

RESEARCH

Open Access



# Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling

Amol D. Vibhute<sup>1\*</sup>, Karbhari V. Kale<sup>2</sup>, Suresh C. Mehrotra<sup>2</sup>, Rajesh K. Dhumal<sup>1</sup> and Ajay D. Nagne<sup>1</sup>

## Abstract

**Introduction:** An accurate and reliable detection of soil physicochemical attributes (SPAs) is a difficult and complicated issue in soil science. The SPA may be varied spatially and temporally with the complexity of nature. In the past, SPA detection has been obtained through routine soil chemical and physical laboratory analysis. However, these laboratory methods do not fulfill the rapid requirements. Accordingly, diffuse reflectance spectroscopy (DRS) can be used to nondestructively detect and characterize soil attributes with superior solution. In the present article, we report a study done through spectral curves in the visible (350–700 nm) and near-infrared (700–2500 nm) (VNIR) region of 74 soil specimens which were agglomerated by farming sectors of Phulambri Tehsil of the Aurangabad region of Maharashtra, India. The quantitative analysis of VNIR spectrum was done.

**Results:** The spectra of agglomerated farming soils were acquired by the Analytical Spectral Device (ASD) Field spec 4 spectroradiometer. The soil spectra of the VNIR region were preprocessed to get pure spectra which were the input for regression modeling. The partial least squares regression (PLSR) model was computed to construct the calibration models, which were individually validated for the prediction of SPA from the soil spectrum. The computed model was based on a correlation study between reflected spectra and detected SPA. The detected SPAs were soil organic carbon (SOC), nitrogen (N), soil organic matter (SOM), pH values, electrical conductivity (EC), phosphorus (P), potassium (K), iron (Fe), sand, silt, and clay. The accuracy of the PLSR model-validated determinant ( $R^2$ ) values were SOC 0.89, N 0.68, SOM 0.93, pH values 0.82, EC 0.89, P 0.98, K 0.82, Fe 0.94, sand 0.98, silt 0.90, and clay 0.69 with root mean square error of prediction (RMSEP) 3.51, 4.34, 2.66, 2.12, 4.11, 1.41, 4.22, 1.56, 1.89, 1.97, and 9.91, respectively. According to the experimental results, the VNIR-DRS was better for detection of SPA and produced more accurate predictions for SPA.

**Conclusions:** In conclusion, the methods examined here offered rapid and novel detection of SPA from reflectance spectroscopy. The outcome of the present research will be apt for precision farming and decision-making.

**Keywords:** VNIR reflectance spectroscopy, First-derivative transformation, Soil physicochemical attributes, Partial least squares regression (PLSR), Spectral pre-treatment, Nondestructive technique, Soil spectral features

\* Correspondence: [amolvibhute2011@gmail.com](mailto:amolvibhute2011@gmail.com)

<sup>1</sup>Geospatial Technology Research Laboratory, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra 431004, India  
Full list of author information is available at the end of the article

## Introduction

An accurate and advanced detection of soil physicochemical attribute (SPA) is a tricky issue in soil science which is requisite for precision farming, environmental monitoring, decision-making, and planning. However, soils are more heterogeneous and dynamic in nature than air and water. In addition, the structure, study and processes of soils are also complicated and formidable both spatially and temporally. With reference to the past experiences, the customary laboratory investigations were made to detect the soil chemical and physical attributes using hazardous compounds. However, these laboratory methods are time-consuming, expensive, and labor-intensive. Consequently, there is a universal need to develop speedy and less expensive methods to detect the SPA (Anne et al. 2014; Bilgili et al. 2010; Das et al. 2015; Rossel et al. 2006, 2016; Vibhute and Gawali 2013). Accordingly, neoteric research has suggested that the usage of VNIR (350–2500 nm)-DRS could be an inexpensive prognostication of SPA that may be used to improve or replace the customary ways of soil analysis overcoming some of the limitations. The VNIR-DRS is productive, rapid, non-wasting, non-destructive, and more accurate as compared to customary ways. The in situ method provides information about the SPA without perilous compounds (Awiti et al. 2008; Bilgili et al. 2010; Brown et al. 2006; Dematté et al. 2010; Rossel et al. 2006, 2016; Srivastava et al. 2017; Udelhoven et al. 2003; Zornoza et al. 2008). In addition, a single spectrum of soils provides information about the various SPA and the methods are adaptable for on-the-go at the field (Rossel et al. 2006). Nevertheless, single reflectance spectra of soil may be diminished by the physicochemical attributes of the soil and the controlled laboratory conditions may provide variable physicochemical attributes of soils which are difficult to determine directly (Ben-Dor et al. 1999, 2002; Ben-Dor and Banin 1994).

The digital soil exploration and its physicochemical attributes detection along with appraisal of it is somewhat formidable task due to availability of various SPA in soils. The varied effect of various attributes of planet surface that can affect the reflectance spectra of soils. The premise of detecting and estimating SPA using VNIR-DRS analysis under laboratory conditions was extensively tested by scientists with some limits. Over the past few decades, many scientists have demonstrated that reflectance spectra of soil in the VNIR region can detect and predict the SPA along with its classification and mapping with hyperspectral datasets (Anne et al. 2014; Ben-Dor et al. 2002; Vibhute et al. 2015). For instance, Rossel et al. (2016) developed a soil spectral library and predicted some soil properties such as soil organic and inorganic carbon, pH, sand, silt, clay, cation exchange capacity (CEC), and iron contents of all land

cover soils. The machine learning approach was used to predict the soil attributes (Rossel et al. 2016). Spectral models were developed based on PLSR to estimate the soil properties at densely covered regions with coastal vegetations. The predicted soil attributes were SOM, labile carbon, labile nitrogen, silt, and clay with good accuracy (Anne et al. 2014). VNIR and mid-infrared or combined DRS were compared simultaneously to determine the combined information generated improved outcomes of soil properties than each of the individual regions. The PLSR model was implemented to calculate calibration models for extraction of soil properties such as soil pH<sub>Ca</sub>, pH<sub>w</sub>, CEC, lime requirements, organic carbon, exchangeable calcium (Ca) and aluminum (Al), nitrogen, phosphorous, potassium, sand, clay, silt, and EC. The results were varied according to the reflectance regions of spectroscopy for the predictions of individual attributes (Rossel et al. 2006). The raw reflectance spectra of soils may give the unacceptable results due to errors found in data while recording the spectra by the devices. Consequently, spectral data of soil can transfer with derivatives and smoothing methods to reconstruct the raw data. The study has been carried out by (Bilgili et al. 2010) using the first-order derivative transformation (FDT) and Savitzky–Golay (SG) methods along with PLSR and multivariate adaptive regression splines (MARS) models for obtaining soil properties.

The methodology is well established in soil science to predict the soil properties with various methods using VNIR-DRS regions for various soil groups. However, the spatial and temporal variability of soils in addition to methods used to detect and predict the soil properties have several limitations to some extent. Moreover, the literature reveals the less attention in the farming sectors for rapid detection of spatial information of SPA which allows management and planning to use fertilizers to the crop growth while increasing the farming productivity with decreasing risks of farming droughts.

Under this constraint, efforts have been made to detect and predict the SPA from the farming soils. The goal of this study is to develop statistical models which predict soil attributes belong to the nutrients and soil textures which are imperative to crop growth from the farming sectors. The main objectives of the study were (1) to study the VNIR-DRS regions to detect and predict the SPA from farming sites, (2) to reconstruct the raw spectra by pre-treatment methods (Savitzky–Golay and derivatives), and (3) to develop PLSR-based statistical models from VNIR-DRS regions which predict soil attributes belong to the nutrients and soil textures. As per literature, very few studies have been done for SPA from farming sites for the region. The present manuscript is arranged in four sections. The first section introduces the challenges of the related study with advancement of VNIR-DRS hyper

spectral remote sensing datasets with background study. The second section highlights the geographic location of the studied site, soil sampling and chemical analysis strategy, spectral measurements by the ASD instrument, spectral pre-treatment with its renovation, statistical modeling, chemometrics analysis using PLSR model, and its precision evaluation. Experimental results were explained in the third section with detailed discussion. Conclusion and future scope have been given in the fourth section.

