Das Bhabani S. (Orcid ID: 0000-0002-3542-7421)
Reddy Nagarjuna (Orcid ID: 0000-0002-9981-7058)

Diffuse Reflectance Spectroscopy (DRS) for Rapid Soil Testing and Soil Quality Assessment in Smallholder Farms

Short running head: Reflectance spectroscopy for smallholder farms

Israr Majeed[1], Kaushal K. Garg[2], A. Venkataradha[2], Naveen K. Purushothaman[1], Sourav Roy[1], Nagarjuna N. Reddy[1], Ramesh Singh[2], K. H Anantha[2], Sreenath Dixit[2], Bhabani S. Das[1]*

[1]Agricultural and Food Engineering Department, Indian Institute of Technology, Kharagpur 721302, West Bengal, India

[2]International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502324, Telangana, India

*Corresponding author. Email: bsdas@agfe.iitkgp.ac.in

Other co-authors' e-mails:

resheisrar@gmail.com (Israr Majeed)

kaushal.garg@icrisat.org (Kaushal K. Garg)

venkataradha.akuraju@icrisat.org (A. Venkataradha)

naveenpurushothaman1098@gmail.com (Naveen K. Purushothaman)

souravroy.ag2012@gmail.com (Sourav Roy)

nreddy187@gmail.com (Nagarjuna N. Reddy)

ramesh.singh@icrisat.org (Ramesh Singh)

anantha.kanugod@icrisat.org (K. H. Anantha)

sreenath.dixit@icrisat.org (Sreenath Dixit)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ejss.13358

ABSTRACT

Rapid soil testing and soil quality assessment are essential to address soil degradation and low farm incomes in smallholder farms. With the objective of testing diffuse reflectance spectroscopy (DRS) to rapidly assess the soil's chemical properties, nutrient content and soil quality index (SQI), samples of surface soil were collected from 1113 smallholder farms in seven districts in Bundelkhand region of Uttar Pradesh, India. A minimum dataset (MDS) approach was followed to estimate SQI using the three chemical parameters of soil pH, electrical conductivity (EC) and soil organic carbon (SOC) and 11 different soil nutrients. Principal component and correlation analyses showed that soil pH, SOC content and three available nutrients -- Cu, Fe and S -- may constitute the MDS. Estimated SQI values showed strong positive correlation with crop yields. Results of chemometric modelling showed that the DRS approach could yield the coefficient of determination (R^2) values in the validation datasets ranging from 0.79 to 0.94 for exchangeable Ca followed by 0.67 to 0.88 for exchangeable K, 0.52 to 0.86 for SOC and 0.53 to 0.81 for available B content. Except in one district, the DRS approach could be used to estimate SQI values with R^2 values in the range of 0.63 to 0.81; R^2 value of 0.71 was obtained in the pooled dataset. We also estimated the three-tier soil test crop response (STCR) ratings to compare DRS and wet chemistry soil testing approaches. Similar STCR ratings were obtained for both these approaches in more than 86% of the samples. Parameters for which both the methods yielded similar ratings in more than 80% of the samples were EC (> 98%), pH and exchangeable Ca (> 81%) and available B (> 89%). With similar ratings, these results suggest that the DRS approach may safely be used for farmers' fields, replacing the traditional wet analysis approach of soil testing.

Keywords: Chemometric models, crop yield, soil nutrients, soil testing, soil quality index, soil test crop response rating

Abbreviations: DRS: Diffuse reflectance spectroscopy; PLSR: Partial least squares regression; SVR: Support vector regression; RF: Random forest; $PLSR_{LW}$: Locally-weighted PLSR; $PLSR_{FS}$: PLSR with feature selection; SQI: soil quality index

Highlights

- Diffuse reflectance spectroscopy is a rapid soil testing approach ideal for smallholder farms.
- Moderate to excellent prediction accuracies were obtained with the DRS approach.
- DRS and wet chemistry-based approaches shows similar soil test crop response ratings.
- Soil quality is a key driver for crop yield, cropping intensity, land use and farm income.

1. Introduction

Land degradation and low farm income are two big challenges in many developing countries. Globally, about 24% of land is degraded and it influences more than 3.2 billion people (Edrisi et al., 2022). Specifically, many in rural communities, smallholder farmers and very poor people in the world are affected by land degradation (GEF, 2019). About 475 million out of the 570 million agricultural farms in the world are smallholder farms (Nature, 2020). Farmers in many of these farms have inadequate access to agricultural inputs (Ricciardi et al., 2021) and agricultural technologies to tackle production challenges (Nature, 2020; Laborde et al., 2020). Balanced nutrient application is an important precision management strategy to increase agricultural production. Similarly, a proper assessment of soil quality in degraded lands is expected to improve resource use efficiencies in different crop production systems. However, the sheer number of smallholder farms with fragmented landholdings, poor infrastructure and the lack of sufficient knowledge and capacity (Kishore et al., 2021) make precision nutrient management a challenging task. Well-developed soil testing laboratories and instrumentation for controlled nutrient application in developed countries allow the successful implementation of soil test-based precision nutrient management protocols. However, efforts to replicate such successes on millions of smallholder farms in developing countries require both rapid and affordable sensing and mapping technologies.

Over the last few decades, diffuse reflectance spectroscopy (DRS) in the visible and near-infrared (VNIR) region has emerged as a promising technique to assess multiple soil properties (Ben-Dor and Banin, 1995). It has been used for estimating soil texture (Benedet et al., 2020; Davari et al., 2021), soil organic carbon (SOC) content (Viscarra Rossel et al., 2006; Gholizadeh et al., 2021; Bai et al., 2022), major nutrients (Johnson et al., 2019), micronutrients (Chang et al., 2001), soil hydraulic properties (Santra et al., 2009; Babaeian et al., 2015), soil engineering properties (Gupta et al., 2016), soil aggregate stability (Gomez et al., 2013), lime requirement (Horf et al., 2022) and weathering indices (Mohanty et al., 2016), among others. Comprehensive spectral libraries are being generated in different countries to develop robust chemometric models (Bellinaso et al., 2010; Stevens et al., 2013; Viscarra Rossel et al., 2016; Shepherd et al., 2022). Large variability in spectral database (Coblinski et al., 2020), robust chemometric models (Ludwig et al., 2019; Wang and Wang, 2022) and different pre-processing approaches (Dotto et al., 2018) have improved the performance of the DRS approach. Improvements in the performance of chemometric models have also been reported by combining both VNIR and the mid-infrared (MIR) spectra (Viscarra Rossel et al., 2006) and also by combining VNIR spectra of different aggregate size fractions (Vasava et al., 2019; Vasava and Das, 2022). A working group has been constituted under the IEEE Standards Association to standardize a DRS approach as a soil testing method (https://sagroups.ieee.org/4005/).

Despite several successes, the use of DRS as a replacement for wet chemistry-based (conventional) soil testing is debated (McBride, 2022; Viscarra Rossel et al., 2022). While the DRS approach works in some scenarios and fails in others (Pasquini, 2018), Viscarra Rossel et al. (2022) observe that its confident use lies in its adequate validation with appropriately sampled datasets. DRS studies often use a small sample size (Barra et al., 2021) drawn from small geographical areas (~0.01-5.0 km^2). Specifically, limited studies have been conducted to calibrate chemometric models using spectral data from smallholder farms. Recently, Johnson et al. (2019) accurately estimated 13 out of 29 soil fertility parameters from 42 different sites across 20 sub-Saharan African countries. Similarly, Singh et al. (2019) have shown that both DRS and conventional soil test results yielded similar nutrient recommendations for cocoa plantations. Based on a detailed cost analysis, Li et al. (2022) recently showed that the dry combustion approach of estimating soil carbon may be adequate if one wishes to test about 50 samples per day; however, the DRS approach is cheaper in comparison when more than 250 samples are to be tested per day. Recently, Viscarra Rossel et al. (2022) stated that the spectroscopic approach is almost 10 times cheaper than the conventional approach.

Though smallholder farms provide huge opportunities to increase agricultural production, they need adequate nutrient application (Nesme, 2022). While soil testing is a significant step to supplement required nutrients through fertilizer and manure application, many smallholder farms use fertilizers without soil testing. A major reason for such a practice is the large number of smallholder farms and the large number of soil samples required to be tested, which invariably leads to delayed delivery of soil test reports to farm owners. Since the DRS approach enables the cost-effective testing of several samples (Li et al., 2022), the timely delivery of soil test reports and recommendations is feasible if the approach is adequately calibrated in farmers' fields. This study examines the DRS approach's potential as a complement to conventional soil testing and soil quality assessment in smallholder farms. Specifically, many smallholder farm owners require soil test crop response (STCR) ratings to make fertilizer recommendations. Thus, a major objective of this study is to examine if the DRS approach may be used for developing STCR recommendations similar to the conventional approaches.

2. Materials and Method

2.1. Study area and soil sampling

This study was conducted in seven districts of Bundelkhand region in Central India (Fig. 1) by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad. Bundelkhand is located in the Ken and Betwa catchment of the Yamuna sub-basin of the Ganga river basin. The region covers an area of 6 Mha in parts of Madhya Pradesh and Uttar Pradesh states and a population of 15 million. Seven clusters (each of 50 km^2) in 20 villages and representing different soil types, topography, land use and cropping system of the region were identified. Soils in Bundelkhand region are generally classified under Entisols, Alfisols, Inceptisols and Vertisols (Kumar et al., 2021). The region is known for its water scarcity, land degradation and poor agricultural and livestock population. Several districts in the region have low SOC content. Large tracts of cultivable land have already been transformed into ravenous land (Singh et al., 2014; Garg et al., 2020; Dev et al., 2022). Undulating topography, extreme temperature regime and changes in land use are the main reasons for such degradation. The region receives about 570-820 mm of annual rainfall (Table S1, Supplementary Materials) with large spatial and temporal variability (Singh et al., 2022). Maximum and minimum temperatures in the region often reach 45-47^{o}C during May-June and 2-4^{o}C during December-January, respectively. Agriculture occupies the bulk of land use (53-90%) while wasteland often constitutes more than 20% of the total geographical area, as in the districts of Lalitpur, Jhansi and Chitrakoot (Table S1). More than 80% of the people in this region live below the poverty line (poverty threshold: daily per capita income of US$ 1.25). Due to water scarcity, about 30-60% of cultivable land is left fallow either during the monsoon (*kharif*) or post-monsoon (*rabi*) seasons. About 52% of the total cultivable area in Bundelkhand region is rainfed (Kumar et al., 2021). Groundnut, black gram, green gram, sesame and pigeonpea are the predominant crops cultivated during the *kharif* season while mustard, chickpea, field pea, barley and wheat are generally grown during the *rabi* season (Table S1).

