Research Article

Estimation of soil nitrogen in agricultural regions by VNIR reflectance spectroscopy

Amol D. Vibhute[1] · Karbhari V. Kale[2] · Sandeep V. Gaikwad[2] · Rajesh K. Dhumal[2]

Abstract

In the present paper, the novel hyperspectral model was developed for the estimation of Soil Nitrogen (SN) in agricultural lands using Partial Least Squares Regression (PLSR) method. In this regard, an effort has been made on predicting and analyzing SN from several agricultural lands of Phulambri Tehsil of Aurangabad district of Maharashtra, India. The spectra of seventy four (74) agricultural soil samples were acquired between 350–2500 nm by Analytical Spectral Device Field Spec-4 Spectroradiometer under controlled laboratory conditions. The preprocessing was done on acquired spectra by First-derivative Transformation (FDT) and Savitzky–Golay (SG) method for getting suitable information. The PLSR approach was derived from correlation analysis between reflectance spectra and SN features. The resulted coefficient of determination (R^2) values was 0.68 and 0.94 before and after pre-treatment with root mean square error of prediction (RMSEP) 4.34 and 1.56, respectively. The identified sensitive wavelength bands of nitrogen content were 480 nm, 511 nm, 653 nm, 997 nm, 1472 nm, 1795 nm, 2210 nm and 2296 nm. In the conclusion, the model is reliable for prediction of SN from agricultural areas. The present research will be useful for decision making in agricultural management.

Keywords Soil nitrogen · Partial least squares regression · First-derivative transformation · Reflectance spectroscopy · Agricultural soil

1 Introduction

Soil properties play a crucial role in efficient farming having various aspects. However, efficient farming practices are possible when soils have sufficient amount of vital macro- and micronutrients. Soil nitrogen is one of the important macronutrient which helps to the growth of crops [1, 2] and healthy practices in agricultural soil. Moreover, the nutrient absorption process of crops/plants is fulfilled by the SNs [1]. Therefore, the precise extraction and estimation of SNs are imperative for precision farming, crop growth and yield estimation, evaluations of soil health, etc. Nevertheless, soils may vary according to spatio-temporal dynamics as compared to water and air. The extractions of soil physicochemical nutrients by conventional methods are risky task due to harmful chemicals. Furthermore, these conventional methods are complicated, laborious, time consuming and expensive. They do not offer the spatio-temporal information of soil nutrients [3–5].

Recently, Remotely Sensed Hyperspectral Reflectance Spectra (RSHRS) has provided valuable source in soil sciences to replace the customary ways of soil study for overcoming several limitations. The RSHRS is generally known as visible-near-infrared (VNIR) reflectance spectroscopy. The RSHRS is more precise, non-wasting, rapid and inexpensive than conventional methods. In addition, the single reflectance spectra of soil provide the additional information of soil physicochemical nutrients [4, 6, 7]. Nonetheless, the physicochemical nutrients of soil may be the reason of

✉ Amol D. Vibhute, amolvibhute2011@gmail.com | [1]School of Computer Science, Central University of Karnataka, Kalaburagi 585367, India. [2]Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MH 431004, India.

reducing the reflectance spectra. Additionally, the measurement of soil reflectance spectra is difficult owing to circumstances of controlled laboratory which gives inconsistent characteristics in soil [8–10].

Therefore, the use of reflectance spectra was mostly carried out for the determination of soil physicochemical nutrients by the research community of soil science. Several soil scientists have focused the usage of RSHRS for predicting the SN from various land sectors. The researchers [1] have used ASD FieldSpec 3 data for monitoring the soil nitrate nitrogen in the apple orchards with FDT and multivariate scatter correction preprocessing methods along with support vector machine. An assessment of various soil properties including soil nitrate nitrogen was done through VNIR-MIR or combined reflectance spectroscopy using PLSR modeling [4]. Similarly, PLSR based model was designed for the modeling of soil parameters like soil labile nitrogen, carbon, organic matter, silt and clay in subtropical coastal wetlands [3]. In addition, several soil properties were mapped using hyperspectral airborne image and ASD data for the prediction of soil attributes [10]. Moreover, quantitative mapping of SN contents was carried out using ASD field Spectroradiometer and Hyperion image with multivariate statistical analysis [2]. However, spatio-temporal inconsistencies of soil groups and methods used for detecting and estimating the SNs have many constrains. Additionally, agricultural soils were often not used for rapid identification of SNs. Under this constraint, an effort has been made on developing a novel model to predict and analyze nitrogen from several agricultural soils. The methodology is found to be better than other reported techniques [1–4, 11].