## Materials and Methods

### The study site

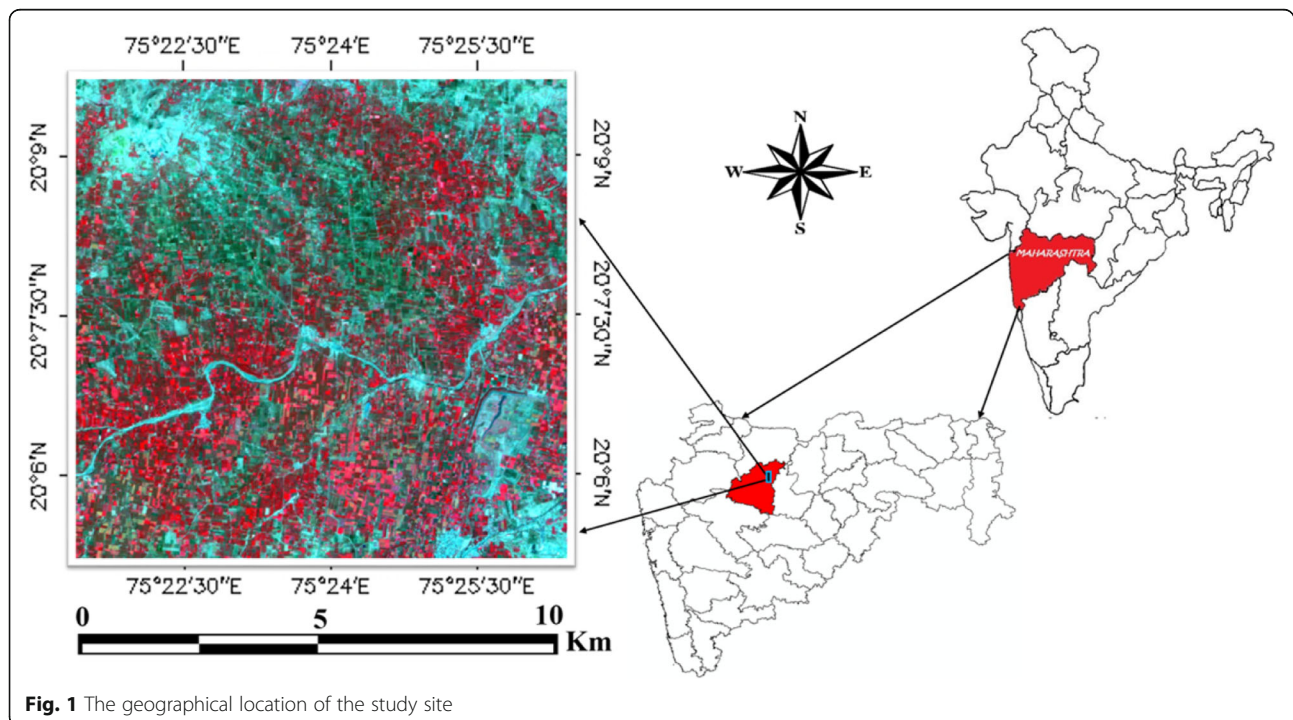
The study site (Fig. 1) was selected which covers near about 72 km<sup>2</sup> area of Kanhori, Pimpalgaon Walan, Pal, and Wanegaon rural areas in Phulambri Taluka of Aurangabad district, Maharashtra, India. The soil specimens were agglomerated from a study region on the experimental farming sectors of said region geologically to be located at 19°28′43.27″–20°24′52.19″ N latitude and 75°13′10.75″–75°30′14.87″ E longitude. The study site has a complex nature with hills (Vibhute et al. 2016). The site has a semiarid atmosphere at an elevation of 580 m above the sea level with an average annual temperature as 17 to 33 °C and an annual average precipitation of 710 mm. The surface soil type of the site is black cotton soil and mostly textured as sandy loam in the studied site.

### Soil sampling and chemical analyses

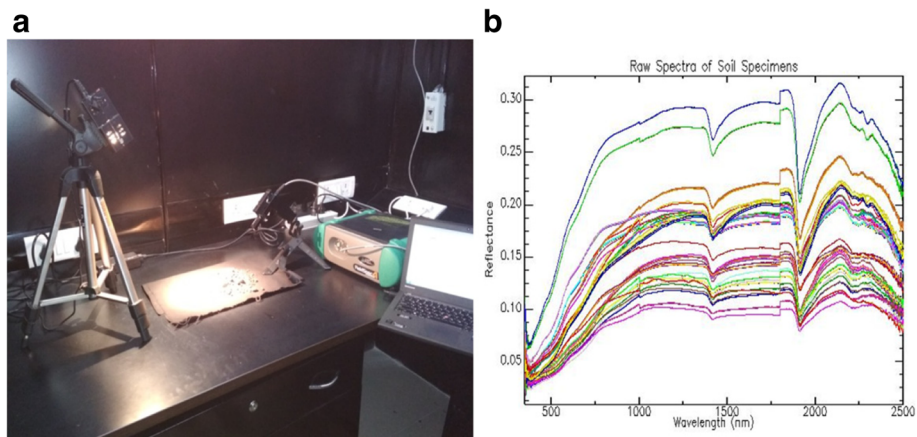
The topsoil (0–20 cm) soil specimens were agglomerated from 72-km<sup>2</sup> farming sectors in airtight containers. The

soil sampling was implemented during the period of 10 February–25 March, 2015, just after the harvest of cotton, wheat, and Jowar crops in between 0800 to 1330 h with clear climate. The surface crop/plant relics were removed by hands while agglomerating the specimens in airtight containers. An exhaustive 74 specimens was agglomerated, air-dried, and passed through a 2-mm sieve. The specimens were uniformly alienated into two portions which were used for determining the reflected spectra of soils and SPA, respectively. The soil in the site consisted of mainly deep and medium black cotton with sandy loam texture (Balpande 2013). A number of 15 soil specimens were used for calibration set and the rest of the 59 specimens were used for a validation set from the random 74 specimens.

The components of SPA for our study were SOC, N, SOM, pH values, EC, P, K, Fe, sand, silt, and clay. The chemical analyses of soil attributes were performed at the “MIT Soil and Water Testing Laboratory”, Aurangabad, Maharashtra, India. The SPAs were analyzed by standard laboratory methods. Soil particle size (soil textures) distributions were determined by the Bouyoucos hydrometer method (Bouyoucos 1927). The potassium dichromate method was used for the analysis of SOM content. Soil pH was determined in water with 1:2 (soil:water ratio) using a glass electrode pH meter (Eckert and Sims 1995) and EC in soil using a conductivity meter (Jackson 1973). The K concentrations of the soil extracts were calculated by a flame photometer (Jackson 1973). The SOC and N were determined by the Thermo Quest EA 1112 elemental



**Fig. 1** The geographical location of the study site



**Fig. 2** **a** The ASD Field spec 4 spectroradiometer with optical setup and **b** acquired raw spectra of all 74 specimens of soils

analyzer. P was analyzed by calorimetrically (molybdenum blue) method.

