Ypred = predicted values; Ymean = mean of measured values; Ymeas = measured values; n= number of predicted or measured values with I = 1, 2, ...n.

2.6.2 Room Mean Square Error (RMSE)

$$RMSE = \sqrt{1/n\Sigma(y - x)2} \quad (4)$$

Where y is a predicted value of soil parameter and x is a measured value.

2.6.3 Ratio of Performance Deviation (RPD)

$$RPD = \frac{SD}{RMSE} \quad (5)$$

where SD is the standard deviation of measured values in the validation dataset; and RMSE= root mean square error of prediction in the validation dataset.

Chang *et al.* (2001) [19] categorized the ability of NIR spectra to predict soil properties into three categories based on the ratio of performance deviation (RPD) and the Correlation coefficient (R^2) values as shown in table (1).

Table 1: Categories of NIR predictability of soil parameters

Category	RPD	R^2	Parameters
A	>2	1-0.8	TC, TN, moisture, sand, silt, exch.Ca and CEC.
B	2-1.4	0.8-0.5	Clay, pH, mineralizable N, extractable K, Ca, Mg, Fe and Mn
C	<1.4	<0.5	Extractable Cu, P, Zn and Na.

3. Results and Discussion

3.1 Soil characterization

Descriptive statistical analysis of the soil properties data was given in table (2). From the obtained data, the soil of the studied area was ranged from slight to strong alkaline whereas soil pH values were ranged from 7.73 to 9.75 with an average of 8.50. Soil ranged from non-saline to strong saline soils. The ECe values were ranged from 0.80 to 11.00 ds.m^{-1} with an average of 4.14 ds.m^{-1}. Soil of the studied area was non-calcareous except in some locations. The soil content of $CaCO_3$ ranged between Nil to 20.00 % with an average of 1.96 %.

Table 2: Descriptive statistics of soil parameters.

Statistical parameters	pH	Ece (ds/m)	$CaCO_3$ (%)
Mean	8.50	4.14	1.96
Standard Error (S.E)	0.07	0.42	0.67
Median	8.41	3.10	0.00
Mode	7.95	1.20	0.00
Standard Deviation (S.D)	0.49	2.85	4.60
Sample Variance	0.24	8.13	21.17
Kurtosis	0.30	-0.16	5.90
Skewness	0.84	0.88	2.55
Range	2.02	10.20	20.00
Minimum	7.73	0.80	Nil
Maximum	9.75	11.00	20.00

3.2. Soil hyperspectral signature

From the figure (2), it was shown that reflectance spectra of soil samples followed the same basic shape as observed by many researchers, with prominent absorption bands around 1400, 1900, and 2200 nm (Shepherd and Walsh 2002). These bands are associated with clay minerals, for example, OH features of free water at 1400 and 1900 nm, and lattice OH features at 1400 and 2200 nm (Hunt 1980).