The present study aims (1) to use the VNIR spectroscopy for SN extraction from agricultural soils, (2) to use the Savitzky–Golay method for eliminating the noise from soil reflectance spectra, (3) to develop the models using PLSR method from RSHRS to extract and predict the SNs and (4) to verify and validate the accuracy of developed models. This paper is organized as follows. Section first gives the background of the study with set objectives for present research. The studied area, laboratory analysis and measurements of reflectance spectra are provided in section two. Section three highlights the methodology along with preprocessing, statistical analysis, development of PLSR models and evaluation of accuracy. The results are discussed in section four. Section five concludes the present study along with its future enhancement.

2 Study area and data used

2.1 Area of interest (AOI)

The agricultural soils of Phulambri Tehsil of Aurangabad District, Maharashtra India were selected as our area of interest. The AOI (Fig. 1) was of size 62.35 km^2 geo-located at 19°28′43.27"–20°24′52.19" N latitude and 75°13′10.75"–75°30′14.87" E longitude. The soil of the present region is black cotton with sandy loam texture. The environmental conditions were semiarid with 580 m elevation above the sea level. The temperature falls between 17° C to 33° C with 710 mm rainfall [12]

2.2 Field campaign and laboratory analysis

The tenure of field campaign was chosen as February–March, 2015, in between 0800 to 1330 (IST) due to clear environmental conditions. The farming sites of AOI were considered for collecting the soil samples from the topsoil's (0–20 cm) in airtight artificial bags. The crops were harvested when soil sampling was done. The number of 74 soil samples has been collected from the studied region. These samples were passed using 2 mm sieve after the air-dried. Subsequently, two groups of samples were similarly generated to predict the reflectance spectra of soil and SNs. The calibration and validation set was done from the random samples on 15 and 59 soil samples correspondingly. The Thermo Quest EA 1112 elemental analyzer was used for analyzing the SNs at standard soil testing laboratory.

2.3 Measurements of reflectance spectra using ASD instrument

The reflectance spectra of soil samples were obtained using ASD field-spec-4 Spectroradiometer instrument [13] in the controlled laboratory conditions. The device provides high spectral resolution at the 350–2500 nm VNIR spectrum. These spectral measurements of soil samples were done instantly after the soil sampling. The 20-cm diameter and 2-cm thickness size Petri dish was used to scan the soil spectra at the bottom of the Field of View (FOV). In our case, we have used 8° off-nadir FOV to scan the samples from the distance of 15 cm above the samples. The soil samples were illuminated using 75 W tungsten quartz halogen lamp. The distance between samples from the lamp was set as 45 cm with 60° of Zenith angle. Prior to scanning the samples, white reference panel was used to optimize and calibrate the device [14]. However, samples were recorded with ten epochs