#### Spectral measurements using the ASD

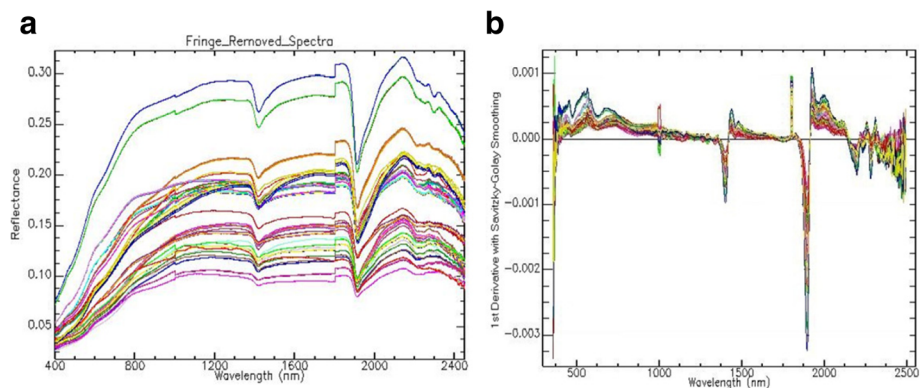
The reflected spectra of each soil specimen were measured using the ASD Field Spec 4 spectroradiometer (Fig. 2a) (<https://www.asdi.com/>) at VNIR wavelengths ranging from 0.35 to 2.5  $\mu\text{m}$ , yielding a total of 2151 data values per spectrum. The measuring was instantly after the specimen agglomeration from farming sectors. The instrument has high spectral resolution with 3 nm for 350–1000 nm and 10 nm for 1000–2500 nm along with sampling intervals 1.4 and 2 nm for each range. Specimens were placed at the bottom of a 4-cm diameter Petri dish using a high-intensity source probe. This instrument uses a 75-W tungsten quartz halogen filament lamp as a light source for illuminating the specimens. The lamp was set as 60° of zenith angle at a distance of 45 cm above the soil specimen. The reflected light was collected in 1-nm bandwidths between 0.35 and 2.5  $\mu\text{m}$  with 8° field-of-view (FOV) fiber-optic cable of spectroradiometer that was kept off-nadir at a distance of 15 cm above the

soil specimen. The instrument was optimized and calibrated using white spectralon for obtaining absolute reflectance for controlled laboratory readings before specimens were recorded (Hatchell 1999). The average of 10 spectra was recorded at each specimen to minimize noise produced by the instrument for obtaining the final spectra. The RS3 (version 6.3) inbuilt software was used to record the reflected spectra.

#### Spectral pre-treatment and renovation

The fringe spectra 350–399 nm and 2451–2500 nm were omitted prior to performing statistical analysis as these were found to be background or low-signal-to-noise produced by the instrument, leaving 2051 waveband predictors between 400 and 2450 nm (Fig. 3a) (Hatchell 1999; Rossel et al. 2006). The reflectance spectra were transferred from .asd file to ASCII and exported in batches by View Spec Pro (6.0.11) designed by ASD Inc.

The spectral data were renovated by FDT (Savitzky and Golay 1964) with SG smoothing in Unscrambler X 10.5 software CAMO, Norway. The mathematical formation of SG smoothing method is formulized by Eq. 1.



**Fig. 3** **a** Fringe removed spectra and **b** FDT with SG smoothing of all soil specimens of the study

$$A_i, SG = \frac{\sum_{j=-m}^{j=m} C_j \cdot A_i + j}{N} \tag{1}$$

where  $A_i$ ,  $SG$  is the smoothed and  $A_i$  is the original reflectance value,  $C_j$  is the filter coefficients,  $j$  is the running index of the original data in the original data table and  $N$  is the number of convoluting integers (Savitzky and Golay 1964).

The mathematical treatments of the SG transformation were 2, 1, 2, and 5, which refer to the order of polynomial, order of derivative, first smoothing point, and second smoothing point, respectively. The FDT and SG smoothing was used to process all the spectral data to smooth the spectra and to eliminate the particle size effects and noise produced by illumination variations (Bilgili et al. 2010; Tsai and Philpot 1998). It is found to be an optimal spectral pre-treatment in alike studies (Awiti et al. 2008; Bilgili et al. 2010; Srivastava et al. 2017). The SPAs were detected by the first derivative transferred spectra which generally augment the absorption features. Finally, combination of FDT with SG smoothing method was used in predicting the soil attributes.

**Statistical modeling**

A number of 74 specimens was randomly separated into two subsets viz. calibration (15) and validation (testing) (59) datasets. The predictions of SPA were based on cross-validation of all specimens used in the calibration and afterward to test the precision of predictions with validation datasets. The correlation studies were performed between SPA and their respective predictor variables (spectral reflectance) using the statistical analysis. The calibration between reflectance of soil and its SPA were carried out in Unscrambler X 10.5 software CAMO, Norway, via PLSR-based chemometric analysis approach.

**Chemometric analyses**

The renovated spectra by pre-treatment methods were used to implement PLSR-based models with leave-one-out cross-validation and were used to calibrate the spectral data with the reference (laboratory) data of soil.

**Partial least squares regression (PLSR)-based modeling**

PLSR is a widely acceptable modeling method in chemometrics and is normally used in quantitative reflectance spectroscopy data analyses. The PLSR is a linear multiple regression method which merges and generalizes the features from multiple regression and principal component regression. The two matrices X and Y are relatively used using a linear multivariate model. It is also used to construct the predictive model using highly correlated spectral variables while reducing the

unwanted spectral bands. The PLSR decomposes the X and Y variables with finding new latent variables and selects successive orthogonal factors that maximizes the covariance between predictor (X-soil reflectance) and response variables (Y-measured SPA by laboratory data). In the fitting of the PLSR model, the selection of the number of the latent variables is critical to prevent the data that explain most of the variation in both predictors and responses. Consequently, proper fitting of the data is very important which were obtained by cross-validation (Bilgili et al. 2010; Rossel et al. 2006; Qiao et al. 2017). The basic idea behind the PLSR model is to look for components  $T$  that allow us to decompose the block of predictors (Eq. 2),

$$X = T \cdot P^T + Residuals(E) \tag{2}$$

and predict the response variables (Eq. 3),

$$Y = T \cdot C + Error(f) \tag{3}$$

where X and Y are the mean-centered matrices that contain the predictor and response variables, respectively.  $P$  and  $C$  are the factor loadings, and  $E$  and  $f$  are the matrices of residuals and errors, respectively (Rossel et al. 2006; Lin et al. 2016).

**Prediction accuracy assessment of models**

The prediction accuracy of the models was used for each soil attributes based on PLSR method and was tested using coefficient of determination ( $R^2$ ) values, RMSEP, and ratio of performance to deviation (RPD) by Eqs. 4, 5, and 6, respectively. The best model was chosen based on higher values of  $R^2$  and lower values of RMSEP. The lower the RMSEP and the higher the  $R^2$  indicate a more accurate and stable model.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{meas} - Y_{pred})^2}{\sum_{i=1}^n (Y_i - Y_{mean})^2} \tag{4}$$

$$RMSEP = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{pred} - Y_{meas})^2} \tag{5}$$

$$RPD = SD_{val} / RMSEP \tag{6}$$

where  $Y_{meas}$  is the measured value (SPA),  $Y_{pred}$  is the predicted theoretical value,  $Y_{mean}$  is the mean of measured values,  $n$  is the number of measured/predicted values with  $i = 1, 2, 3, \dots, n$ ,  $SD_{val}$  is standard deviation of measured values in the validation set, and RMSEP is the root mean square error of prediction of validation dataset (Bilgili et al. 2010; Srivastava et al. 2017; Zornoza et al. 2008).

**Table 1** Statistical summary of the SPA analyzed using conventional laboratory methods

Soil attributes	Min	Max	Mean	Std. deviation
SOC (g kg <sup>-1</sup> )	0.12	0.16	0.13	0.72
N (mg kg <sup>-1</sup> )	212.23	250.88	226.66	14.98
SOM (g kg <sup>-1</sup> )	9.12	134.87	50.46	39.09
pH values	7.14	8.42	7.95	0.35
EC (d Sm <sup>-1</sup> )	0.25	0.46	0.36	0.07
P (mg kg <sup>-1</sup> )	21.27	81.29	37.41	15.26
K (mg kg <sup>-1</sup> )	773.12	3308.94	1937.06	800.12
Fe (mg kg <sup>-1</sup> )	0.926	2.994	1.85	0.89
Sand (g kg <sup>-1</sup> )	89	765	360	230.37
Silt (g kg <sup>-1</sup> )	97	876	388.06	236.11
Clay (g kg <sup>-1</sup> )	109	398	193	77.63

## Results and discussion

### Raw spectra and pre-treatment

The fringe removed spectra of 74 specimens as discussed in the earlier “Spectral pre-treatment and renovation” section are exposed in Fig. 3a and b. Figure 2b shows the raw spectra derived by the ASD instrument (Fig. 2a) with the wavelength 350–2500 nm of 74 soil specimens which are having unwanted noise in fringe spectra. Consequently, firstly fringe spectra were eliminated (Fig. 3a and b) from the processing to avoid the said noises which are generally present in the wavelength 350–399 nm and 2451–2500 nm. The fringe removed hyper spectral data has also much more spectral channels which is very complicated for computation.