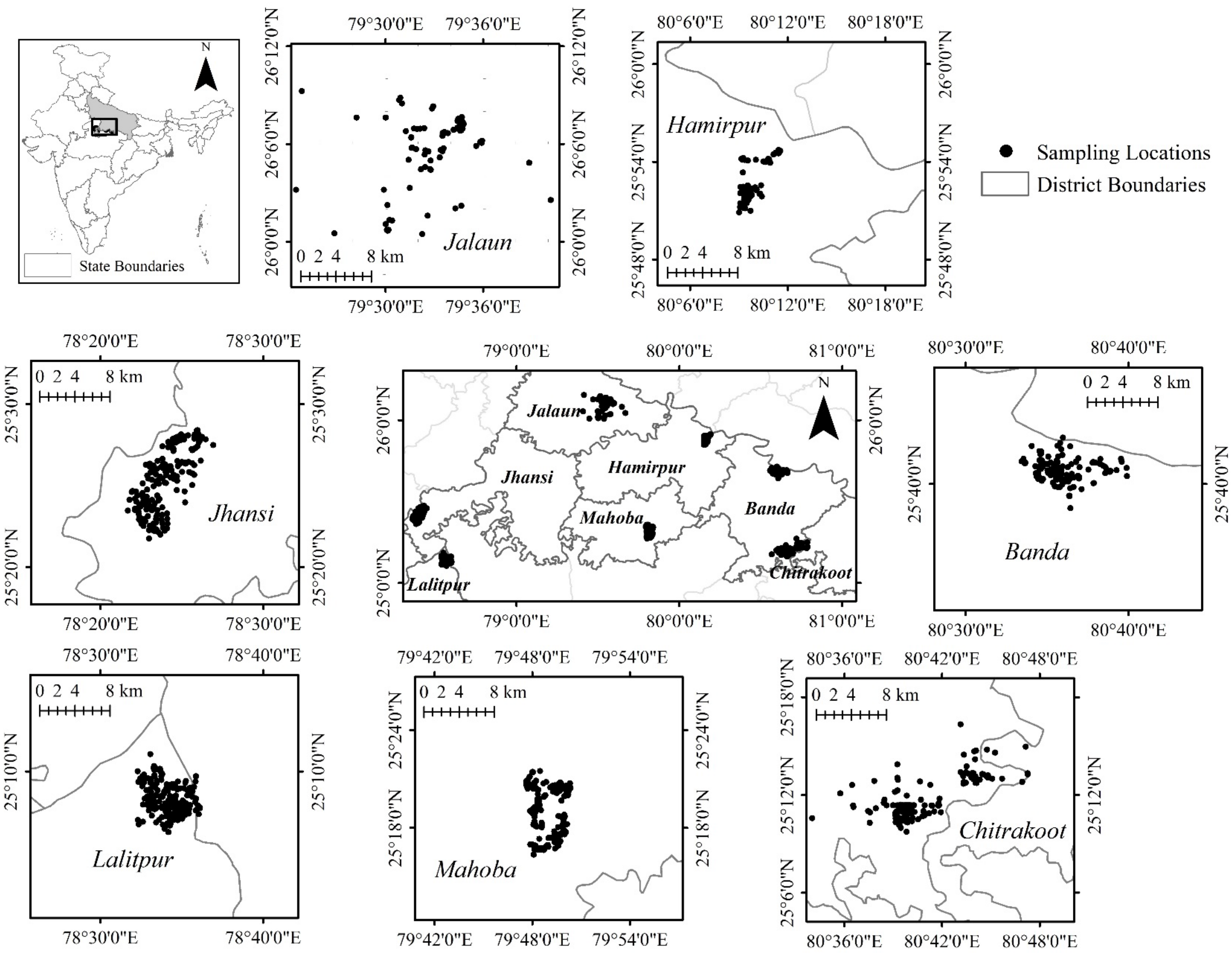

Fig. 1: Locations of soil samples collected from smallholder farms in selected districts of Bundelkhand region

2.2. Determining soil properties, nutrient content and soil quality index (SQI)

A total of 1113 surface soil samples were collected (depth of 0-15 cm) during May-June, 2018 following stratified sampling. The samples were air-dried, ground and sifted through a 2 mm sieve and used for physical, chemical and spectral analyses. Each sample was analyzed to estimate soil pH, electrical conductivity (EC), SOC content, macronutrients, secondary nutrients and micronutrients. Soil pH and EC were measured in 1:2 soil:water slurry (Thomas, 1996; Rhoades, 1996). Chromic acid digestion method (Walkley and Black, 1934) was followed to estimate SOC content. Available P was estimated using the Olsen's and Bray's method using a continuous auto-analyzer (Olsen and Sommers, 1982). The ammonium acetate method was used to extract available cations -- calcium (Ca), magnesium (Mg), sodium (Na) and potassium (K) -- for assessing soil cation exchange capacity (CEC) (Hanway and Heidel, 1952). A flame photometer (Systronics India Ltd., Ahmedabad, India) was used to measure exchangeable cation concentrations. Available boron (B) was estimated using the hot water-soluble extraction method (Keren, 1996).

The SQI value for each sample was estimated following three main steps: selection of appropriate soil properties to construct a minimum dataset (MDS), development of parameter scores and the aggregation of scored parameters into weighted SQI values (Andrews et al., 2002). To construct the MDS, all the soil parameters (pH, EC, SOC content and 11 nutrient contents) were standardized and the principal component analysis (PCA) was applied to select principal components (PC) with eigenvalue ≥ 1 (Kaiser, 1960; Girden and Kabacoff, 2010). Variables within 10% of the highest component loading (absolute values) within each PC were chosen as indicator variables to construct the MDS (Armenise et al., 2013). When multiple indicators appeared within a particular PC, Pearson correlation coefficients (r) among these indicators were used to remove redundant variables. Specifically, only the variable with the highest factor loading was retained in the MDS when the correlation coefficient between two indicators exceeded 0.7 (Andrews and Carroll, 2001). Selected indicators in the MDS were then transformed to unitless scores using a linear scoring (less is better, more is better and optimum is better) function (Vasu et al., 2016). Table S2 shows the scoring functions, parameters and relationships for scoring functions for each soil indicator in our study area. With the selected indicators in the MDS transformed into dimensionless scores, a weighted additive method was employed to estimate SQI values:

$$SQI = \sum_{i=1}^{n} W_i S_i \qquad (1)$$

Where, W_i and S_i are the i^{th} weighting factor (derived from PCA outcomes) and score of the individual parameter, respectively. All the statistical analyses was carried out in MATLAB R2021b (The MathWorks, Inc., Natick, MA, USA).

2.3. Collection of soil spectra and chemometric modelling

Spectral reflectance for each processed soil sample was measured in the laboratory in proximal mode over the VNIR region (350–2,500 nm) using a portable spectroradiometer (Model: Field spec®4 Hi-Res NG; Malvern Panalytical, UK) and a turntable fitted with a halogen bulb as its light source. About 100 g of processed soil sample was used to fill a glass petri dish and the soil surface in each petri dish was levelled with a thin glass plate. The Spectralon® white reference panel (Lab sphere, USA) was used to obtain reference spectra. Soil spectra were collected following the protocol of: (a) warming up the instrument for an hour before data collection, (b) optimization and collection of reference spectra after every 30 samples and (c) averaging 30 scans per sample. Smoothing of individual spectra was done by a third-order Savitzky-Golay smoothing method with a span length of 9 nm (Savitzky and Golay, 1964). Preliminary modelling studies using five chemometric models (PLSR: partial least squares regression, SVR: support vector regression, RF: random forest, $PLSR_{LW}$: locally-weighted PLSR, $PLSR_{FS}$: PLSR with feature selection approach) showed that the $PLSR_{FS}$ and SVR models were the two top performers in our dataset. Based on a meta-analysis of soil spectroscopic studies conducted between 1990 and 2019, Ahmadi et al. (2021) have observed that about 69.6% of research papers used PLSR (62.3%) and modified PLSR models followed by SVR (9.9%) models. Modified PLSR models such as $PLSR_{FS}$ (Sarathjith et al., 2016) and $PLSR_{LW}$ (Gupta et al., 2018) have also been known to perform better than PLSR models (Ng et al., 2018; Dorantes et al., 2022). Therefore, we used the SVR and $PLSR_{FS}$ models in the remainder of this study to estimate 14 nutrient contents and corresponding SQI values. In the SVR approach, training samples are mapped so as to maximize the width between the observed and predicted responses (Smola and Schölkopf, 2004). For the $PLSR_{FS}$ approach, soil properties are estimated in the PLSR approach after selecting important feature variables (Teófilo et al., 2009; Sarathjith et al., 2016). For this study, we considered 35 predictor variables consisting of PLSR regression-coefficients (β), variable influence on projection (VIP), AMI (Sarathjith et al., 2014) and their combinations (Sarathjith et al., 2016) as listed in Table S3. The ordered predictor selection (OPS) approach (Teófilo et al., 2009) was used to identify a parsimonious set of predictors in this algorithm. Details of the $PLSR_{FS}$ model along with the OPS algorithm are shown in the Supplementary Materials.

To comprehensively test these modelling approaches, we examined five transformation schemes for individual soil spectra and two transformation schemes for soil properties. Thus, a total of six cases were examined for spectral transformation: (a) untransformed spectra, (b) log-transformed absorbance [log(A)], (c) first derivative of spectra (FD), (d) second derivative of spectra (SD), (e) standard normal variate (SNV) and (f) Kubelka-Munk transformation (KM). Similarly, three cases were examined for the transformation of soil properties: (a) no transformation when a property is normally distributed and/or if the property does not conform to normal distribution even after log or Box-Cox transformation, (b) log transformation and (c) Box-Cox transformation. A two-tailed Kolomogorov-Smirnov test with 5% level of significance was followed to test for the normality of each soil property. A partition sorting approach (Viscarra Rossel et al., 2006) was used to divide our whole dataset into 75% calibration and 25% validation samples. Thus, a total of 36 sets of calibration and validation statistics (2 models × 3 transformations for soil parameter × 6 transformations for spectra) were estimated for each parameter for a given cluster (district). All the spectral transformations and modelling work were carried out in MATLAB R2021b (The MathWorks, Inc., Natick, MA, USA). The prediction performance of the DRS algorithms was assessed using root-mean-squared error (RMSE) and coefficient of determination (R^2):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \quad (3)$$

Where, Y_i is the measured soil parameter with its mean value of $\bar{Y}_i$ and predicted value of $\hat{Y}_i$ at the i^{th} location and N is the number of locations. The ratio of performance to deviation (RPD) values were obtained by dividing the standard deviation (SD) of the reference data by the standard error of prediction (SEP) values (Williams, 1987). We also estimated the bias values as the mean error (average difference between $\hat{Y}_i$ and Y_i). All the transformed variables were back-transformed before evaluating the performance statistics. Consideration of $PLSR_{FS}$ and SVR models and 18 different transformations (3 transformations for soil properties and 6 transformations for soil spectra) resulted in 36 different sets of performance statistics for each of the 15 soil parameters in 7 districts and in the pooled datasets (a total 120 test cases: 15 properties at each of the 7 locations and pooled data). For each of these test cases, the best chemometric model and the best transformation of spectra and individual analytes were selected based on the R^2, RMSE, bias and RPD values. We also classified both the observed and estimated soil parameters using a three-tier STCR rating scheme generally used for Indian soils (Sendhil et al., 2018). The classification accuracy for a given parameter in the specific district and in the pooled dataset was estimated by calculating the percentage of locations for which estimated and observed values of a parameter yielded identical STCR ratings.