Fig. 1 Area of interest and region of field campaign
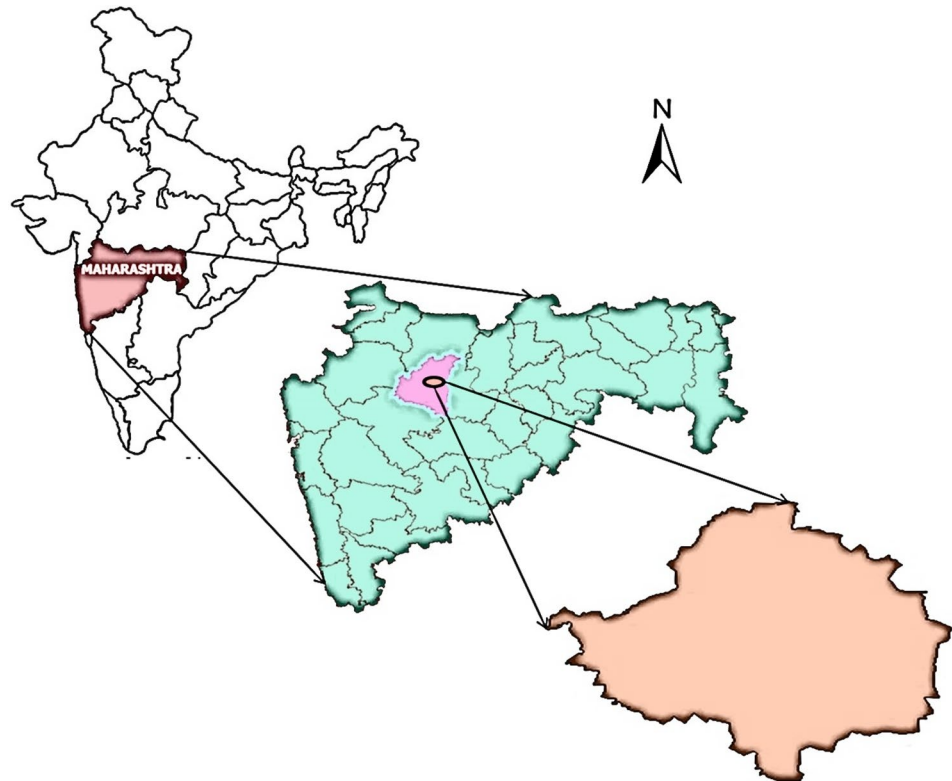
and subsequently averaged to obtain pure reflectance using RS3 software.

3 The methodology

3.1 Spectral preprocessing

Generally, the device produces background or low signal-to-noise; hence, spectral pre-treatment is essential to remove the said noises. Moreover, said causes were eliminated while removing the fringe spectra related to 350–399 nm and 2451–2500 nm wavebands. We obtained 2051 pure wavelength for the each spectra ranged between 400–2450 nm [14]. The ASCII files were generated from files with .asd format corresponding to each reflectance spectra for further analysis. The spectral smoothing and its renovation were done using SG method along with FDT method [15] in Unscrambler X 10.5 software CAMO, Norway. The order of polynomial was set as 2 for SG method. The first and second smoothing point was 2 and 5. The order of derivative was set as one. The absorption features of SNs were increased by FDT method. Both the methods were jointly used to predict the SNs.

3.2 Statistical analysis

Two separate groups were generated from 74 random soil samples for calibration (15) and validation (59) sets. The leave-one-out cross-validation (LOOCV) approach was used on all samples which were used in calibration set. Subsequently, correlation study was carried out between nitrogen and their predictor variables (reflectance spectra) using statistical analysis. The PLSR-based models were generated for the rest of the processing.

3.3 Development of models based on PLSR method

The PLSR-based models were developed after the spectral renovations with LOOCV method. Subsequently, the spectral data and reference (laboratory) data of soil were calibrated using the developed models. The PLSR is linear multiple regression method which requires fewer latent variables and used when independent variables are more. The models were designed using Eqs. 1 and 2 for decomposing the block of predictors (Eq. 1)

$$X = T \cdot P^T + \text{Residuals}(E) \tag{1}$$

and predict the response variables (Eq. 2)