### Chemical composition (laboratory analysis) and correlation of soil attributes

The statistics of the SPA of the processed soil specimens studied under laboratory conditions are given in Table 1.

Table 1 shows that the SOC was ranged between 0.123 and 0.161 g kg<sup>-1</sup> with more soil specimens having fewer

values than 0.13 g kg<sup>-1</sup>. The soil pH values ranged from 7.14 to 8.42 which mean more soil specimens were having less than 7.95 pH values, which indicated that the soil from the studied region was good for farming practices. The total average nitrogen was below than 226 mg kg<sup>-1</sup>, whereas SOM concentration was average 50.46 g kg<sup>-1</sup>. The EC values vary from 0.25 to 0.46 d Sm<sup>-1</sup> where values were less than 0.36 d Sm<sup>-1</sup>. The average values of P, K, and Fe were 37.41, 1937.06, and 1.85 mg kg<sup>-1</sup>, respectively. The textural classes of soils were also calculated in the terms as sand, silt, and clay with average values 230.37, 236.11, and 77.63 g kg<sup>-1</sup>, respectively. The silt concentrations were higher values than sand and clay textures. The textural class silt was varied from 97 to 876 g kg<sup>-1</sup> and sand was varied between 89 and 765 g kg<sup>-1</sup> along with clay varied between 109 and 398 g kg<sup>-1</sup>. The correlation matrix of measured soil attributes is given in Table 2.

The soil attributes such as SOC, N, SOM, pH, EC, P, K, Fe, sand, silt, and clay were correlated to each other with both positive and negative sides. The SOC is strongly correlated with N, SOM, pH, and EC, and negatively correlated to P and Fe. The SOM is positively correlated to pH values, EC, Fe, silt, and clay, and negatively correlated to P, K, and sand. The pH is negatively correlated to EC values, K, and sand, whereas Fe is moderately correlated to sand and silt while negatively correlated to clay. The P is positively correlated to Fe, sand, and clay whereas K is positively correlated to Fe, silt, and clay. Sand is negatively correlated to silt and clay.

### Spectral reflectance characteristics of soil and its relationship with attributes in soil

The soil reflectance curve is affected by the physico-chemical attributes of soil. The major physical attributes of the soil which affect the spectral reflectance curve are soil moisture, soil structure, soil texture, soil color, soil surface conditions, etc. along with soil chemical

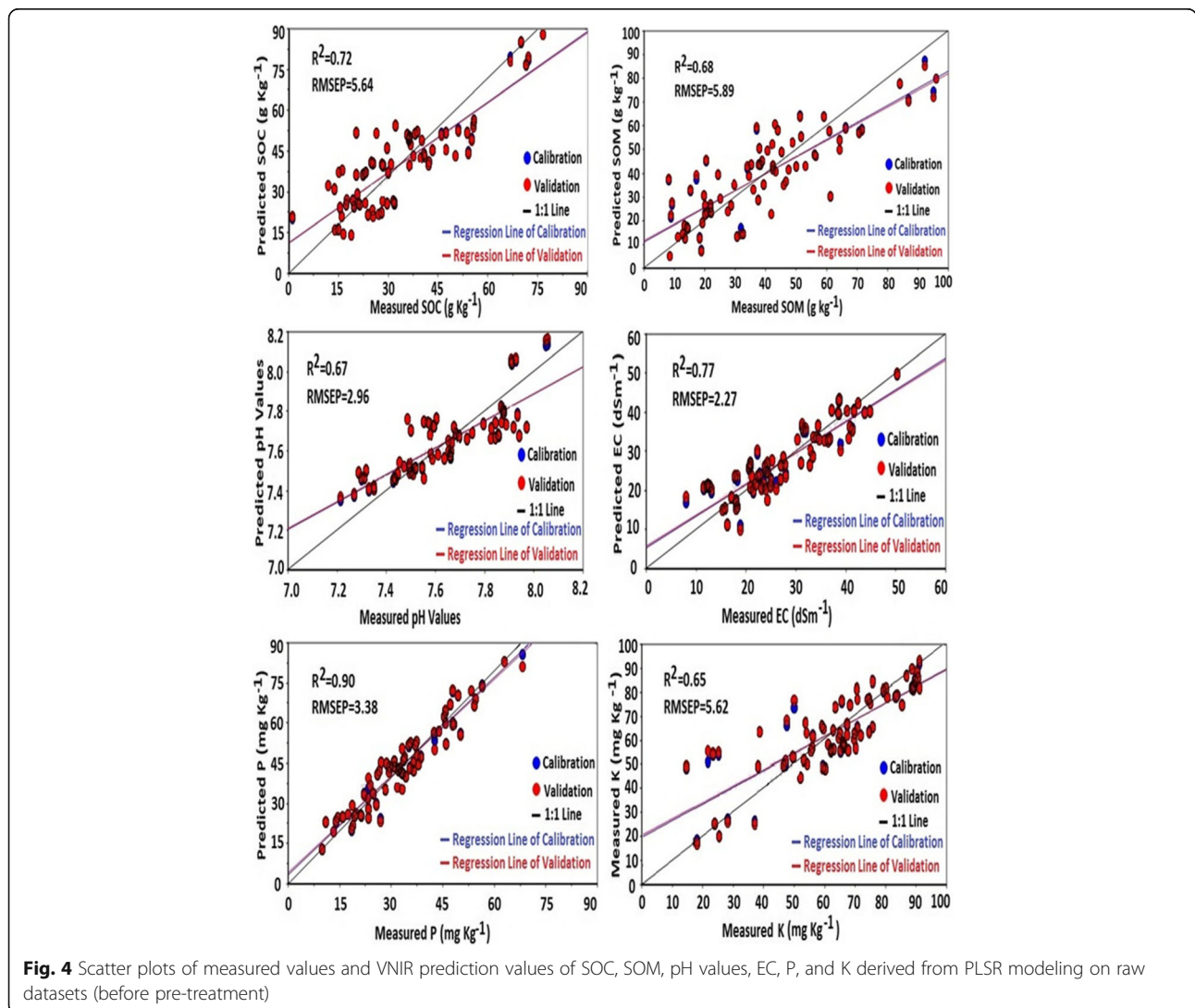
**Table 2** Correlation matrix of measured soil attributes

	SOC	N	SOM	pH	EC	P	K	Fe	Sand	Silt	Clay
SOC	1	0.95	-0.99	0.85	0.86	-0.90	0.90	-0.87	0.88	0.90	0.96
N	0.95	1	-0.97	0.87	0.88	0.88	-0.89	0.85	-0.76	0.93	-0.87
SOM	-0.99	-0.97	1	-0.86	0.89	-0.87	-0.67	0.89	-0.65	0.94	0.96
pH	0.85	0.87	-0.86	1	-0.90	0.84	-0.77	0.86	-0.53	0.97	0.95
EC	0.86	0.88	0.89	-0.90	1	0.87	0.45	0.94	-0.76	-0.76	0.81
P	-0.90	0.88	-0.87	0.84	0.87	1	-0.65	0.76	0.89	-0.85	0.84
K	0.90	-0.89	-0.67	-0.77	0.45	-0.65	1	0.98	-0.91	0.76	0.89
Fe	-0.87	0.85	0.89	0.86	0.94	0.76	0.98	1	0.92	0.78	-0.67
Sand	0.88	-0.76	-0.65	-0.53	-0.76	0.89	-0.91	0.92	1	-0.98	-0.56
Silt	0.90	0.93	0.94	0.97	-0.76	-0.85	0.76	0.78	-0.98	1	0.99
Clay	0.96	-0.87	0.96	0.95	0.81	0.84	0.89	-0.67	-0.56	0.99	1