2.4. Assessment of agricultural productivity

We considered crop yield, land use and farm income as major indicators of agricultural productivity. Due to water scarcity, farmers in Bundelkhand region grow crops that require less water. Rainfed and short-duration (<100 days) crops such as black gram, green gram, pearl millet, sorghum and sesame are grown in the *kharif* season. Farmers also grow crops such as groundnut (90-100 days) and pigeonpea (180-250 days) where supplemental irrigation is available. For the *rabi* season, farmers grow slightly longer-duration crops (120-150 days) such as mustard, chickpea, field pea, lentil, barley and wheat, which require several irrigations during the growing period. Before land development interventions were implemented in the study area, a baseline household survey was conducted during March-June, 2018 by randomly selecting 1,403 farmers covering all the 20 villages to collect data on household demographic characteristics, household occupation, livelihood activities, farm resources, assets and yields of different crops. Using these responses, we estimated average yield, cropping intensity and the contribution of farm income (only from crop cultivation) to the total income for each district in order to evaluate their connection with soil quality. The extent of agricultural land cultivated and left fallow during the *kharif* and *rabi* seasons were also estimated from the farmers' responses.

3. Results

3.1. Soil chemical properties, nutrient content and SQI

Table 1 shows the average values for three soil chemical properties and 11 nutrient contents. Although the number of soil samples within a district is relatively small, the average soil properties show clear characteristics of soil pH control on the availability of secondary and micronutrients for different districts. Specifically, samples collected from Hamirpur, Banda and Jalaun had alkaline reactions, with Hamirpur showing the maximum average soil pH of 8.29. Soil samples in the remaining districts were acidic to neutral with the Lalitpur samples showing the lowest average pH value of 6.66.

In terms of acidic soil reaction, Lalitpur's soil samples showed the highest average available soil phosphorous (P), iron (Fe), zinc (Zn) and manganese (Mn) and lowest exchangeable Ca and Mg concentrations among the 7 districts. Opposite trends may be seen for these nutrients in the soils of Hamirpur, Jalaun and Banda (pH > 7.0). Particularly, low available P in soil samples of these three districts may be due to the precipitation of P in the form of calcium phosphates making the soils chemically degraded. Despite the higher concentration of basic elements such as Ca and Mg, almost all the soil samples had low electrical conductivity. In general, soils with high soil pH showed high exchangeable K and available copper (Cu), with Banda soils registering the highest average for both these nutrients. Soil samples from Mahoba showed the highest average concentration for available sulphur (S). With lowest concentrations for S, Zn, Mn, P and SOC content, Jalaun soils may be considered the most degraded among the 7 districts of this study. In general, SOC content was low with an average value of just 0.53% in the pooled dataset as is often observed in most Indian soils (Rai et al., 2009; Reddy et al., 2021). From a variability perspective, available Zn in Chitrakoot's soil samples showed a maximum coefficient of variation (CV) value of 296% followed by 214% for available S in Mahoba's soil samples. Soil pH was the least variable soil parameter with a CV range of just 3% in Banda and Hamirpur soils to about a maximum of 10% in Lalitpur's soils. Such low variability in soil pH for sampled areas as large as 377 km^2 in Jalaun and 206 km^2 in Chitrakoot suggest that large tracts of geographical area in the study site may be degraded – a condition expected with widespread gullies and ravines in Bundelkhand region.

Results obtained from the PCA revealed that only four PCs have eigenvalues ≥ 1 (Table 2). Together, these four PCs explained more than 65% of total variability in the dataset. Under PC1, the component loading matrix (Table 2) showed that soil pH was the only highly weighted variable and was retained to constitute MDS. Based on the factor loading values greater than 90% of the maximum factor loading within a PC, both SOC and available Cu content were selected as indicator variables for PC2. Both these indicators were included in MDS because of the lack of a strong correlation between them (Table 3). Along with available Fe under PC3 and available S under PC4, SQI values were estimated using a 5-parameter MDS (pH, SOC, Cu, Fe and S) in Eq. 1. Estimated SQI values ranged from 0.37 to 0.75, with the average value ranging from 0.43 for Hamirpur and Jalaun to 0.49 for Lalitpur and Jhansi (Fig. 2A). Figure 2 also shows the scatter plots between SQIs in the pooled samples as functions of pH (Fig. 2B), SOC content (Fig. 2C), log-transformed available Cu (Fig. 2D), Fe (Fig. 2E) and S (Fig. 2F).

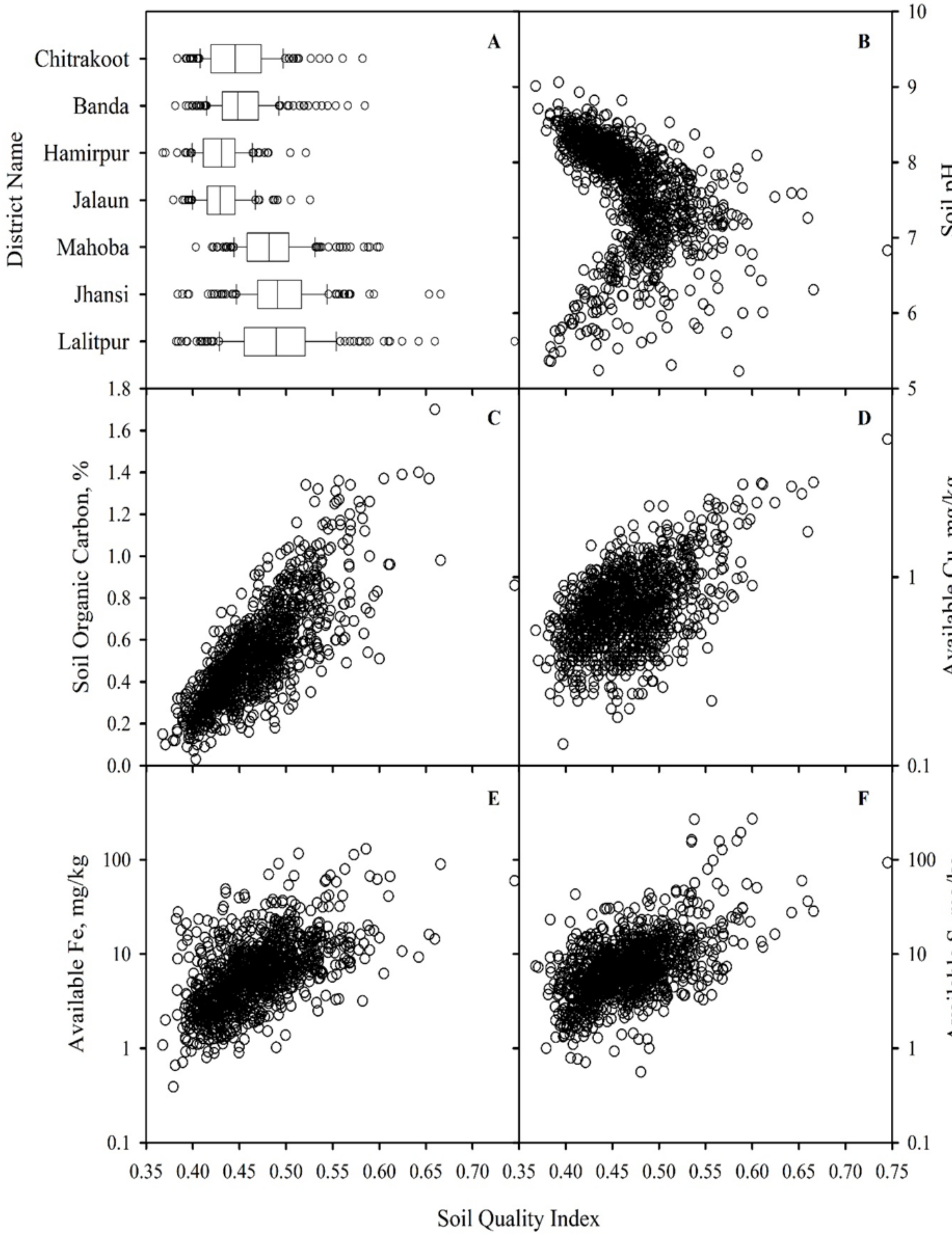

Fig. 2. (A) Average soil quality index values for different districts and (B – F) the relationship between soil quality indicators and soil quality indices in selected smallholder farms in Bundelkhand region

Table 1. Mean and the coefficient of variation (CV) expressed in percentage (in parentheses) for different soil parameters of smallholder farms in selected districts in Bundelkhand region