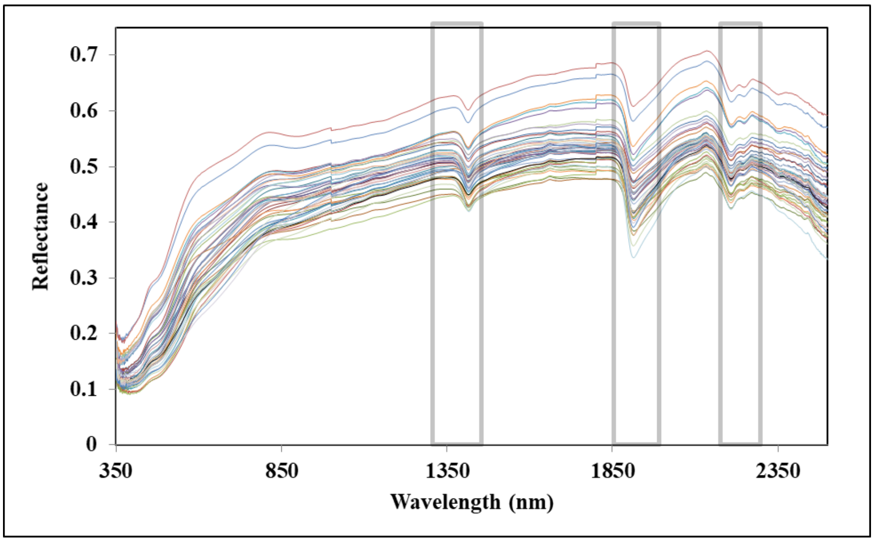

Fig 2: Soil spectral signature of soil samples.

3.3. PLSR-ASD modeling

Table (4) showed the obtained results of PLSR-ASD modeling for the calibration and validation model using the data set of the study area. The obtained data of the validation model were plotted against the measured data of the predicted soil parameters and shown in figure (3).

Table 4: The performance assessment of the PLSR calibration and validation model of ASD data of the study area.

Parameter	Calibration			Validation		
	R^2	RMSE	RPD	R^2	RMSE	RPD
PH	0.92	0.14	3.65	0.69	0.27	1.70
ECe (ds/m)	0.82	1.10	2.39	0.31	2.72	1.20
$CaCO_3$ (%)	0.99	0.59	8.52	0.83	1.83	2.06

From the obtained data, the PLSR calibration model was performing well for all soil parameters whereas values of R^2 >0.50 and RPD >1.40. The R^2 values of calibration were 0.92, 0.82 and 0.99, while RPD values of calibration were 3.65, 2.39 and 8.52 for pH, ECe and $CaCO_3$, respectively.
For the validation model, soil pH and soil $CaCO_3$ showed high predictability using the PLSR model. The values of R^2 were 0.69 and 0.83 while RPD values were 1.70 and 2.06 for these soil parameters, respectively. These results are consistent with the findings of Miloš and Bensa (2017) [16] for $CaCO_3$ and Rossel *et al.* (2006) [1] for soil pH parameters. In the case of the ECe parameter, the R^2 and RPD were 0.31 and 1.20, respectively. The PLSR could not predict the ECe parameter in a good way. Similar findings were recorded by Islam *et al.* (2003) [4].
According to the developed criteria of Chang *et al.* (2001) [19] of the ability of Vis-NIR spectra to predict soil properties, $CaCO_3$ soil parameter was under the 'A' category whereas R^2 values between 0.80 and 1.00 and RPD values more than 2.00. The soil $CaCO_3$ be well predicted using the PLSR prediction model. Soil pH parameter was under the 'B' category which R^2 values between 0.50 and 0.80, and RPD values between 1.40 and 2.00. These soil parameters can be moderately predicted using the PLSR prediction model. The ECe soil parameter could be under the 'C' category whereas R^2 less than 0.50 and RPD less than 1.40. This parameter cannot be predicted with a good result.

4. Conclusion

Partial Least Square Regression model was applied for developing the calibration and validation models for predicting the different soil parameters. Soil parameters (i.e. pH and $CaCO_3$) were predicted well by the PLSR prediction while the ECe parameter was having low predictability. Hyperspectral reflectance data in the range of Viz-NIR (350-2500nm) which integrated with the partial least square regression PLSR model as an empirical technique, showed promising performance for soil parameters' prediction. Further studies can be done with an application of several algorithms to enhance the prediction of soil parameters.