**Table 3** The statistical summary of the spectral models for the soil attributes using PLSR on raw datasets (before pre-treatment)

Soil attributes	Calibration set		Validation set			
	$R^2$	RMSE	$R^2$	RMSEP	SD	RPD
SOC (g Kg <sup>-1</sup> )	0.72	5.55	0.72	5.64	22.53	3.99
N (mg Kg <sup>-1</sup> )	0.54	3.75	0.48	4.03	10.36	2.57
SOM(g Kg <sup>-1</sup> )	0.71	5.69	0.68	5.89	14.76	2.50
pH Values	0.68	2.87	0.67	2.96	11.02	3.72
EC (d Sm <sup>-1</sup> )	0.80	2.07	0.77	2.27	7.95	3.50
P (mg Kg <sup>-1</sup> )	0.92	3.07	0.90	3.38	5.37	1.58
K (mg Kg <sup>-1</sup> )	0.69	5.35	0.65	5.62	8.77	1.56
Fe (mg Kg <sup>-1</sup> )	0.59	4.38	0.58	4.49	5.24	1.16
Sand (g Kg <sup>-1</sup> )	0.80	8.95	0.80	9.31	8.39	0.90
Silt (g Kg <sup>-1</sup> )	0.87	2.20	0.85	2.42	3.08	1.27
Clay (g Kg <sup>-1</sup> )	0.51	9.28	0.49	9.58	10.69	1.11

attributes SOM, Fe, carbonates, soil minerals, and salinity of soil. The spectral reflectance of soils is decreased by the soil attributes particularly with increase of organic matter, soil water contents, and soil clay contents with Fe contents (Bowers and Hanks 1965; Mitran et al. 2015). Figure 3a and b illustrates the fringe bands removed raw spectra derived by the ASD instrument (Fig. 2a) which was used to carry the present study along with the FDT with the SG smoothing method. The soil reflectance with the fringe removed spectra (Fig. 3a and b) clearly shows that the reflectance curve of spectra was smoothed and reflected very well and noise-free. The visible region (400–700 nm) of the spectrum has lower reflectance (Islam et al. 2003; Shepherd and Walsh 2002) as compared to the near-infrared (1000–2450 nm) region which reflects high for soil. The three notable absorption peaks were identified at 1417-, 1914-, and 2209-nm bands by reflectance spectra as well as FDT-SG methods along with clay concentrations at

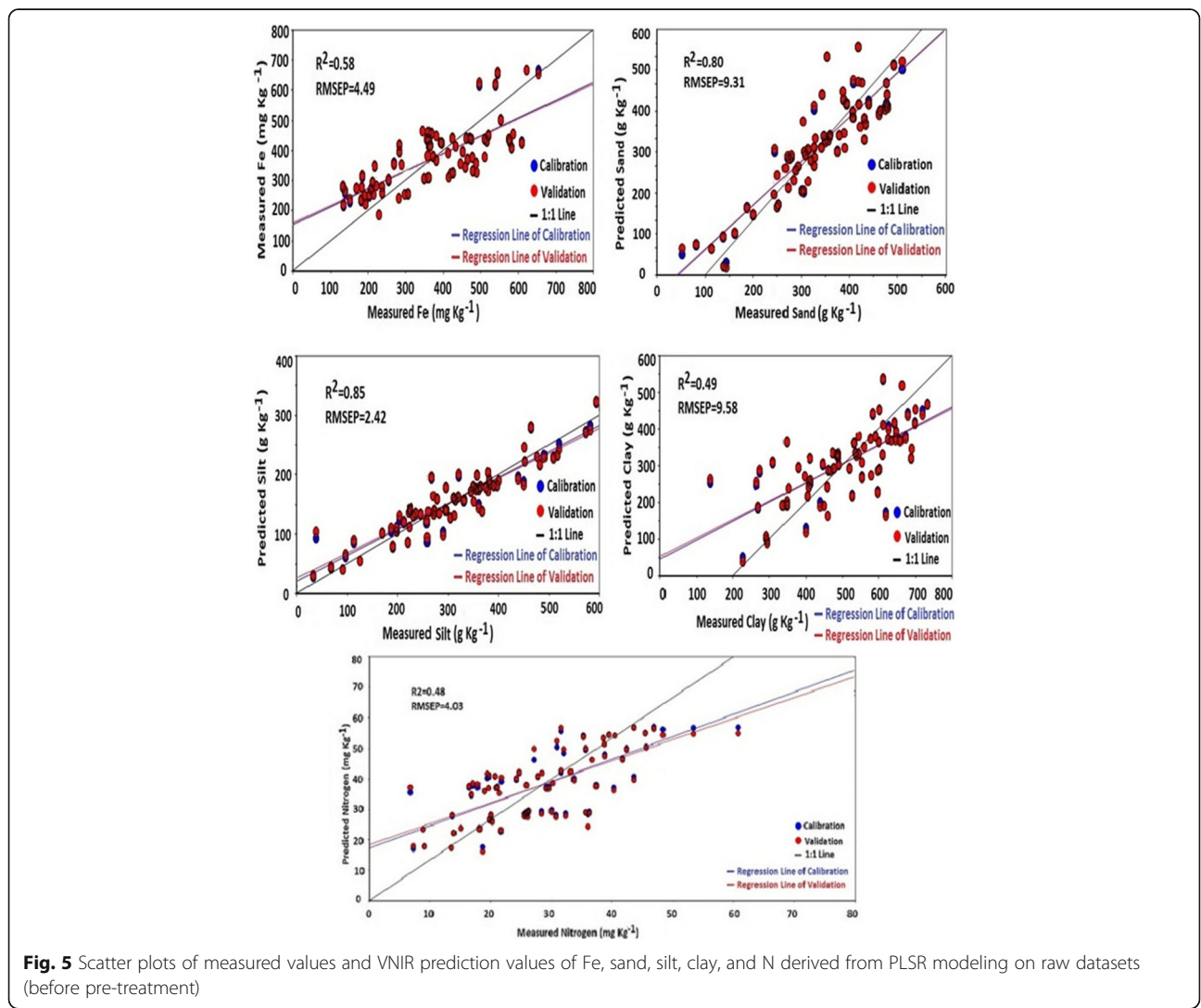


2200–2329 nm (Ben-Dor and Banin 1994; Islam et al. 2003). The absorption peaks at 1400 and 1900 nm are related to water (moisture and hydroxyl ions) and clay lattice OH features at 1400 and 2200 nm (Bilgili et al. 2010; Srivastava et al. 2017). The water absorption peaks were usually enhanced in the FDT with SG smoothing method (Fig. 3a and b). The correlation matrix (Table 2) showed that SOC, N, SOM, pH values, and EC values have significant correlation with the reflectance at 417, 436, 441, 444, 447, 480, 517, 549, 572, 617, 747, 760, 832, 847, and 980 nm. Accordingly, the results show that there are both positive and negative correlations between soil attributes and their reflectance spectra at various wavelengths. The highest correlation for the soil N material ( $R^2 = 0.95$ ) was found at 980 nm. Some of the other high correlations were also observed for pH values ( $R^2 = 0.87$ ), EC values ( $R^2 = 0.89$ ), and SOM ( $R^2 = 0.89$ ). The reference of Wei et al. (2017) reported the positive correlations with reflectance spectra of N contents at

visible wavebands. The visible waveband at 417 nm was determined as the significant wavelength along with 1000 nm (Qiao et al. 2017). The more sensitive wavelengths for N, SOC, and SOM were found at 403, 470, 687, and 846 nm for N; at 400, 409, 441, and 907 nm for SOC; and at 400, 441, 832, and 907 nm for SOM (Gmur et al. 2012).