Indicators	Lalitpur *n* = 176	Jhansi *n* = 195	Mahoba *n* = 193	Jalaun *n* = 111	Hamirpur *n* = 101	Banda *n* = 160	Chitrakoot *n* = 177	Pooled *n* = 1113
pH	6.66 (10)	7.25 (6)	7.59 (7)	8.13 (4)	8.29 (3)	8.14 (3)	7.91 (6)	7.63 (9)
EC (dS/m)	0.20 (104)	0.23 (89)	0.22 (59)	0.12 (32)	0.23 (29)	0.19 (58)	0.14 (56)	0.19 (77)
SOC (%)	0.66 (41)	0.64 (36)	0.58 (30)	0.32 (31)	0.43 (42)	0.51 (38)	0.42 (47)	0.53 (44)
Av. P	24.4 (63)	15.7 (64)	7.9 (78)	3.4 (133)	5.5 (149)	6.3 (93)	8.9 (88)	11.1 (103)
Ex. K	77 (129)	91 (80)	167 (47)	179 (40)	214 (66)	240 (55)	147 (77)	152 (76)
Ex. Ca	1150 (56)	1296 (54)	3278 (35)	2800 (26)	3723 (34)	2930 (21)	2033 (35)	2340 (53)
Ex. Mg	206 (47)	244 (44)	294 (42)	400 (38)	455 (41)	416 (35)	261 (43)	309 (50)
Av. S	10.7 (108)	11.7 (116)	17.2 (214)	4.3 (48)	9.8 (45)	8.5 (76)	5.9 (88)	10.2 (174)
Av. Zn	1.21 (73)	0.60 (122)	0.42 (48)	0.35 (58)	0.35 (68)	0.63 (88)	0.61 (296)	0.62 (150)
Av. B	0.43 (60)	0.41 (44)	0.50 (34)	0.65 (31)	0.84 (46)	0.64 (38)	0.45 (62)	0.53 (52)
Av. Fe	20.6 (99)	8.1 (89)	9.3 (90)	4.8 (84)	3.5 (54)	4.8 (66)	6.8 (99)	8.9 (126)
Av. Cu	0.75 (85)	0.70 (63)	0.83 (47)	0.81 (33)	0.63 (44)	0.88 (35)	0.75 (57)	0.77 (56)
Av. Mn	21.7 (55)	11.1 (60)	10.4 (59)	5.1 (65)	6.1 (50)	9.8 (106)	7.5 (71)	10.8 (85)
Ex. Na	67.3 (43)	64.4 (73)	110.0 (105)	218.8 (37)	200.3 (12)	191.0 (50)	110.8 (104)	126.1 (83)

EC = Electrical conductivity; SOC = Soil organic carbon; Av. = Available; and Ex.: Exchangeable. All nutrients are in mg/kg of soil

With pH = 7 as optimum soil reaction and 'more is better' for SOC and micronutrient content, both high and low pH values yielded low SQI values. Figure 2 shows a strong linear relationship between SOC content and SQI while log-linearity may be observed for three other indicator variables. Although SQI values show a wide variability in the Lalitpur, Jhansi and Mahoba soil samples compared to those of Hamirpur and Jalaun, overall low SQI values (low median SQI) for the latter two districts once again support the conjecture that soils in Hamirpur and Jalaun are more degraded than those in the remaining districts. Soil reclamation and the addition of organic manures may be practiced immediately to restore soil quality in these two districts. Furthermore, supplementing S and Cu may also be suggested because these two parameters serve as key indicator variables and are low in these high pH soils.

Table 2. Principal components (PC) and factor loadings for different soil quality indicators

Principal components	PC1	PC2	PC3	PC4
Eigenvalues	3.84	2.99	1.32	1.06
Variance explained, %	27.46	21.33	9.44	7.57
Cumulative variance, %	27.46	48.79	58.22	65.79
Indicators		Factor loadings		
pH	**<u>0.870</u>**	0.175	0.176	0.081
EC (dS/m)	0.085	0.593	0.224	0.498
SOC (%)	0.189	**<u>0.674</u>**	0.381	0.145
Av. P	0.501	0.539	0.362	0.160
Ex. K	0.545	0.509	0.054	0.070
Ex. Ca	0.706	0.068	0.209	0.005
Ex. Mg	0.647	0.381	0.293	0.147
Av. S	0.020	0.447	0.055	**<u>0.715</u>**
Av. Zn	0.218	0.396	0.356	0.414
Av. B	0.625	0.495	0.186	0.131
Av. Fe	0.541	0.413	**<u>0.572</u>**	0.064
Av. Cu	0.085	**<u>0.654</u>**	0.488	0.159
Av. Mn	0.593	0.451	0.291	0.007
Ex. Na	0.700	0.212	0.127	0.045

*Bold face underlined factor loadings were retained in the minimum dataset; EC = Electrical conductivity, SOC = Soil organic carbon, Av. = Available, and Ex. = Exchangeable. All nutrients are in mg/kg of soil

Table 3. Pearson correlation coefficients for different soil parameters in Bundelkhand region

Indicators	pH	EC (dS/m)	SOC (%)	Av. P	Ex. K	Ex. Ca	Ex. Mg	Av. S	Av. Zn	Av. B	Av. Fe	Av. Cu	Av. Mn
EC (dS/m)	-0.02												
SOC (%)	-0.20	0.33											
Av. P	-0.44	0.24	0.53										
Ex. K	0.39	0.28	0.20	0.04									
Ex. Ca	0.57	0.06	-0.10	-0.39	0.46								
Ex. Mg	0.38	0.12	0.04	-0.16	0.41	0.41							
Av. S	-0.13	0.37	0.20	0.13	0.16	0.03	0.08						
Av. Zn	-0.15	0.12	0.31	0.34	0.08	-0.13	-0.04	0.07					
Av. B	0.44	0.27	0.29	0.05	0.49	0.32	0.57	0.12	0.06				
Av. Fe	-0.61	0.07	0.18	0.31	-0.14	-0.23	-0.05	0.10	0.11	-0.18			
Av. Cu	-0.05	0.21	0.29	0.13	0.37	0.20	0.36	0.18	0.14	0.22	0.41		
Av. Mn	-0.60	0.16	0.26	0.40	-0.07	-0.30	-0.15	0.14	0.21	-0.17	0.56	0.26	
Ex. Na	0.52	0.22	-0.12	-0.23	0.32	0.33	0.59	0.05	-0.09	0.56	-0.17	0.13	-0.21

EC = Electrical conductivity; SOC = Soil organic carbon; Av.= Available; and Ex. = Exchangeable. All nutrients are in mg/kg of soil

3.2. Relationship between SQI and agricultural productivity

Table 4 shows the average crop yield for different *kharif* and *rabi* season crops in 7 districts estimated from the questionnaire survey data. A wide variation in crop yield (CV = 10-43%) may be seen across different districts for both the seasons. Specifically, yields of sesame, mustard and chickpea showed more than 30% CV while green gram and black gram yields were almost similar across different districts. Average crop yields for crops grown in the region were found to be 30-40% lower than the state and national averages (GoI, 2020). Interestingly, farmers in Jalaun and Hamirpur grow a range of crops even with low soil quality and end up with low average yields for most crops except sesame and sorghum. Crop diversification appears to be a coping strategy against the loss of soil quality (land degradation) by smallholder farmers while sorghum and sesame ensure high returns because of their hardy nature. Farmers in Jhansi and Lalitpur, on the other hand, are selective in growing crops and, generally, get better yields than those in other districts.

Table 4. Average crop yields (kg/ha) during *kharif* (monsoon) and *rabi* (post-monsoon) seasons estimated from a household survey (n = 1403) in different districts of Bundelkhand region

District	Lalitpur	Jhansi	Mahoba	Jalaun	Hamirpur	Banda	Chitrakoot	Mean
				Kharif (monsoon) season				
Gram*	600	670	680	530	-	-	630	620
Groundnut	1280	1800	1550	-	1200	-	-	1450
Millet	-	-	-	630	-	710	420	600
Pigeonpea	-	-	530	630	840	900	770	730
Sesame	330	340	840	800	670	390	360	530
Sorghum	-	-	-	1050	880	760	740	860
				Rabi (post-monsoon) season				
Mustard	-	-	1920	1090	1350	1050	780	1240
Chickpea	2000	1650	1810	1360	970	1030	980	1400
Field pea	-	1880	1700	1470	-	1250	1220	1500
Lentil	-	-	-	950	1270	2000	-	1410
Barley	-	2990	2000	-	-	-	-	2500
Wheat	2350	3000	2530	2590	2470	2770	1670	2500

*Combined yield of green gram and black gram

Soil samples from these two districts showed high SQI values suggesting that greater agricultural income generated from high yields may empower farmers to invest more in managing their crops. For both sorghum and wheat, yield variations across the 7 districts were low. These are used as staple food and, possibly, farmers invest more on managing these crops to ensure good yields despite poor quality soils in some of the districts. Thus, similar yields from these crops may have resulted from the compulsion to manage crops that provide staple food. Figure 3 shows the relationship between average SQI and crop yields in addition to agricultural income and other parameters linked to agricultural productivity. Although we developed linear regression relationships between average SQI and yield for all the 12 crops covering *kharif* and *rabi* seasons, yield data were not available for every crop in each district (Table 4). For crops with yield data in at least four districts, we observed a moderately strong linear relationship between SQI and yield of *kharif* crops grams (R^2 = 0.52) and sorghum (R^2 = 0.73) and for *rabi* crops chickpea (R^2 = 0.67) and field pea (R^2 = 0.61). Poor relationships between SQI and crop yield were observed for groundnut, pigeonpea, sesame, mustard and wheat. In general, the increase in SQI values tends to increase crop yield for all the *rabi* crops and for legumes such as green gram, black gram and groundnut grown during *kharif* season. For the remaining *kharif* crops (pigeonpea, sesame and sorghum), low crop yields were observed with increasing SQI values. Being a deep-rooted and leguminous crop, pigeonpea may not show any relationship with SQI values obtained only for our surface soil samples. We did not find a good reason for SQI not contributing to yields in sesame and sorghum except that these are hardy crops. Positive correlation may also be observed between average SQI and the average household income across different districts (Fig. 3E). Specifically, almost 60% of the total income in Jhansi (with high average SQI) came from the agricultural sector while Chitrakoot had just 19% of total income generated coming from agriculture. Figure 3 also shows a decreasing trend (R^2 = 0.55) between average SQI values and the extent of fallow land left during *kharif* season (data source: NRSC, 2017). Farmers are expected to utilize better quality soils leaving low quality soils fallow.