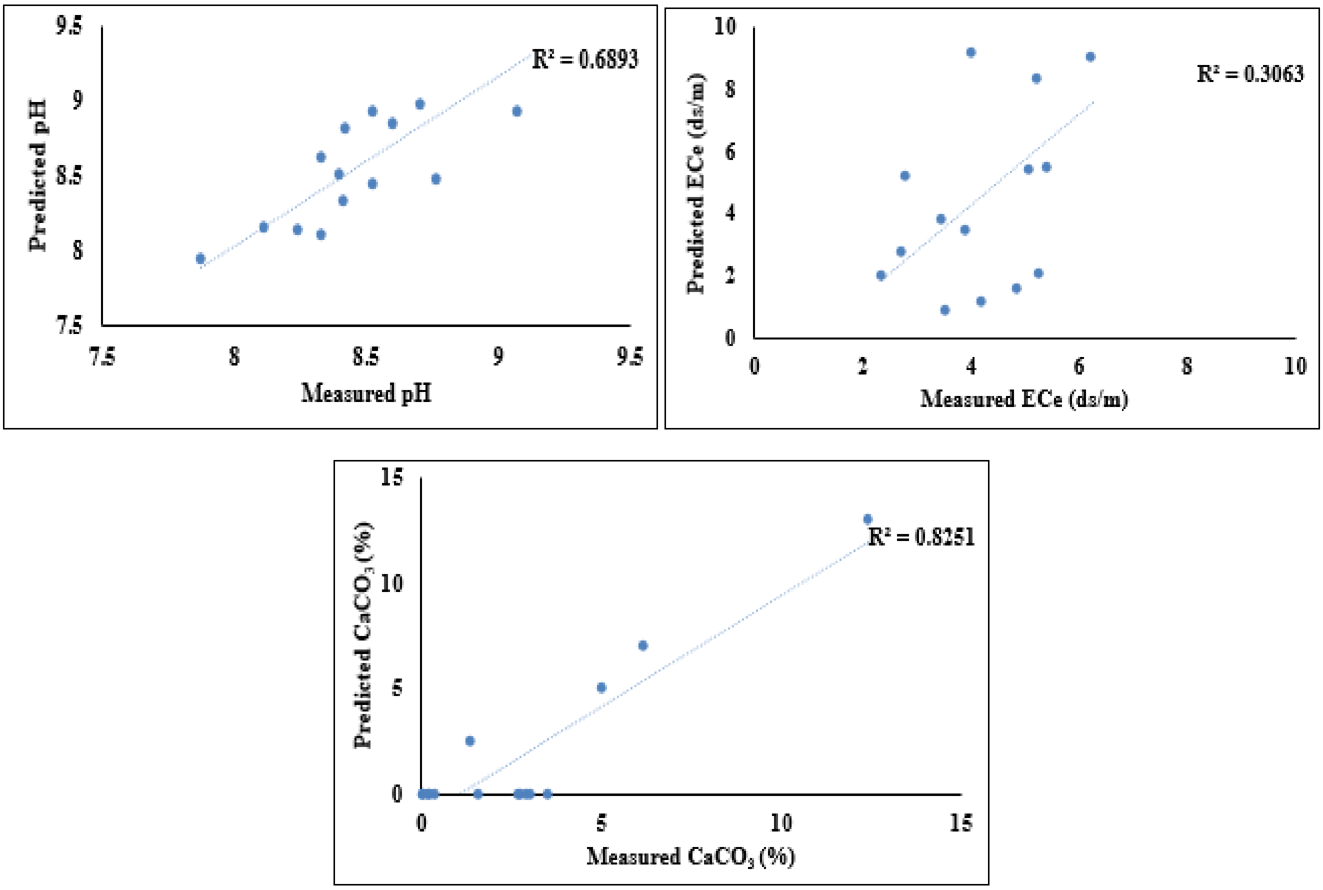

Fig 3: The measured values plotting against the predicted values of soil parameters using the PLSR prediction model; pH, ECe and $CaCO_3$.