**Prediction of attributes in soil based on PLSR modeling**

The random cross-validation method along with kernel PLS of PLSR was used on both raw datasets (Table 3, Figs. 4 and 5) and FDT with the SG smoothing datasets (Table 4, Figs. 6 and 7) to test the prediction accuracy for soil attributes. The soil attributes such as SOC ( $R^2 = 0.72$  and RMSE = 5.55), SOM ( $R^2 = 0.71$  and RMSE = 5.69), pH values ( $R^2 = 0.68$  and RMSE = 2.87), EC ( $R^2 = 0.80$  and RMSE = 2.07), P ( $R^2 = 0.92$  and RMSE = 3.07), K ( $R^2 = 0.69$  and RMSE = 5.35), sand ( $R^2 = 0.80$  and RMSE = 8.95), and silt ( $R^2 = 0.87$  and RMSE = 2.20)



**Fig. 5** Scatter plots of measured values and VNIR prediction values of Fe, sand, silt, clay, and N derived from PLSR modeling on raw datasets (before pre-treatment)

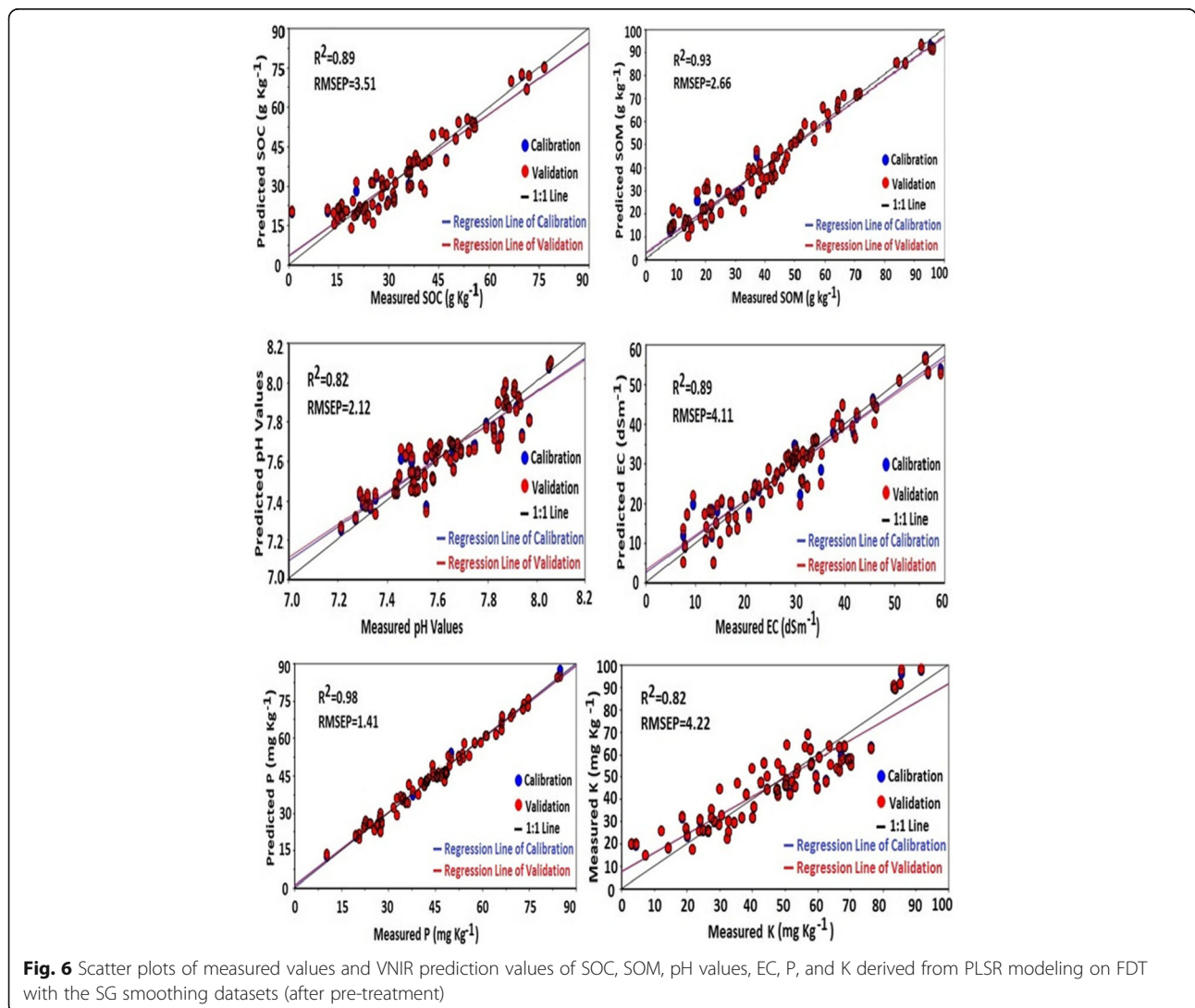


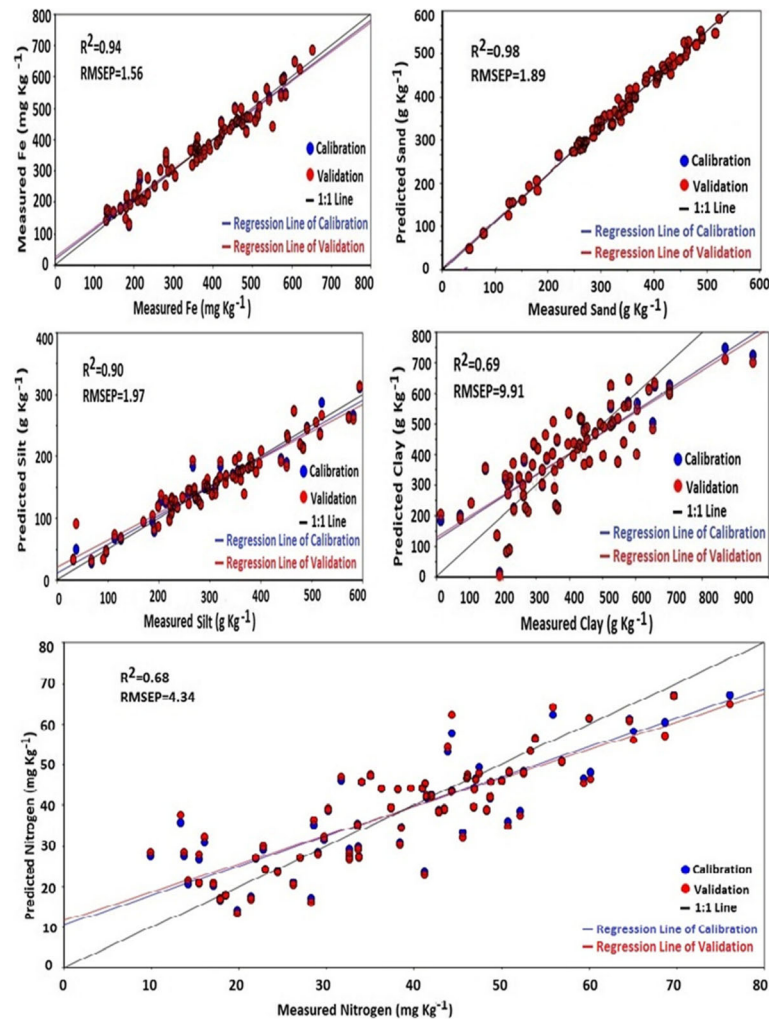
**Table 4** Statistics summary of the spectral models for the soil attributes using PLSR on FDT with the SG smoothing datasets (after pre-treatment)

Soil attributes	Calibration set		Validation set			
	$R^2$	RMSE	$R^2$	RMSEP	SD	RPD
SOC (g kg <sup>-1</sup> )	0.90	3.28	0.89	3.51	7.72	2.19
N (mg kg <sup>-1</sup> )	0.72	3.95	0.68	4.34	9.25	2.13
SOM(g kg <sup>-1</sup> )	0.94	2.47	0.93	2.66	6.15	2.31
pH values	0.86	1.90	0.82	2.12	4.31	2.03
EC (d Sm <sup>-1</sup> )	0.90	3.65	0.89	4.11	5.47	1.33
P (mg kg <sup>-1</sup> )	0.98	1.39	0.98	1.41	2.56	1.81
K (mg kg <sup>-1</sup> )	0.84	4.05	0.82	4.22	6.25	1.48
Fe (mg kg <sup>-1</sup> )	0.95	1.48	0.94	1.56	3.34	2.14
Sand (g kg <sup>-1</sup> )	0.98	1.79	0.98	1.89	3.37	1.78
Silt (g kg <sup>-1</sup> )	0.93	1.62	0.90	1.97	3.43	1.74
Clay (g kg <sup>-1</sup> )	0.70	9.48	0.69	9.91	13.18	1.32

were well calibrated before its pre-treatments. Table 3 describes the statistics resulted from PLSR modeling with its calibration ( $R^2$  values and RMSE) and validation ( $R^2$  values, RMSEP, SD, and RPD) before the use of FDT with the SG smoothing methods. Figures 4 and 5 depict the scatter plot of measured values and VNIR prediction values of all soil attributes derived from PLSR models before its pre-treatment.