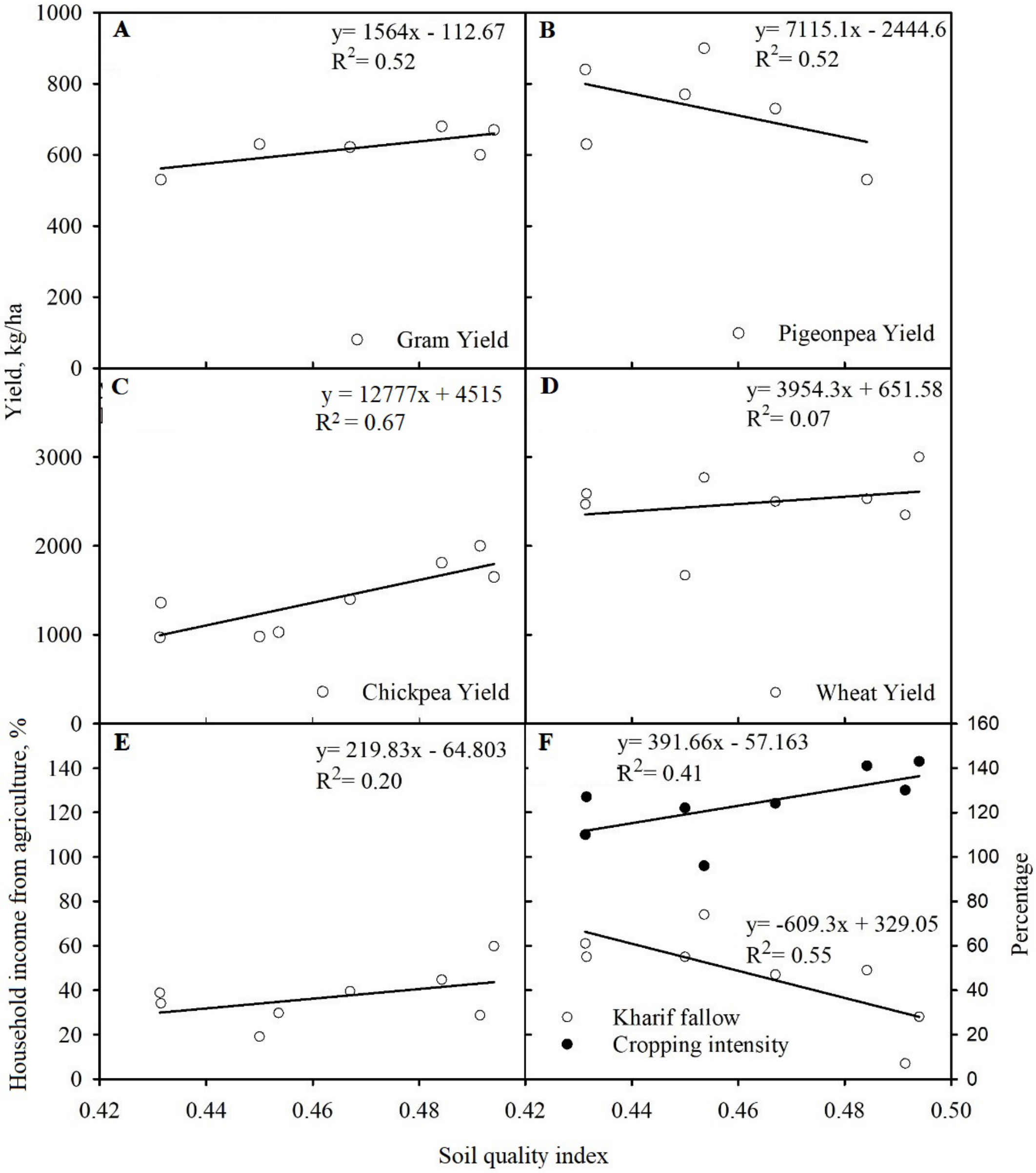

Fig. 3. Relationships between soil quality index and crop yields for (A) gram (green gram and black gram combined), (B) pigeonpea, (C) chickpea, (D) wheat and with (E) fraction of total income from agriculture and (F) cropping intensity and fallow land during monsoon season for Bundelkhand region. Legends show linear regression equations and corresponding coefficient of determination (R^2) values

Estimated SQI values for Jalaun and Hamirpur districts were the least (average SQI = 0.43) and both of these districts showed 55-61% of their agricultural land left fallow during the *kharif* season. On the other hand, Lalitpur and Jhansi districts with high average SQI values (i.e., 0.49) have only 7-28% of their land fallow during the *kharif* season. A positive trend between SQI and cropping intensity may also be seen in Fig. 3F with the highest average cropping intensity of 143% observed in Jhansi. These results suggest that soil quality may be a key driver of both land use (cropping and cropping intensity) and farm income. A rapid assessment approach such as DRS may provide ways and means to quickly assess this composite soil parameter in smallholder farms.

3.3. Assessment of soil parameters and SQI in smallholder farms using DRS approaches

Processed reflectance spectra, scree plot and biplots depicting spectral characteristics for the collected soil samples are shown in Figure 4. In general, soil samples from Lalitpur and Jhansi showed greater overall spectral reflectance values than those of Jalaun and Hamirpur districts despite the later having low SOC content (Table 1). With high available Fe and Mn content, soils from Lalitpur and Jhansi may be more reflective than soils of Jalaun and Hamirpur with low Fe and high Na. Spectral reflectance data for the pooled sample showed reflectance characteristics similar to those of Chitrakoot and Mahoba districts (Fig. 4A). With wide variability in spectral reflectance data, four leading PCs for the pooled dataset could explain only 67% of the spectral variance (Fig. 4B); the first two PCs explained only 51% of the total spectral variation. Johnson et al. (2019) showed that three leading PCs could explain about 79% of variance in spectral reflectance data collected from farmers' fields across three different production systems in Africa. The biplot between standardized PC1 and PC2 (Fig. 4C) shows the extent of similarity among soil spectra across different districts. Specifically, samples from Banda, Chitrakoot and Hamirpur districts show similar spectral characteristics. With close proximity to the biplot origin, these samples have less spectral variability compared to those of other districts. Samples from Lalitpur, Jhansi and Mahoba cover a large area of the biplot space indicating the presence of large variability in their spectral signatures. The biplot for the Jalaun spectra showed both high and low variability (Fig. 4D).

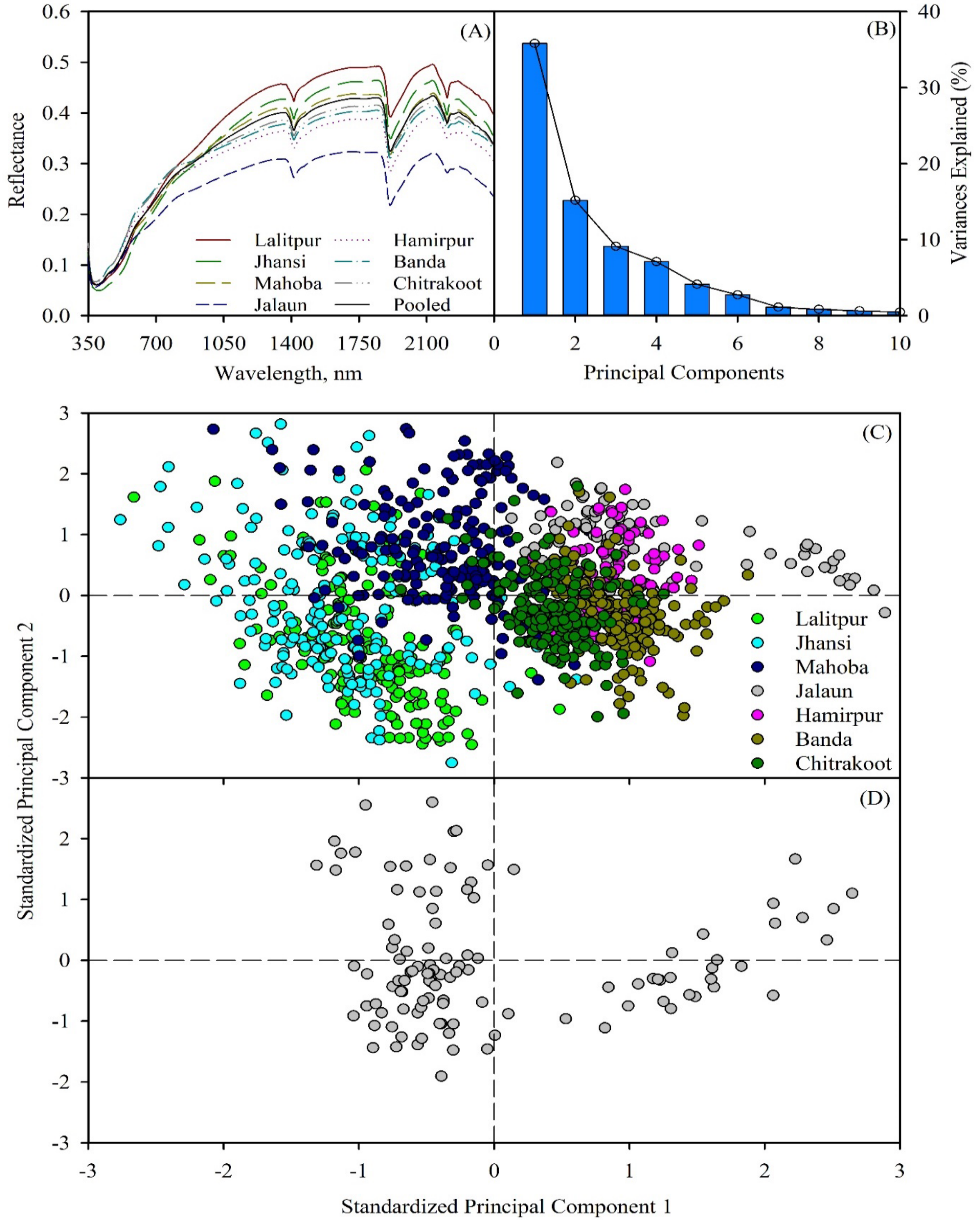

Fig. 4. (A) Average reflectance spectra, (B) scree plot showing the cumulative variance explained by principal components of spectra, (C) biplots showing soil spectral variability across all the 7 districts and (D) for Jalaun district in Bundelkhand region

The total area covered in Jalaun (377 km^2) was more than those in Lalitpur (52 km^2) and Jhansi (74 km^2). Overall, Jalaun soil samples showed low CV values for several soil properties compared to those of other districts (Table 1). With such a large area, it may be possible that Jalaun's samples have two different groups of soil characteristics: one with high and the other with low spectral variability. While such distinctly diverse variability is ideal to develop chemometric models for estimating soil properties, it also captures the fact that the creation of multiple smallholder farms because of land fragmentation eventually lead to varied soil properties as each piece of landholding is managed differently by different farmers.

Table 5 shows the best validation R^2 and bias values for the 120 test cases (15 parameters in 7 districts and in the pooled dataset). Corresponding RMSE and RPD values are provided in Table S4 (Supplementary Materials). Similarly, the best performing spectral transformation and property transformation for these 120 test cases are listed in Table S5. Table 5 shows that exchangeable Ca and K were the best estimated soil parameters with the R^2 values ranging from 0.79 (Jalaun) to 0.94 (Lalitpur) and from 0.67 (Jhansi) to 0.88 (Lalitpur), respectively. The RPD values for exchangeable Ca were high (ranging from 2.21 to 4.31). Soil organic carbon (R^2 range = 0.52 to 0.86; RMSE range = 0.05-0.14%) and available B (R^2 range = 0.53 to 0.82; RMSE range = 0.08-0.15 mg/kg) were the next best estimated soil parameters. Pooled datasets also showed that chemometric models performed best for exchangeable Ca (R^2 = 0.85; bias = 63.17 mg/kg), followed by soil pH (R^2 = 0.83; bias = 0.005). Available S (R^2 range = 0.22-0.56) and EC (R^2 range = 0.24-0.51) values were the poorly estimated soil parameters. The SQI values showed high validation statistics across different districts with the Banda soils showing the highest R^2 value of 0.81 (RMSE = 0.01; RPD = 2.31). The Mahoba soil samples showed the minimum estimation accuracy for SQI (R^2 = 0.41; RPD = 1.32) even though this district had a large number of soil samples and almost all the indicator variables were generally well predicted through the DRS approach. Estimated bias values for SQI ranged from -0.0014 in Banda's soil samples to +0.0045 in Mahoba's samples. With such low bias values and the R^2 value of 0.71 in the pooled dataset, the DRS approach may be conveniently used for estimating SQI in smallholder farms. Overall, the $PLSR_{FS}$ approach showed the best performance in 60% of the 120 cases; the SVR model performed better than $PLSR_{FS}$ in the remaining cases.