5. References

1. Rossel RV, Walvoort DJ, McBratney AB, Janik LJ, Skjemstad JO. Visible, near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Elsevier 2006; 2:59-75.
2. Disla SJ, Janik LS, Rossel RV, MacDonald LM, McLaughlin MJ. The Performance of Visible, Near, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. Applied Spectroscopy Reviews 2014; 2:139-186.
3. Demattê AM, Marcelo RV, Bruna CG, Caio TF, Marcus VS. Hyperspectral remote sensing as an alternative to estimate soil attributes. Revista Ciência Agronômica 2015; 2:223-232.
4. Islam K, Singh B, McBratney AB. Simultaneous estimation of various soil properties by ultra-violet, visible and near-infrared reflectance spectroscopy. Australian Journal of Soil Research 2003; 41:1101-1114.
5. Kadupitiya HK, Sahoo RN, Ray SS, Chakraborty D, Ahmed N. Quantitative assessment of soil chemical properties using visible (VIS) and near-infrared (NIR) proximal hyperspectral data. Tropical Agriculture. 2010; 158:41-60.
6. Jackson ML. Soil Chemical Analysis. Prentice-Hall of India Pvt. Ltd., New Delhi 1967, 498.
7. Martens H, Næs T. Multivariate Calibration. John Wiley and Sons, Chichester 1989, 419.
8. Ogen Y, Zaludab J, Francosb N, Goldshlegerc N, Ben-Dor E. Cluster-based spectral models for a robust assessment of soil properties. Geoderma. 2019; 340:175-184.
9. Chabrillat S, Ben-Dor E, Rossel RV, Dematte JA. Quantitative soil spectroscopy. Applied and Environmental Soil Science. 2013. https://doi.org/10.1155/2013/616578.
10. Stenberg B. Effects of soil sample pretreatments and standardized rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma. 2010; 158:15-22.
11. Wold S, Antti H, Lindgren F, Ohman J. Orthogonal signal correction of near-infrared spectra. Chemom. Intell. Lab. Syst. 1998; 44:175-185.
12. Woodcock CE. Uncertainty in Remote Sensing. Uncertainty in Remote Sensing and GIS. Chichester, UK: John Wiley and Sons 2006, 19-24.
13. Srivastava R, Prasad J, Saxena R. Spectral reflectance properties of some shrink-swell soils of Central India as influenced by soil properties. Agropedology. 2004; 14:45-54.
14. Fikrat F, Maharram B, Simone P, Giovanni L. Using Visible-Near Infrared Spectroscopy to Predict Soil Properties of Mugan Plain, Azerbaijan. Open Journal of Soil Science 2016; 3:52-58.
15. Ostovari Y, Ghorbani-Dashtaki S, Bahrami HA, Abbasi M, Dematte JM, Arthur E *et al*. Towards prediction of soil erodibility, SOM and CaCO3 using laboratory Vis-NIR spectra: A case study in a semi-arid region of Iran. Geoderma. 2018; 314:102-112.
16. Miloš B, Bensa A. Estimation of calcium carbonate in anthropogenic soils on Flysch deposits from Dalmatia (Croatia) using Vis-NIR spectroscopy. Agriculture & Forestry/Poljoprivreda i Sumarstvo, 2017, 63(4).
17. Williams DE. A Rapid Manometric Method for Determination of Calcium Carbonate in Soil. Soil

Science Society of America. 1949; 13:127-129.
18. Liu WD, Baret F, Gu XF, Tong QX, Zheng LF, Zhang B. Relating soil surface moisture to reflectance. Remote Sens. Environ. 2002; 81:238-246.
19. Chang CW, Liard DA, Mausbach MJ, Hurburgh CR. Near-infrared reflectance spectroscopy-principal components regression analysis of soil properties. Soil Science Society of American Journal. 2001; 65:480-90.
20. Hunt GR. Spectroscopic properties of rocks and minerals. In 'Handbook of physical properties of rocks'. Ed. CR Stewart, CRC Press: Boca Raton, FL, 1980, 295-385.
21. Shepherd KD, Walsh MG. Development of reflectance spectral libraries for characterization of soil properties. American Journal of Soil Science Society. 2002; 66:988-98.