Higher regression coefficients ( $R^2$ ) were obtained for all the PLSR models developed for predicting the soil attributes in the calibration sets. The prediction results of raw datasets (before the pre-treatment explained in the “Spectral pre-treatment and renovation” section) obtained satisfactory values for SOC ( $R^2 = 0.72$  and RMSEP = 5.64), SOM ( $R^2 = 0.68$  and RMSEP = 5.89), pH values ( $R^2 = 0.67$  and RMSEP = 2.96), EC ( $R^2 = 0.77$  and RMSEP = 2.27), P ( $R^2 = 0.90$  and RMSEP = 3.38), K ( $R^2 = 0.65$  and RMSEP = 5.62), sand ( $R^2 = 0.80$  and RMSEP = 9.31), and silt ( $R^2 = 0.85$  and RMSEP = 2.42).





**Fig. 7** Scatter plots of measured values and VNIR prediction values of Fe, sand, silt, clay, and N derived from PLSR modeling on FDT with the SG smoothing datasets (after pre-treatment)

The soil attributes such as N, Fe, and clay were moderately calibrated and validated (Table 3) as compared to other soil attributes. Table 4 summarizes PLSR cross-validation statistics for the soil attributes after the use of pre-treatment methods (FDT with the SG smoothing). The PLSR method was evaluated using cross-validation  $R^2$  and RMSEP with RPD (Table 4) for various soil attributes. Figures 6 and 7 demonstrate the scatter plot of measured values and VNIR prediction values of all soil attributes derived from PLSR models after its pre-treatment by FDT-SG methods.

The evaluation of the accuracy for prediction performances of the PLSR models for various soil attributes was estimated based on  $R^2$ , RMSEP, and RPD values achieved in the independent validation set, both of raw datasets and transferred datasets by SG-FDT method (Tables 3 and 4). The value of  $R^2$  demonstrates the

percentage of the variation in the dependent variable that is accounted for the independent variables. The RMSEP value estimates the variation of predicted values around measured values, and as this approaches to zero, prediction performance of the equation improves (Srivastava et al. 2017).

The prediction of accuracy has been improved by the RPD factor as compared to using the mean composition for all samples (Srivastava et al. 2017). Chang et al. (2001) have predicted the quality of soil attributes using VNIR-DRS is based on RPD values. The RPD factor has categorized into three ranges of RPD values as  $> 2.0$ ,  $1.4$ – $2.0$ , and  $< 1.4$  which indicate the excellent, acceptable, and poor predictions, respectively. They recommended that the predictions between the ranges  $1.4$  and  $2.0$  could be gained by various calibration approaches. On the other hand, the properties of the ranges  $< 1.4$  may not be reliably predicted using DRS (Chang et al. 2001). As per the

result, the developed calibration models of soil attributes were performed better for RPD values for all soil attributes excluding Fe, sand, silt, and clay on the raw datasets and EC with clay on the FDT-SG datasets as indicated in Tables 3 and 4. The high RPD values of SOC, N, SOM, pH, and EC were 3.99, 2.57, 2.50, 3.72, and 3.50, respectively, for the raw datasets, whereas high RPD values of SOC, N, SOM, pH, and Fe were 2.19, 2.13, 2.31, 2.03, and 2.14, respectively, for FDT-SG datasets. The acceptable RPD values were 1.58 and 1.56 for P and K, respectively, on raw datasets, whereas 1.81, 1.48, 1.78, and 1.74 for P, K, sand, and silt, respectively, on processed datasets. Poor predictions were obtained for Fe, sand, silt, and clay with RPD values between 0.90 and 1.27 on raw datasets and EC (1.33) and clay (1.32) with RPD values on transferred datasets. The calibration and prediction accuracy ( $R^2$  values) of all soil attributes were increased significantly after the pre-treatment by FDT-SG method as the results detailed in the Table 4. The prediction results may vary based on different statistical processing methods and derivatives of spectra (Reeves et al. 2002). The best predictive models were obtained for SOC ( $R^2 = 0.89$  and RMSEP = 3.51), SOM ( $R^2 = 0.93$  and RMSEP = 2.66), pH values ( $R^2 = 0.82$  and RMSEP = 2.12), EC values ( $R^2 = 0.89$  and RMSEP = 4.11), P ( $R^2 = 0.98$  and RMSEP = 1.41), K ( $R^2 = 0.82$  and RMSEP = 4.22), Fe ( $R^2 = 0.94$  and RMSEP = 1.56), sand ( $R^2 = 0.98$  and RMSEP = 1.89), and silt ( $R^2 = 0.90$  and RMSEP = 1.97). The moderate predictions were obtained for N and clay only with  $R^2 = 0.68$  and RMSEP = 4.34 and  $R^2 = 0.69$  and RMSEP = 9.91, respectively. The results found in this study are superior to the previous study by Bilgili et al. (2010) for SOM, K, pH values, EC values, sand, and silt with  $R^2 = 0.93, 0.82, 0.82, 0.89, 0.98,$  and  $0.90,$  respectively. The  $R^2$  values of the present study for SOC (0.89), N (0.68), sand (0.98), silt (0.90), clay (0.69), P (0.98), K (0.82), and EC (0.89) are slightly better than those of the reference (Rossel et al. 2006) which were revealed for the simultaneous assessment of various soil properties using visible, near-infrared, mid-infrared, or combined diffuse spectroscopy. The classification of the soil condition was estimated using infrared spectroscopy and resulted coefficient of determination ( $R^2$ ) of pH values (0.72), sand (0.75) and silt (0.77) which were smaller than the values reported in Awiti et al. (2008) as compared to the values reported here.

## Conclusions

The soil attributes were detected through the VNIR-DRS with soil specimens accumulated from farming sectors. It is concluded that the soil attributes can be detected successfully by the VNIR hyper spectral (DRS) datasets at wavelength 400–2450 nm with superior outcome. Our outcome demonstrated that the potential use of VNIR spectroscopy for

detecting the soil attributes can be used in an effective manner which may be useful in precision farming practices along with farming planning and management. According to the current study, spectral pre-treatment is essential to correlate the soil attributes in a better way. It may be concluded that the pre-treatment by SG-FDT method produced positive response on the robustness of PLSR models. Major soil nutrients were well predicted using hyper spectral VNIR spectroscopy. Our results revealed the importance of VNIR spectroscopy for the measurement of soil properties than the conventional laboratory approaches which are costly, time-consuming, and tedious. The PLSR-based model was developed for detecting the soil nutrients with better accuracy for all attributes with less RMSEP values as mentioned in Tables 3 and 4. The sensitive bands were detected for SOC at wavelengths 436, 444, 447, 549, and 616 nm; for N at wavelengths 480, 980, and 1702 nm; and for SOM at wavelengths 417, 441, 832, and 907 nm. The bands at 517, 657, 747, 1477, and 1492 nm showed good correlations with soil pH values, whereas bands at 572, 847, 992, 1202, 1537, 1622, and 1767 nm for EC values. The K sensitive bands were correlated at wavelengths 535, 1542, 1862, and 1947 nm. In conclusion, the pre-treatment is imperative for developing PLSR models to extract the soil attributes from the VNIR spectrum. The detected soil attributes were analyzed by spectroscopy using the PLSR model which may be useful in precision farming and management. In the future scope, more soil samples will be considered for developing better predictive models. The detected soil attributes will be quantitatively mapped through satellite or airborne hyper spectral imagery for the large area.