$$Y = T \cdot C + \text{Error}(f) \qquad (2)$$

where X and Y are the mean-centered matrices which contain the predictor and response variables, respectively. P and C are the factor loadings; E and f are the matrices of residuals and errors, respectively [3, 4, 12, 16].
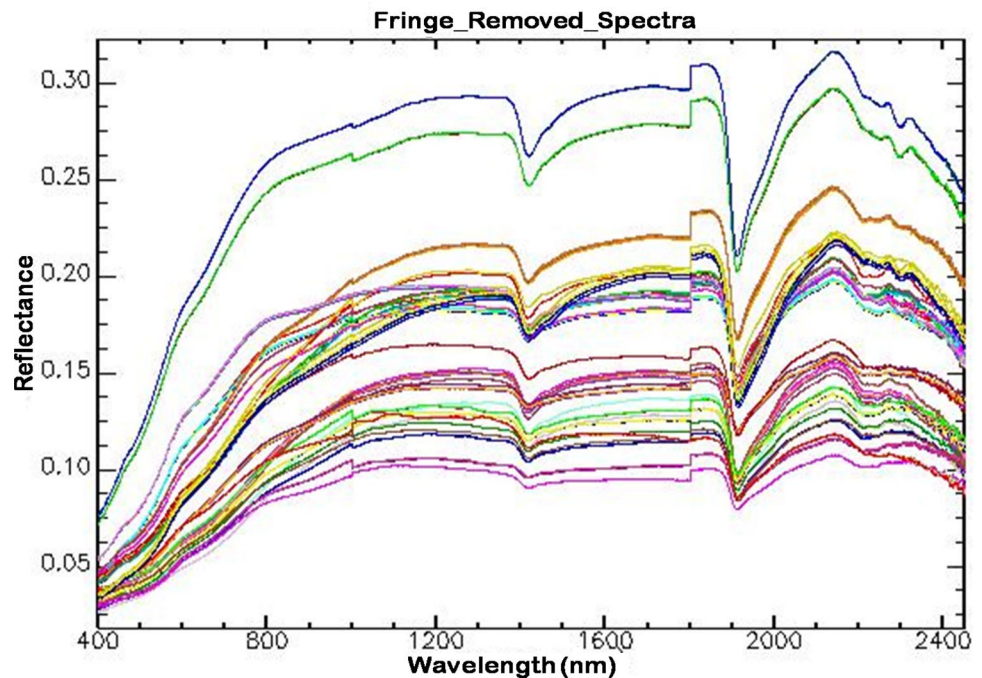
The LOOCV method was used to identify and analyze the nitrogen sensitive bands of spectra and to resolve the number of features for conserving the calibration models. The PLSR method generated number of features and models while developing the suitable model. Each model provided one additional feature from the earlier one. Hence, we obtained more models; among those, we have tested eight features with the computation of RMSEP. The suitable cross-validated calibration model was selected using RMSEP. The final model was selected based on highest accuracy of predicted R^2 and lowest RMSEP of cross-validation model. Finally, the optimum number of PLS features were determined. The RMSEP, R^2 values and LOOCV outcomes were used for evaluating the results obtained through PLS with all the nitrogen sensitive bands.

Table 1 Statistics of soil nitrogen

Soil nitrogen	Number of soil samples	Min	Max	Mean	SD
N (mg Kg $^{-1}$)	74	212.23	250.88	226.66	14.98

Max maximum, *Min* minimum, *SD* standard deviation

3.4 Validation of prediction models

The performance evaluation of developed prediction models were verified using R^2 values, ratio of performance to deviation (RPD) and RMSEP by Eqs. (3–5), respectively. Hence, the highest R^2 values and lowest RMSEP were used to select the suitable model. The highest accuracy of the model shows stability and correctness of the model [12].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y\text{meas} - Y\text{pred})^2}{\sum_{i=1}^{n}(Yi - Y_{\text{mean}})^2} \qquad (3)$$

$$\text{RMSEP} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_{\text{pred}} - Y_{\text{meas}})^2} \qquad (4)$$

$$\text{RPD} = \text{SD}_{\text{val.}}/\text{RMSEP} \qquad (5)$$

where Y_{meas} is the measured values (SN), Y_{pred} is the predicted theoretical values, Y_{mean} is the mean of measured values, n is the number of measured/predicted values with $i = 1, 2, 3, \ldots, n$, $\text{SD}_{\text{val.}}$ is standard deviation of measured values in the validation set, and RMSEP is the root mean square error of prediction of validation dataset [12, 17].

4 Experimental results and its discussion

The statistical outcomes of the determination of nitrogen were derived through laboratory methods. The laboratory analysis was done on 74 soil samples resulted in Table 1.

Fig. 2 The edge-band-removed spectra (400–2450 nm)