Since several test cases were considered for modelling, we summarize some of the key results in Figure 5 for illustration. Interestingly, the best validation statistics were seen when soil parameters were log-normally distributed (48% of the 120 cases) or showed normal distribution (Fig. 5A) without any transformation (33% of the 120 cases). Similarly, untransformed spectra, SNV and FD transformations performed equally well showing best validation statistics accounting for almost 78% of the 120 cases (Fig. 5B). Almost 58% of the 120 best performing validation models showed R^2 values exceeding 0.60; about 20% cases showed R^2 values in excess of 0.8, which are considered as an excellent performance for the DRS approach. This suggests that normality in estimated soil parameters is an important trait even when complex feature selection-based models are employed in chemometric modelling. Figures 5C to 5F show observed vs. predicted soil properties. Among all the parameters, exchangeable Ca showed the best prediction for the Lalitpur samples (Fig. 5C) and in the pooled datasets (Fig. 5F). Widely scattered data along 1:1 line in Figure 5D shows the extent of error in the predicted data for one of the most poorly predicted soil properties in our dataset. We also examined the effects of sample number on R^2 statistic. There was no systematic improvement in validation R^2 values when sample number increased from 101 in the Hamirpur soils to 193 in the Mahoba soils. However, the sampling density appeared to slightly increase the validation R^2 values although low R^2 values were also seen when sampling density increased because of the decrease in the total sampling area, as in the case of Hamirpur (21 km^2) and Mahoba (42 km^2) soils. This again illustrates the importance of variability in the dataset as a requirement for better performance of the DRS approach.

Table 5. Coefficient of determination (R^2) and bias values for the partial least square regression with feature selection and support vector regression (denoted by an asterisk) models in the validation datasets of Bundelkhand region

Soil Indicators	Lalitpur n = 176		Jhansi n = 195		Mahoba n = 193		Jalaun n = 111		Hamirpur n = 101		Banda n = 160		Chitrakoot n = 177		Pooled n = 1113	
	R^2	Bias	R^2	Bias	R^2	Bias	R^2	Bias	R^2	Bias	R^2	Bias	R^2	Bias	R^2	Bias
pH	0.77*	-0.01	0.47	-0.03	0.76	-0.03	0.56	-0.04	0.53	-0.01	0.63*	-0.03	0.68*	0.01	0.83	0.00
EC (dS/m)	0.33*	-0.04	0.30*	-0.02	0.24*	-0.001	0.26*	0.00	0.39	0.00	0.39	-0.01	0.51	0.00	0.40*	0.00
SOC (%)	0.64	0.02	0.86	0.00	0.66*	0.02	0.52	0.02	0.81*	-0.01	0.86	0.01	0.75*	0.00	0.75*	0.00
Av. P	0.37	-0.96	0.35	0.65	0.43	-0.10	0.38	-0.30	0.79	-0.71	0.86	-0.65	0.63	-0.03	0.62*	0.29
Ex. K	0.88	4.03	0.67	-2.04	0.85	-3.34	0.71*	-8.36	0.87*	-18.22	0.85	10.35	0.82	-0.77	0.72*	0.06
Ex. Ca	0.94	19.05	0.92*	6.24	0.91	101.40	0.79	5.57	0.93*	-1.60	0.84	46.55	0.80	20.44	0.85*	63.17
Ex. Mg	0.76*	9.77	0.82	-7.18	0.68	-2.48	0.60*	9.19	0.43*	44.33	0.69*	4.57	0.59*	6.83	0.73	-0.02
Av. S	0.29*	-0.15	0.50*	-0.46	0.56*	-1.17	0.22	0.14	0.28	0.04	0.50	-0.86	0.55*	-0.89	0.45	-0.87
Av. Zn	0.83	-0.04	0.66	-0.06	0.65*	-0.02	0.24	0.00	0.33	-0.01	0.78	-0.02	0.63	0.03	0.54	-0.00
Av. B	0.82	-0.02	0.77*	0.01	0.65*	0.00	0.53	0.00	0.76	-0.07	0.72*	0.01	0.72*	0.00	0.61*	-0.02
Av. Fe	0.81	-0.42	0.50*	0.19	0.78*	0.17	0.75*	-0.45	0.26	0.10	0.72*	-0.12	0.68	0.14	0.67	-1.03
Av. Cu	0.90	0.01	0.75*	-0.02	0.54*	-0.02	0.61*	-0.04	0.20	0.03	0.61	-0.02	0.80	0.01	0.43	-0.04
Av. Mn	0.34	0.14	0.49	-0.21	0.29	-0.35	0.69	0.21	0.29	-0.03	0.52	-0.43	0.48	-5.00	0.61	-0.95
Ex. Na	0.58*	1.96	0.54	-0.38	0.49*	-0.62	0.47	-5.46	0.56*	-0.74	0.51	-3.58	0.38*	-12.63	0.48	-3.28
SQI	0.63	0.00	0.64*	0.00	0.41	0.00	0.68*	0.00	0.66	0.00	0.81*	0.00	0.64	0.00	0.71	0.00

EC = Electrical conductivity; SOC = Soil organic carbon; Av.= Available; Ex.= Exchangeable and SQI = Soil quality index. All nutrients are in units of mg/kg of soil.

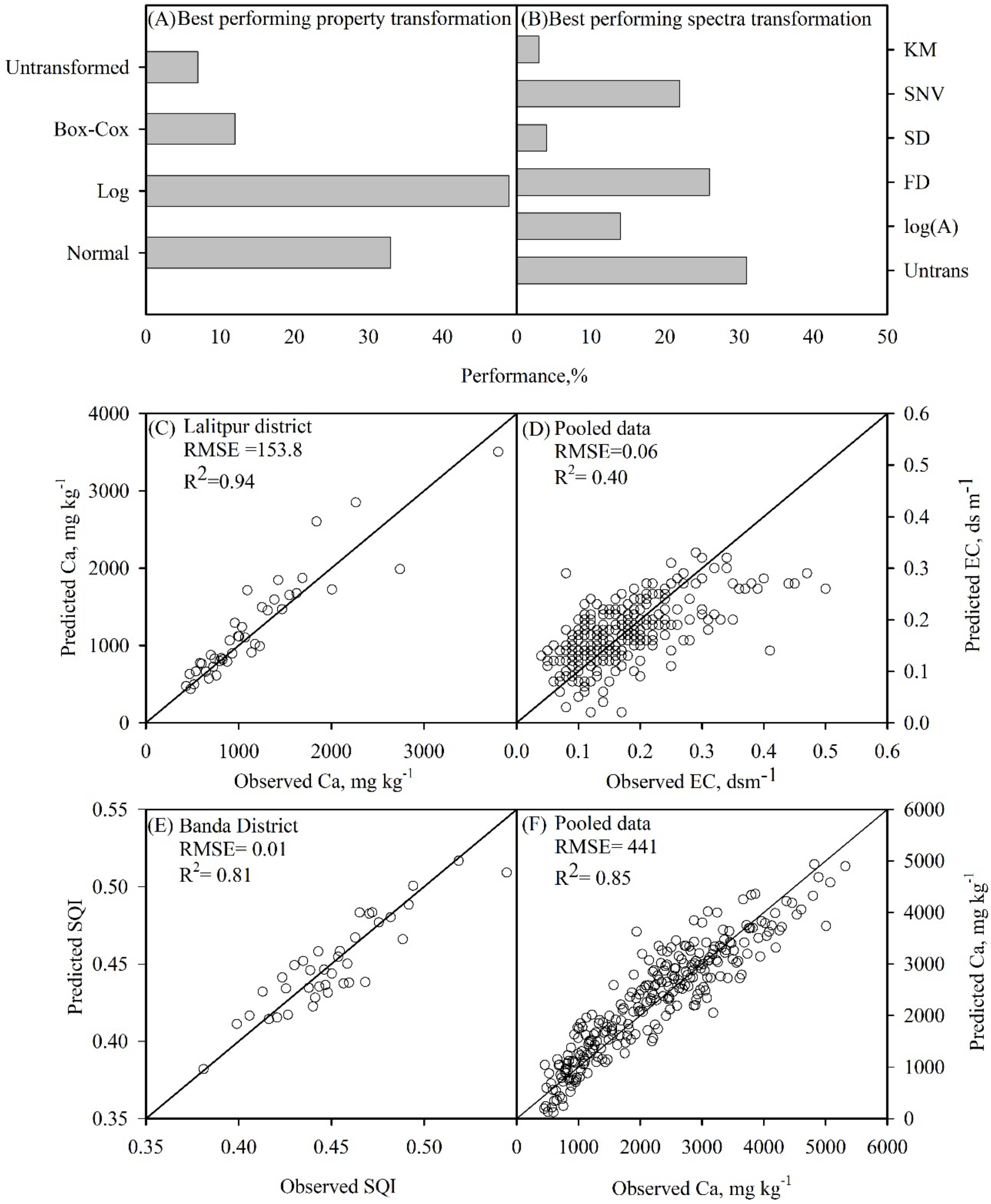

Fig. 5. Percentage of samples showing best transformations for (A) soil properties and (B) soil spectra; and (C, D, E, F) selected scatter plots between observed vs. predicted parameters in Bundelkhand region

Table 6. Agreement (expressed in %) between the soil test crop response classes for soil parameters measured by the wet chemistry approach and estimated using diffuse reflectance spectroscopy for the validation datasets in Bundelkhand region