## Abbreviations

Al: Aluminum; ASD: Analytical spectral device; Ca: Exchangeable calcium; CEC: Cation exchange capacity; DRS: Diffuse reflectance spectroscopy; EC: Electrical conductivity; FDT: First order derivative transformation; Fe: Iron; FOV: Field of view; K: Potassium; MARS: Multivariate adaptive regression splines; N: Nitrogen; P: Phosphorus; PLSR: Partial least squares regression; RMSEP: Root mean square error of prediction; RPD: Ratio of performance to deviation; SG: Savitzky–Golay; SOC: Soil organic carbon; SOM: Soil organic matter; SPA: Soil physicochemical attribute; VNIR: Visible and near-infrared

## Acknowledgements

The authors would like to acknowledge the UGC for providing BSR Fellowship and lab facilities under UGC SAP (II) DRS Phase-I F.No.-3-42/2009, Phase-II 4-15/2015/DRS-II, DeitY, Government of India, under Visvesvaraya PhD Scheme, DST-MRP-R No. BDID/01/23/2014-HSRS/35(ALG-IV), and also extend their gratitude to the DST-FIST program of the Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S. India. The authors would also like to acknowledge the support of the Ramanujan Geospatial Chair from the NRDMS, New Delhi. We are also thankful to Prof. D. T. Bornare and his team for the physicochemical analysis of the soil specimens at "MIT Soil and Water Testing Laboratory, Aurangabad," Maharashtra, India.

**Availability of data and materials**

All data generated and analyzed during this study are included in this published article.

**Authors' contributions**

ADV proposed the idea and conducted the original study, acquired the vital data, developed the model and performed the experiments, and wrote the manuscript. KVK and SCM critically evaluated the scientific validity of the proposed system and provided technical advice and guide for the final manuscript. RKD and ADN edited the manuscript and performed its final edits. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not Applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Geospatial Technology Research Laboratory, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra 431004, India. <sup>2</sup>Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra 431004, India.

Received: 31 January 2018 Accepted: 25 June 2018

Published online: 19 July 2018

**References**

- Anne NJ, Abd-Elrahman AH, Lewis DB, Hewitt NA (2014) Modeling soil parameters using hyperspectral image reflectance in subtropical coastal wetlands. *Int J Appl Earth Obs Geoinf* 33:47–56.
- Awiti AO, Walsh MG, Shepherd KD, Kinyamario J (2008) Soil condition classification using infrared spectroscopy: a proposition for assessment of soil condition along a tropical forest-cropland chronosequence. *Geoderma* 143(1):73–84.
- Balpande U. S. (2013). Ground water information Aurangabad District Maharashtra, Central Region Nagpur. Govt of India Ministry of Water Resources Central Ground Water Board.
- Ben-Dor E, Banin A (1994) Visible and near-infrared (0.4–1.1  $\mu\text{m}$ ) analysis of arid and semiarid soils. *Remote Sens Environ* 48(3):261–274.
- Ben-Dor E, Irons JR, Epema GF (1999) Soil reflectance. *Man Remote Sens Earth Science* 3:111.
- Ben-Dor E, Patkin K, Banin A, Karnieli A (2002) Mapping of several soil properties using DAIS-7915 hyperspectral scanner data—a case study over clayey soils in Israel. *Int J Remote Sens* 23(6):1043–1062.
- Bilgili AV, Van Es HM, Akbas F, Durak A, Hively WD (2010) Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *J Arid Environ* 74(2):229–238.
- Bouyoucos GJ (1927) The hydrometer as a new and rapid method for determining the colloidal content of soils. *Soil Sci* 23(4):319–331.
- Bowers SA, Hanks RJ (1965) Reflection of radiant energy from soils. *Soil Sci* 100:130–137.
- Brown DJ, Shepherd KD, Walsh MG, Mays MD, Reinsch TG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132(3):273–290.
- Chang CW, Laird DA, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy—principal components regression analyses of soil properties. *Soil Sci Soc Am J* 65(2):480–490.
- Das BS, Sarathjith MC, Santra P, Sahoo RN, Srivastava R, Routray A, Ray SS (2015) Hyperspectral remote sensing: opportunities, status and challenges for rapid soil assessment in India. *Curr Sci* 108(5):860.
- Demattè JA, Fiorio PR, Araújo SR (2010) Variation of routine soil analysis when compared with hyperspectral narrow band sensing method. *Remote Sens* 2(8):1998–2016.
- Eckert D, Sims JT (1995) Recommended soil pH and lime requirement tests. Recommended soil testing procedures for the northeastern United States. *Northeast Reg Bull* 493:11–16.
- Gmur S, Vogt D, Zabowski D, Moskal LM (2012) Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees. *Sensors* 12(8):10639–10658.
- Hatchell D. C., Analytical spectral devices, Field Spec™ user's guide. Inc. (ASD) technical guide, 3rd Ed. 1999.
- Islam K, Singh B, McBratney A (2003) Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust J Soil Res* 41:1193–1202.
- Jackson ML (1973) Soil chemical analysis. Prentice Hall of India, New Delhi.
- Lin L, Wang Y, Teng J, Wang X (2016) Hyperspectral analysis of soil organic matter in coal mining regions using wavelets, correlations, and partial least squares regression. *Environ Monit Assess* 188(2):97.
- Mitran T, Ravisankar T, Fyzee MA, Suresh JR, Sujatha G, Sreenivas K (2015) Retrieval of soil physicochemical properties towards assessing salt-affected soils using hyperspectral data. *Geocarto Int* 30(6):701–721.
- Qiao XX, Wang C, Feng MC, Yang WD, Ding GW, Sun H et al (2017) Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques. *Spectrosc Lett* 50(3):156–163.
- Reeves J, McCarty G, Mimmo T (2002) The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environ Pollut* 116:S277–S284.
- Rossel RV, Behrens T, Ben-Dor E, Brown DJ, Demattè JAM, Shepherd KD et al (2016) A global spectral library to characterize the world's soil. *Earth Sci Rev* 155:198–230.
- Rossel RV, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131(1):59–75.
- Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem* 36(8):1627–1639.
- Shepherd KD, Walsh MG (2002) Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci Soc Am J* 66(3):988–998.
- Srivastava R, Sethi M, Yadav RK, Bundela DS, Singh M, Chattaraj S et al (2017) Visible-near infrared reflectance spectroscopy for rapid characterization of salt-affected soil in the Indo-Gangetic Plains of Haryana. *India J Indian Soc Remote Sensing* 45(2):307–315.
- Tsai F, Philpot W (1998) Derivative analysis of hyperspectral data. *Remote Sens Environ* 66(1):41–51.
- Udelhoven T, Emmerling C, Jarmer T (2003) Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. *Plant Soil* 251(2):319–329.
- Vibhute AD, Dhumal RK, Nagne AD, Rajendra YD, Kale KV, Mehrotra SC (2016) Analysis, classification, and estimation of pattern for land of Aurangabad Region using high-resolution satellite image. In: Proceedings of the Second International Conference on Computer and Communication Technologies. Springer, New Delhi, pp 413–427.
- Vibhute AD, Gawali BW (2013) Analysis and modeling of agricultural land use using remote sensing and geographic information system: a review. *Int J Eng Res Appl* 3(3):81–91.
- Vibhute AD, Kale KV, Dhumal RK, Mehrotra SC (2015) Soil type classification and mapping using hyperspectral remote sensing data. In *Man and Machine Interfacing (MAMI)*, 2015 International Conference on IEEE, pp 1–4.
- Wei Y, Zhu X, Li C, Cheng L, Wang L, Zhao G, Jiang Y (2017) Monitoring soil nitrate nitrogen based on hyperspectral data in the apple orchards. *Agric Sci* 8(1):21–32.
- Zornoza R, Guerrero C, Mataix-Solera J, Scow KM, Arcenegui V, Mataix-Beneyto J (2008) Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol Biochem* 40(7):1923–1930.