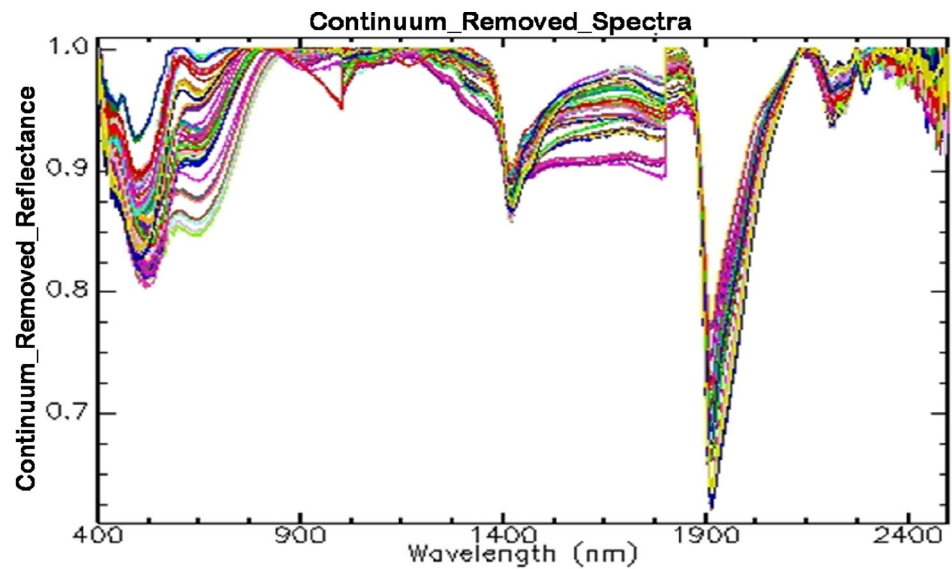**Fig. 3** The reflectance spectra of soil samples using continuum-removed method

Fig. 2 shows redundant noise eliminated spectra of 74 samples as discussed in section III-A. The edge-removed spectra were used for further processing to avoid the redundant noises. However, there was also some confusion in fringe-removed spectra due to more spectral bands. This task is complicated for computation. Hence, we have used continuum-removal (CN) method to detect the sensitive bands related to nitrogen (Fig. 3).

The CN method is a means of stabilizing reflectance spectra for allowing the assessment of particular absorption characteristics from regular baseline [18]. The detection of absorption characteristics of soils is done by the CN method. Also the calculation of band depth absorption was carried out using this method. As a result, the

absorption features were directly identified from every spectrum along with its CN spectra. The water absorption channels were identified at 1417, 1914 and 2209 nm bands. Moreover, the clay of soil was also detected at 2200–2329 nm [9, 19]. Conversely, nitrogen sensitive bands were observed at the wavelengths 451–491, 511–521, 536, 612–658, 980–997, 1209–1376, 1472–1795, 1901, 2210, 2296 and 2382 nm [9]. Hence, further analysis was done on nitrogen sensitive bands.

Consequently, preprocessed (edge-removed and FDT-SG smoothing) spectra were used for developing the PLSR models. The overlapped and nonchemical affected bands were eliminated using FDT-SG method to smooth the spectra. The de-noised and renovated spectra of 74

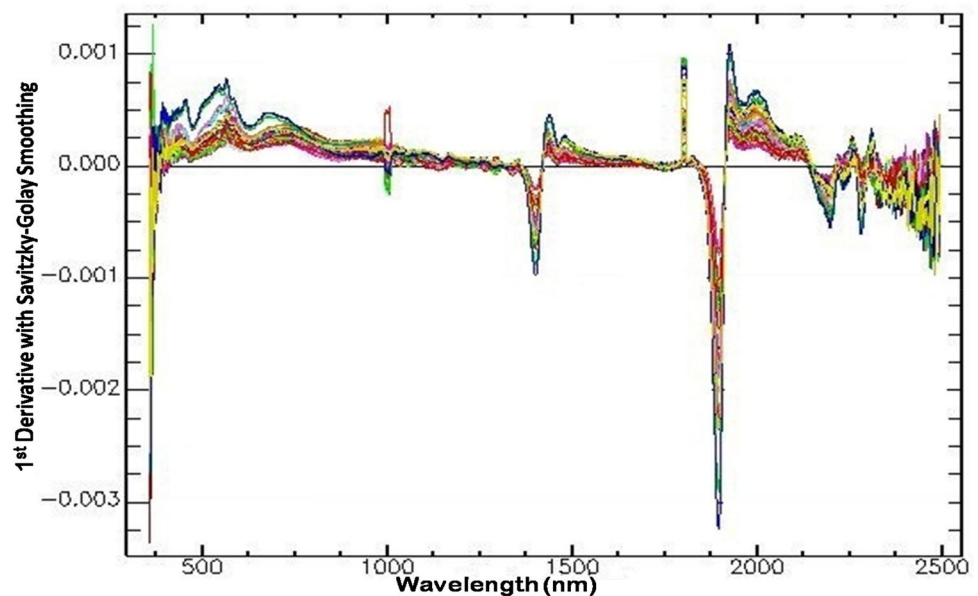**Fig. 4** Renovated spectrum by FDT-SG smoothing method