Indicators	Lalitpur	Jhansi	Mahoba	Jalaun	Hamirpur	Banda	Chitrakoot	Pooled	Soil classification ratings		
									Low	Medium	High
pH	86	95	98	100	96	100	98	81	< 6.5	6.5 - 8.7	> 8.7
EC (dS/m)	100	100	100	100	100	100	98	99	< 0.8	0.8 - 2.5	> 2.5
SOC (%)	63	81	83	92	92	87	86	83	< 0.40	0.40 - 0.75	> 0.75
Av. P	85	56	63	81	80	66	66	58	< 5.0	5.0 - 12.5	> 12.5
Ex. K	98	98	81	62	76	61	79	97	< 151	151 - 250	> 250
Ex. Ca	95	96	87	89	100	97	81	95	< 2000	2000 - 4000	> 4000
Ex. Mg	98	96	89	73	58	74	93	100	< 396	396 - 996	> 996
Av. S	72	78	69	100	60	87	86	78	< 10	10 - 20	> 20
Av. Zn	76	98	100	100	100	100	100	100	< 1.21	1.21 -2.40	> 2.40
Av. B	98	100	100	96	92	89	92	100	< 1	1 - 2	> 2
Av. Fe	76	80	70	96	100	100	88	83	< 9.1	9.1 - 27	> 27
Av. Cu	71	79	85	85	100	90	86	72	< 0.41	0.41 - 1.2	> 1.2
Av. Mn	76	79	89	81	75	69	30	76	< 4.10	4.1 - 16	> 16

EC = Electrical conductivity; SOC = Soil organic carbon; Av. = available; and Ex. = exchangeable. All nutrients are in mg/ kg of soil

Most smallholder farm owners in developing countries look for soil test results and corresponding STCR ratings to decide the amount of fertilizer to be applied to a specific crop. Although the DRS approach is generally evaluated using R^2 and RMSE statistics (Table 5 and Table S4), we also converted measured (wet chemistry) and DRS-estimated (dry chemistry) soil parameters into their respective STCR ratings using a 3-tier classification scheme. Table 6 shows the ratings and the degree of agreement between the wet and dry chemistry-based ratings in the validation datasets. Similar STCR ratings were obtained for both approaches in 30-100% of the samples (average = 86%) for validation datasets. Except for available S, P and Mn, the DRS-estimated STCR ratings generally agreed with the conventional soil test rating in more than 70% of the soil samples. With such a close agreement in STCR ratings, the DRS approach may safely be used in farmers' fields to make nutrient recommendations much akin to the traditional wet chemistry approach of soil testing.

4. Discussion

Smallholder farmers in developing countries focus largely on applying plant nutrients (with attention generally on NPK) with limited soil testing and without information on the nature and extent of nutrient availability in the soil and/or its connection with other factors of crop production. We examined the DRS approach as an alternative to wet chemistry-based (conventional) STCR ratings. We calibrated two efficient chemometric models by using a soil spectral library compiled from 1113 smallholder farms representing a large geographical area. We analyzed chemical properties and nutrient content and developed SQIs. The SQI framework revealed the complex linkages among soil parameters. For example, results from Lalitpur and Hamirpur clearly showed that variations in pH make an important difference in Fe and Ca availability, which is expected to directly influence grain and fodder yields. Thus, soil pH may be a key driver for crop productivity in these districts. Similarly, a key understanding of Cu, Fe and S nutrition both in terms of content and distribution may also be needed for ensuring balanced nutrition in smallholder farms. Poorly managed farms in an inherently degraded landscape provided us an opportunity to exclusively link SQI with agricultural productivity parameters (yield, cropping intensity and fraction of agricultural income). Most studies on soil quality assessment show poor linkage between SQI values and agricultural productivity parameters such as crop yield (Roper et al., 2017), possibly because management factors may have masked the effects of soil quality on yield traits. Cultivation in poorly managed smallholder farms (without the external input of plant nutrients) in our study provided the opportunity to capture the effects of soil quality alone on yield parameters. Results showed that *rabi* crops grown in good quality soils generally produced more yield whereas *kharif* crops did not show such effects unless farmers cultivated short-duration legumes. Positive correlation between SQI and cropping intensity suggested that farmers with good quality soils may find it lucrative to grow more than one crop during a year, deriving a larger fraction of income from agriculture while negative correlation between SQI and fallow land suggested that farmers may find it less remunerative to grow crops in farms having low quality soils. Therefore, while good quality soils may empower smallholder farmers to better manage their crops with increased cropping intensity, poor quality soils may be compelling them to diversify cropping to ensure some return to support their livelihood. These results suggest that SQI may be a comprehensive soil parameter providing insights into land use, crop yield and overall agricultural productivity in smallholder farms.

High validation statistics for 14 different soil parameters for the chemometric models in this study clearly suggest that this non-invasive and rapid soil testing technology has the potential to reach millions of smallholder farmers though it is currently expensive to buy a spectroradiometer. A critical advantage of the DRS approach over the conventional soil testing approach is that the former is environmentally friendly because it does not need any chemical reagent. Specifically, strong agreement in the STCR ratings obtained for the conventional and

DRS-based soil testing clearly suggests that the DRS approach may safely be used for making nutrient recommendations. Not only individual soil parameters, the DRS approach also showed high validation statistics for SQI values for all the districts and in the pooled datasets. We propose that the DRS approach be promoted as a service similar to wet chemistry-based STCR ratings provided by conventional soil testing laboratories.

5. Conclusion

Analyses of soil property data showed a strong pH control on nutrient availability and resulting SQI values across different districts in Bundelkhand region. Similarly, estimated SQI values showed that soil quality strongly influences crop productivity in smallholder farms. With strong positive correlation with cropping intensity and agricultural income and negative correlation with the extent of fallow land left during the monsoon season, SQI may be considered a key driver for crop yield and farm income in smallholder farms. We evaluated two efficient chemometric models, 6 spectral transformation and 3 variable transformation approaches for all the 15 parameters across 7 districts and for the pooled dataset. Analysis of best performing transformations and chemometric models showed that exchangeable Ca (R^2 range = 0.79 to 0.94) and K (R^2 range = 0.67 to 0.88) are the best estimated soil parameters followed by SOC (R^2 range = 0.52 to 0.86) and available B content (R^2 range = 0.53 to 0.81). Except for one district, SQI values showed high R^2 values (0.63 to 0.81) with an overall R^2 value of 0.71 in the pooled dataset. Comparison of STCR ratings for the wet chemistry and DRS-estimated soil parameters showed a high degree of agreement between these two methods of obtaining soil test recommendations. On an average, 86% of the samples showed identical STCR ratings for the wet chemistry and DRS approaches, suggesting that the latter approach may be used as an alternative to conventional soil testing on smallholder farms. The DRS approach for estimating both individual soil properties and the collective indicator of soil quality in terms of SQI provides an opportunity to empower land managers to identify corrective and exploitative measures to manage smallholder farms. With many smallholder farms facing the daunting challenge of land degradation, the opportunity to rapidly test large number of soil samples may help smallholder farmers in many developing countries. Specifically, nutrient recommendations through the DRS approach could be a game changer given that soil quality management is a top priority of governments and a key agenda of the United Nations Sustainable Development Goals.

Acknowledgement:

We acknowledge the funding support received from the Rastriya Krishi Vikas Yojana (RKVY) through the Department of Agriculture, Government of Uttar Pradesh, toundertake the Doubling Farmers' Income (DFI) initiative in Uttar Pradesh's Bundelkhand region, Central India.

References

Ahmadi, A., Emami, M., Daccache, A. and He, L., 2021. Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis. Agronomy, 11(3), p.433.

Andrews, S.S. and Carroll, C.R., 2001. Designing a soil quality assessment tool for sustainable agroecosystem management. Ecological Applications, 11(6), pp.1573-1585.

Andrews, S.S., Karlen, D.L. and Mitchell, J.P., 2002. A comparison of soil quality indexing methods for vegetable production systems in Northern California. Agriculture, ecosystems & environment, 90(1), pp.25-45.

Armenise, E., Redmile-Gordon, M.A., Stellacci, A.M., Ciccarese, A. and Rubino, P., 2013. Developing a soil quality index to compare soil fitness for agricultural use under different managements in the Mediterranean environment. Soil and Tillage Research, 130, pp.91-98.

Babaeian, E., Homaee, M., Montzka, C., Vereecken, H. and Norouzi, A.A., 2015. Towards retrieving soil hydraulic properties by hyperspectral remote sensing. Vadose Zone Journal, 14(3), pp.1-17.

Bai, Z., Xie, M., Hu, B., Luo, D., Wan, C., Peng, J. and Shi, Z., 2022. Estimation of Soil Organic Carbon using Vis-NIR Spectral Data and Spectral Feature Bands Selection in Southern Xinjiang, China. Sensors, 22(16), p.6124.

Barra, I., Haefele, S.M., Sakrabani, R. and Kebede, F., 2021. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. Trends in Analytical Chemistry, 135, p.116166.

Bellinaso, H., Demattê, J.A.M. and Romeiro, S.A., 2010. Soil spectral library and its use in soil classification. Revista Brasileira de Ciência do Solo, 34, pp.861-870.

Ben-Dor, E. and Banin, A., 1995. Near- infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal, 59(2), pp.364-372.

Benedet, L., Faria, W.M., Silva, S.H.G., Mancini, M., Demattê, J.A.M., Guilherme, L.R.G. and Curi, N., 2020. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. Geoderma, 376, p.114553.

Chang, C.W., Laird, D.A., Mausbach, M.J. and Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Science Society of America Journal, 65(2), pp.480-490.

Coblinski, J.A., Giasson, É., Demattê, J.A., Dotto, A.C., Costa, J.J.F. and Vašát, R., 2020. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. Catena, 189, p.104485.

Davari, M., Karimi, S.A., Bahrami, H.A., Hossaini, S.M.T. and Fahmideh, S., 2021. Simultaneous prediction of several soil properties related to engineering uses based on laboratory Vis-NIR reflectance spectroscopy. Catena, 197, p.104987.

Dev, I., Singh, R., Garg, K.K., Ram, A., Singh, D., Kumar, N., Dhyani, S.K., Singh, A., Anantha, K.H., Akuraju, V. and Dixit, S., 2022. Transforming livestock productivity through watershed interventions: A case study of Parasai-Sindh watershed in Bundelkhand region of Central India. Agricultural Systems, 196, p.103346.

Dorantes, M.J., Fuentes, B.A. and Miller, D.M., 2022. Calibration set optimization and library transfer for soil carbon estimation using soil spectroscopy—A review. Soil Science Society of America Journal, 86(4), pp.879-903.

Dotto, A.C., Dalmolin, R.S.D., ten Caten, A. and Grunwald, S., 2018. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma, 314, pp.262-274.

Edrisi, S.A., Sarkar, P., Son, J., Prakash, N.T. and Bara, H., 2022. Assessing the realization of global land restoration: A meta-analysis. Anthropocene Science, pp.1-16.

Garg, K.K., Singh, R., Anantha, K.H., Singh, A.K., Akuraju, V.R., Barron, J., Dev, I., Tewari, R.K., Wani, S.P., Dhyani, S.K. and Dixit, S., 2020. Building climate resilience in degraded agricultural landscapes through water management: a case study of Bundelkhand region, Central India. Journal of Hydrology, 591, p.125592.