Table 2 Accuracy evaluations of nitrogen features based on PLSR model

Soil nitrogen (mg Kg $^{-1}$)	Calibration set		Validation set			
	R^2	RMSE	R^2	RMSEP	SD	RPD
Before FDT-SG method	0.72	3.95	0.68	4.34	9.25	2.13
After FDT-SG method	0.95	1.48	0.94	1.56	3.34	2.14

Fig. 5 The predicted and measured SN values scatter plot based on PLSR modeling before preprocessing
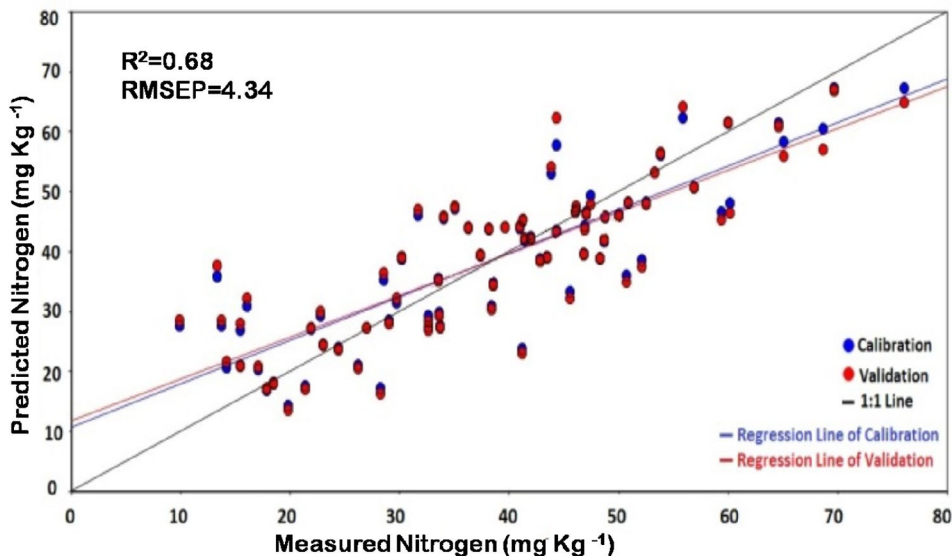
soil samples are shown in Fig. 4. The spectral characteristics of these spectra have been preserved. It was noticed that the spectral smoothness was enhanced with the use of FDT-SG method. Moreover, the nitrogen sensitive bands were also preserved significantly.

Subsequently, the correlation study was done between spectra and nitrogen of the samples. Since the correlation analysis is the reliable approach for studying the correlation between dependent and independent variables. It was observed by the obtained results, the SN features were enhanced by the renovated spectra. The suitable bands were identified using the correlation study of nitrogen features and transferred spectra. Hence, the correlation study provided suitable bands at the wavelengths 480 nm, 511 nm, 653 nm, 997 nm, 1472 nm, 1795 nm, 2210 nm and 2296 nm. The highest correlation for the SN ($R^2 = 0.95$) was found at wavelength 997 nm.

The random cross-validation method with PLS kernel were used for developing the models. Afterward, the prediction performance (Table 2) of models was evaluated for both (raw and renovated) the spectra. However, the raw spectra were not sufficient for SN calibration. The preprocessed spectra have given satisfactory SN calibrations. The calibration accuracy was only 72% ($R^2 = 0.72$) on the raw spectra with 3.95 RMSE values. The obtained calibration (R^2 values and RMSE) and validation (R^2 values, RMSEP, SD and
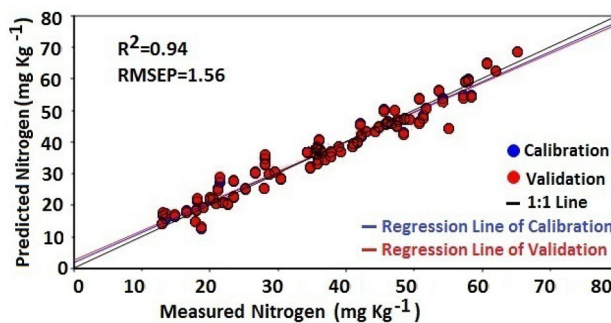

Fig. 6 The predicted and measured SN values scatter plot based on PLSR modeling after preprocessing

RPD) statistics of raw and renovated spectra have been summarized in Table 2.

The scatter plots of the measured values and RSHRS predicted values of SNs on both the spectra are shown in Figs. 5 and 6 correspondingly. The R^2 values were higher for all the PLSR models in calibration sets. However, the prediction accuracy ($R^2 = 0.68$ and RMSEP = 4.34) was found to be low for raw spectra. Since the nitrogen was moderately calibrated and validated (Table 2) after the spectral renovation. Furthermore, the verification performances of developed models for the prediction of SN were evaluated by cross-validation R^2, RMSEP and RPD values (Table 2). The percentage of variation in the

dependent variable was demonstrated using R^2 values which was accounted for the independent variables. The RMSEP value estimates the variation of predicted values around measured values, and this approaches to zero [17].

The ability of RSHRS was shown by the RPD values to predict the values of soil properties [20]. The RPD is grouped into three ranges > 2.0, 1.4–2.0, and < 1.4 which indicate the excellent, acceptable and poor predictions, respectively. It was recommended that the predictions between the ranges 1.4–2.0 could be achieved by various calibration approaches [20]. Conversely, the properties of the ranges having < 1.4 may not be reliably predicted using reflectance spectra [20]. The results indicate that the RPD values were better for this calibration model for both types of the spectra. The RPD values were 2.13 and 2.14 for raw and renovated spectra, respectively. Alternatively, the calibration and prediction accuracy was enhanced for renovated spectra. The prediction accuracy was 94% ($R^2 = 0.94$) for renovated spectra along with RMSEP = 1.56 and RPD = 2.14 which is satisfactory for agricultural areas. Hence, it may be highlighted according to the verification results, the pretreatment is essential for the estimation of soil nutrients. Our results were found to be satisfactory to predict the SNs for large area with acceptable accuracy.

Thus, the model is accurate and reliable to estimate the SNs from farming sectors. Our results were obtained with accuracy of ($R^2 = 0.68$ and 0.94) for both classes of the datasets. The result ($R^2 = 0.95$) obtained in previous study [1] was similar to our results. However, the methods used in study [1] were different and may be the reason for related outcome. The prediction accuracy ($R^2 = 0.94$) of the present study for SN is better than the reference [3] which was studied for densely covered area with coastal wetlands. The study [11] demonstrated that the use of PLSR along with back propagation neural network (BPNN) and genetic algorithm (GA)-BPNN was sufficient for the prediction of soil nitrogen. However, the prediction accuracies (R^2) were 0.25, 0.65 and 0.82 for PLSR, BPNN and GA-BPNN methods, respectively [11], which are inferior to our results. Similarly, total nitrogen was accurately predicted [21] using local correlation maximization-complementary superiority method (LCMCS), PLSR, local correlation maximization (LCM) de-noising method and complementary superiority (CS) method with 0.88, 0.74, 079 and 0.81 prediction (R^2) results, respectively. It is clearly seen that R^2 values of our study are slightly superior to the values reported in [21]. The performance of our model is superior to the results ($R^2 = 0.92$ and 0.93) obtained in references [22] using PLS and extreme learning machine models. Thus, the present methodology performed better than results reported earlier [2, 4, 23–25].