GEF, 2019. Global Environment Facility. 2019. The GEF Monitoring Report 2019. Washington DC.

Gholizadeh, A., Neumann, C., Chabrillat, S., van Wesemael, B., Castaldi, F., Borůvka, L., Sanderman, J., Klement, A. and Hohmann, C., 2021. Soil organic carbon estimation using VNIR–SWIR spectroscopy: The effect of multiple sensors and scanning conditions. Soil and Tillage Research, 211, p.105017.

Girden, E.R. and Kabacoff, R. eds., 2010. Evaluating research articles from start to finish. Sage.

GoI, 2020. Government of India. 2020. Agricultural Statistics at a glance 2019. Ministry of Agriculture and Farmers Welfare, Department of Agriculture Cooperation and Farmers Welfare. Directorate of Economics and Statistics, Government of India.

Gomez, C., Le Bissonnais, Y., Annabi, M., Bahri, H. and Raclot, D., 2013. Laboratory Vis–NIR spectroscopy as an alternative method for estimating the soil aggregate stability indexes of Mediterranean soils. Geoderma, 209, pp.86-97.

Gupta, A., Das, B.S., Kumar, A., Chakraborty, P. and Mohanty, B., 2016. Rapid and noninvasive assessment of Atterberg limits using diffuse reflectance spectroscopy. Soil Science Society of America Journal, 80(5), pp.1283-1295.

Gupta, A., Vasava, H.B., Das, B.S. and Choubey, A.K., 2018. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. Geoderma, 325, pp.59-71.

Hanway, J.J. and Heidel, H., 1952. Soil analysis methods as used in Iowa state college soil testing laboratory. Iowa agriculture, 57, 1-31.

Horf, M., Bönecke, E., Gebbers, R., Kling, C., Kramer, E., Rühlmann, J., Schröter, I., Schwanghart, W. and Vogel, S., 2022. Utility of visible and near-infrared spectroscopy to predict base neutralizing capacity and lime requirement of quaternary soils. Precision Agriculture, pp.1-22.

Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D. and Saito, K., 2019. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. Geoderma, 354, p.113840.

Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educational and psychological measurement, 20(1), pp.141-151.

Keren. R., 1996. Boron. In: Methods of Soil Analysis, Part 3 Chemical methods. (Sparks, D.L., Page, A.L. Eds.) Soil Science Society of America and American Society of Agronomy. Madison, Wisconsin, USA. Pp. 603-626.

Kishore, A., Alvi, M. and Krupnik, T.J., 2021. Development of balanced nutrient management innovations in South Asia: perspectives from Bangladesh, India, Nepal, and Sri Lanka. Global Food Security, 28, p.100464.

Kumar, D., Meena, M.K., Yadav, R.S., and Pramanik, M., 2021. Exploring Conservation Agricultural Practices in Bundelkhand Region, Central India. *In.* S. Jayaraman et al. (eds.), Conservation Agriculture: A Sustainable Approach for Soil Health and Food Security, https://doi.org/10.1007/978-981-16-0827-8_9

Laborde, D., Porciello, J., Smaller, C., Murphy, S. and Parent, M., 2020. Ceres2030: Sustainable Solutions to end Hunger Summary Report.

Li, S., Viscarra Rossel, R.A. and Webster, R., 2022. The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon. European Journal of Soil Science, 73(1), p.e13202.

Ludwig, B., Murugan, R., Parama, V.R. and Vohland, M., 2019. Accuracy of estimating soil properties with mid-infrared spectroscopy: Implications of different chemometric approaches and software packages related to calibration sample size. Soil Science Society of America Journal, 83(5), pp.1542-1552.

McBride, M.B., 2022. Estimating soil chemical properties by diffuse reflectance spectroscopy: Promise versus reality. European Journal of Soil Science, 73(1), p.e13192.

Mohanty, B., Gupta, A. and Das, B.S., 2016. Estimation of weathering indices using spectral reflectance over visible to mid-infrared region. Geoderma, 266, pp.111-119.

Nature Editorial, 2020. Ending hunger: Science must stop neglecting smallholder farmers. Nature, 586, p.336.

Nesme, T., 2022. Fertilising smallholder crops. Nature Sustainability, 5(1), pp.11-12.

Ng, W., Minasny, B., Malone, B. and Filippi, P., 2018. In search of an optimum sampling algorithm for prediction of soil properties from infrared spectra. PeerJ, 6, p.e5722.

NRSC, 2017. National Remote Sensing Center. 2017. Land use land cover 2017-18. https://bhuvan-app1.nrsc.gov.in/2dresources/thematic/LULC250/MAP/1718.pdf

Olsen, S.R. and Sommers, L.E. 1982. Phosphorus. Pages 403-430 in Methods of Soil Analysis. Part II (Page, A.L., Miller, R.H., and Keeney, D.R., eds.). 2nd edition. Madison, Wisconsin, USA: American Society of Agronomy and Soil Science Society of America.

Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives–A review. Analytica chimica acta, 1026, pp.8-36.

Rai, M., Acharya, S.S., Virmani, P.K., Aggarwal, P.K., 2009. State of Indian agriculture. NAAS Publ, New Delhi.

Reddy, N.N., Chakraborty, P., Roy, S., Singh, K., Minasny, B., McBratney, A.B., Biswas, A. and Das, B.S., 2021. Legacy data-based national-scale digital mapping of key soil properties in India. Geoderma, 381, p.114684.

Rhoades, J.D., 1996. Salinity, Electrical Conductivity and total dissolved solids. In: Methods of Soil analysis, part 3 Chemical methods (Sparks, D.L., Page, A.L. Eds.). Soil Science Society of America and American Society of Agronomy. Madison, Wisconsin, USA. pp. 417-435.

Ricciardi, V., Mehrabi, Z., Wittman, H., James, D. and Ramankutty, N., 2021. Higher yields and more biodiversity on smaller farms. Nature Sustainability, 4(7), pp.651-657.

Roper, W.R., Osmond, D.L., Heitman, J.L., Wagger, M.G. and Reberg-Horton, S.C., 2017. Soil health indicators do not differentiate among agronomic management systems in North Carolina soils. Soil Science Society of America Journal, 81(4), pp.828-843.

Santra, P., Sahoo, R.N., Das, B.S., Samal, R.N., Pattanaik, A.K. and Gupta, V.K., 2009. Estimation of soil hydraulic properties using proximal spectral reflectance in visible, near-infrared, and shortwave-infrared (VIS–NIR–SWIR) region. Geoderma, 152(3-4), pp.338-349.

Sarathjith, M.C., Das, B.S., Wani, S.P. and Sahrawat, K.L., 2014. Dependency measures for assessing the covariation of spectrally active and inactive soil properties in diffuse reflectance spectroscopy. Soil Science Society of America Journal, 78, pp.1522-1530.

Sarathjith, M.C., Das, B.S., Wani, S.P. and Sahrawat, K.L., 2016. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. Geoderma, 267, pp.1-9.

Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.

Sendhil, R., Kumar, A., Sharma, A.K., Jasrotia, P., Gupta, O.P., Meena, R.P., Singh, S. and Singh, G.P., 2018. Strengthening value chain in wheat and barley for doubling farmers income. Directorate of Extension, Department of Agriculture Cooperation & Farmers Welfare and ICAR-Indian Institute of Wheat and Barley Research, pp.1-144.

Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J. and Ge, Y., 2022. A global soil spectral calibration library and estimation service. Soil Security, 7, p.100061.

Singh, K., Vasava, H.B., Snoeck, D., Das, B.S., Yinil, D., Field, D., Sanderson, T., Fidelis, C., Majeed, I. and Panigrahi, N., 2019. Assessment of cocoa input needs using soil types and soil spectral analysis. Soil Use and Management, 35(3), pp.492-502.

Singh, R., Akuraju, V., Anantha, K.H., Garg, K.K., Barron, J., Whitbread, A.M., Dev, I. and Dixit, S., 2022. Traditional rainwater management (haveli cultivation) for building system level resilience in a fragile ecosystem of Bundelkhand Region, Central India. Front. Sustain. Food Syst. 6:826722.

Singh, R., Garg, K.K., Wani, S.P., Tewari, R.K. and Dhyani, S.K., 2014. Impact of water management interventions on hydrology and ecosystem services in Garhkundar-Dabar watershed of Bundelkhand region, Central India. Journal of Hydrology, 509, pp.132-149.

Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and computing, 14(3), pp.199-222.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L. and van Wesemael, B., 2013. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. PloS one, 8(6), p.e66409.

Teofilo, R.F., Martins, J.P.A. and Ferreira, M.M., 2009. Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. Journal of Chemometrics: A Journal of the Chemometrics Society, 23(1), pp.32-48.

The MathWorks, Inc. (R2021b). Natick, Massachusetts, United States. Matlab.

Thomas, G.W., 1996. Soil pH and soil acidity. In: Methods of Soil Analysis, Part 3: Chemical methods. (Sparks, D.L., Page, A.L. Eds.) Soil Science Society of America and American Society of Agronomy. Madison, Wisconsin, USA. pp. 475-490.

Vasava, H.B. and Das, B.S., 2022. Assessment of soil properties using spectral signatures of bulk soils and their aggregate size fractions. Geoderma, 417, p.115837.

Vasava, H.B., Gupta, A., Arora, R. and Das, B.S., 2019. Assessment of soil texture from spectral reflectance data of bulk soil samples and their dry-sieved aggregate size fractions. Geoderma, 337, pp.914-926.

Vasu, D., Singh, S.K., Ray, S.K., Duraisami, V.P., Tiwary, P., Chandran, P., Nimkar, A.M. and Anantwar, S.G., 2016. Soil quality index (SQI) as a tool to evaluate crop productivity in semi-arid Deccan plateau, India. Geoderma, 282, pp.70-79.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V. and Aïchi, H., 2016. A global spectral library to characterize the world's soil. Earth-Science Reviews, 155, pp.198-230.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J. and Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma, 131(1-2), pp.59-75.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Ge, Y., Gomez, C., Guerrero, C., Peng, Y., Ramirez-Lopez, L. and Shi, Z., 2022. Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. European Journal of Soil Science, 73(4), p.e13271.

Walkley, A. and Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil science, 37(1), pp.29-38.

Wang, L. and Wang, R., 2022. Determination of soil pH from Vis-NIR spectroscopy by extreme learning machine and variable selection: a case study in lime concretion black soil. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, p.121707.

Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. *In*: Near-Infrared Technology in the Agricultural and Food Industries (eds P. Williams & K. Norris), pp. 143–167. American Association of Cereal Chemists Inc., Saint Paul, MN.