5 Conclusions and future scope

The RSHRS approach was successfully implemented on 74 soil sample for the estimation of nitrogen contents in agricultural areas. The contribution of the present study is the new model for estimating the nitrogen for agricultural sites. The absorption bands due to nitrogen were successfully identified using the (CN) method. Hence, it is observed that the (CN) method is needed to detect the absorption bands for soil nutrients especially nitrogen. The correlation study is also needed for identifying nitrogen sensitive bands. Additionally, the preprocessing is significant for better correlation of nitrogen features and models. The present approach highlights the significance of reflectance spectra as compared to the conventional methods. It is found that the nitrogen sensitive spectral bands are at the wavelengths 480 nm, 511 nm, 653 nm, 997 nm, 1472 nm, 1795 nm, 2210 nm and 2296 nm. The prediction and calibration accuracy are enhanced with the use of FDT-SG method. The present PLSR model was robust and more appropriate for the estimation of nitrogen features. It is observed that the SN can be estimated very well for large areas with better results. The present study is expected to be useful for healthy farming practices. In future, the model can be implanted using satellite or airborne hyperspectral images in large area.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Wei Y, Zhu X, Li C, Cheng L, Wang L, Zhao G, Jiang Y (2016) Monitoring soil nitrate nitrogen based on hyperspectral data in the apple orchards. Agric Sci 8(01):21
2. Wu J, Liu Y, Chen D, Wang J, Chai X (2009) Quantitative mapping of soil nitrogen content using field spectrometer and hyperspectral remote sensing. In: Environmental science and information application technology, 2009. ESIAT 2009. International conference on (vol 2, pp 379–382). IEEE
3. 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
4. Rossel RV, Walvoort DJ, 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

5. 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 (IJERA) 3(3):81–91

6. Rossel RV, Behrens T, Ben-Dor E, Brown DJ, Demattê JA, Shepherd KD, Shi Z, Stenberg B, Stevens A, Adamchuk V, Aïchi H (2016) A global spectral library to characterize the world's soil. Earth Sci Rev 155:198–230

7. Brown DJ, Shepherd KD, Walsh MG, Mays MD, Reinsch TG (2006) Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132(3):273–290

8. Ben-Dor E, Irons JR, Epema GF (1999) Soil reflectance. Man Remote Sens Remote Sens Earth Sci 3:111

9. Ben-Dor E, Banin A (1994) Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. Remote Sens Environ 48(3):261–274

10. 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

11. Peng Y, Zhao L, Hu Y, Wang G, Wang L, Liu Z (2019) Prediction of soil nutrient contents using visible and near-infrared reflectance spectroscopy. ISPRS Int J Geo-Inf 8(10):437

12. Vibhute AD, Kale KV, Mehrotra SC, Dhumal RK, Nagne AD (2018) Determination of soil physicochemical attributes in farming sites through visible, near-infrared diffuse reflectance spectroscopy and PLSR modeling. Ecol Processes 7(1):26

13. https://www.asdi.com/. Accessed 03 April 2015

14. Hatchell DC (1999) Analytical spectral devices, Field Spec™ User's Guide. Inc. (ASD) Technical Guide, 3rd Ed

15. Savitzky A, Golay MJ (1964) Smoothing and differentiation of data by simplified least squares procedures. Anal Chem 36(8):1627–1639

16. 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

17. Srivastava R, Sethi M, Yadav RK, Bundela DS, Singh M, Chattaraj S, Singh SK, Nasre RA, Bishnoi SR, Dhale S, Mohekar DS, Barthwal AK (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 Sens 45(2):307–315

18. Clark RN, Roush TL (1984) Reflectance spectroscopy: quantitative analysis techniques for remote sensing applications. J Geophys Res: Solid Earth 89(B7):6329–6340

19. 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

20. 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

21. Lin L, Wang Y, Teng J, Xi X (2015) Hyperspectral analysis of soil total nitrogen in subsided land using the local correlation maximization-complementary superiority (LCMCS) method. Sensors 15(8):17990–18011

22. Li H, Jia S, Le Z (2019) Quantitative analysis of soil total nitrogen using hyperspectral imaging technology with extreme learning machine. Sensors 19(20):4355

23. Veum S, K, A Parker P, A Sudduth K, H Holan S, (2018) Predicting profile soil properties with reflectance spectra via bayesian covariate-assisted external parameter orthogonalization. Sensors 18(11):3869

24. Kusumo BH, Hedley CB, Hedley MJ, Hueni A, Tuohy MP, Arnold GC (2008) The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen analysis of pastoral soils. Soil Res 46(7):623–635

25. Kusumo BH, Hedley MJ, Hedley CB, Tuohy MP (2011) Measuring carbon dynamics in field soils using soil spectral reflectance: prediction of maize root density, soil organic carbon and nitrogen content. Plant Soil 338(1–2):233–245