This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

Optimizing Temporal Window for Constructing Efficient Bare Soil Composites from Sentinel-2 Data for Estimating Soil Properties in Semi-Arid Tropics

Israr Majeed[1], Debi P Sahoo[1], Sourav Roy[1], Nagarjuna N. Reddy[1], Nagaraju Budama[1], A. Venkataradha[1], Ramesh Singh[1], Bhabani S. Das[2], Kaushal K. Garg[1,*]

Abstract **- Bare soil composites (BSCs) derived from Sentinel-2 (S2) imagery offer a promising approach for estimating soil properties. However, the influence of temporal window length and spatial heterogeneity on predictive performance remains insufficiently understood. This study evaluated BSCs developed from short-term (one-year), medium-term (three-year), and long-term (six-year) temporal windows for estimating soil properties in a semi-arid agricultural region of Maharashtra, India. A total of 443 surface soil samples were collected from ground reference locations distributed across five districts representing diverse soil conditions and land-use practices. Four regression models, namely support vector regression (SVR), partial least squares regression (PLSR), random forest (RF), and feature selection-based PLSR ($PLSR_{FS}$), were evaluated using Sentinel-2 spectral data. Laboratory reflectance spectra were also compared with each BSC using the Spectral Angle Mapper (SAM) to identify the Sentinel-2 dataset with the highest spectral agreement. Moreover, the fuzzy c-means (FCM) membership function was explored as a novel measure of soil spectral heterogeneity for use as a predictor variable along with S2 data. Results showed that S2 data derived from the long-term temporal window (six-year BSC) had the highest spectral agreement with laboratory spectra (mean SAM = 0.09), with ~50% of samples exhibiting SAM < 0.1. Yet, the coefficient of determination (R^2) values in the test datasets developed for a given temporal window varied with the target soil property. For example, for sand, R^2 increased from 0.22 to 0.33, RMSE decreased from 13.0 to 12.1, and RPIQ increased from 1.69 to 1.88; for clay, R^2 increased from 0.24 to 0.37, RMSE (root mean square error) decreased from 11.1 to 10.2, and RPIQ increased from 1.82 to 1.96; and for SOC, R^2 increased from 0.15 to 0.28, RMSE decreased from 0.13 to 0.12, and ratio of performance to interquartile range (RPIQ) increased from 1.56 to 1.72, respectively. Among the evaluated soil properties, Mn showed the highest predictive performance (R^2 = 0.76, RMSE = 8.86, RPIQ = 2.13). Incorporating FCM membership values as an additional predictor consistently improved model performance, increasing R^2 from 0.33 to 0.56, 0.37 to 0.53, and 0.61 to 0.72, and RPIQ from 1.88 to 2.20, 1.96 to 2.23, and 1.41 to 1.80 for sand, clay, and available Fe, respectively. Moderate improvements were also observed for pH and SOC. Clearly, longer temporal windows for BSCs and a soil heterogeneity measure provide an improved pathway for high-resolution soil assessment in heterogeneous landscapes.**

Index Terms **- Reflectance spectroscopy, temporal window, Fuzzy *c*-means, soil properties, Bare soil**

I. INTRODUCTION

Soil is a key component of the terrestrial ecosystem and is known to regulate agricultural productivity, water and nutrient cycling, and carbon sequestration among others [1], [2], [3], [4]. Hence, accurate and spatially-continuous information on soil properties is essential for sustainable land management, climate change mitigations, and precision agriculture [5], [6]. In semi-arid tropics, soils are particularly sensitive to climatic variability, repeated tillage and intensive cultivation [7]. For instance, variation in farm-level management in an agricultural landscape leads to substantial variability in surface soil properties even over short distances [8]. Although conventional soil characterization methods based on field sampling and laboratory analysis are accurate, capturing such landscape-scale soil heterogeneity becomes costly, time-consuming, and spatially-sparse [9], [10].

Several recent studies show that soil properties across extensive spatial extents may be estimated using multi-spectral imaging (MSI) data [11] ,[3], [12], [13]. Among current MSI missions, Sentinel-2 (S2) has emerged as a particularly valuable data source for soil-related studies [14] because of high spatial resolution (10–20 m) and short revisit intervals (5 days). Over the past decade, S2 MSI data have been used for estimating soil texture [15], [12], soil organic carbon (SOC) content [16], [17], [18], [19], pH [3], Fe content [3], soil moisture [20], cation exchange capacity [3] and even composite indices such as soil degradation index [6]. However, reliable retrieval of soil information from MSI data remains fundamentally constrained by vegetation cover [21]. Even low levels of green vegetation, senescent biomass, or crop residues significantly modify soil reflectance, reducing the interpretability of resulting spectral signatures [10].While unmixing remains a potential approach, unmixing algorithms are often inefficient to quantify the spectral reflectance of soil alone in a pixel containing multiple endmembers because of fewer bands in MSI data [22], [23]. Hence, most MSI-based soil estimation algorithms use the thresholding approach to identify bare soil fractions [24]. With increasing cropping intensities [25], finding sufficient bare soil fractions poses a challenge to implement such soil retrieval algorithms. In semi-arid agricultural systems, this challenge is exacerbated by short fallow periods, rapid crop rotations, and residue retention practices, which limit the temporal availability of fully exposed soil surfaces [26].

The authors acknowledge the support of the Groundwater Surveys and Development Agency (GSDA), Government of Maharashtra, under the Atal Bhujal Yojana for developing benchmark sites. The Multifunctional Landscape (MFL) Science Program of CGIAR is also acknowledged for partially supporting the time of the lead and corresponding authors.

[1]International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502324, Telangana, India
[2]Agricultural and Food Engineering Department, Indian Institute of Technology Kharagpur 721302, West Bengal, India
*Corresponding author. Email: kaushal.garg@icrisat.org

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

To overcome vegetation-induced spectral interference, the use of bare soil composites (BSCs) is increasingly adopted in soil remote sensing studies [17], [19], [27], [28], [29]. In this approach, soil reflectance data for all pixels in a given MSI scene are generated by stitching together exposed (bare) soil pixels available from MSI scenes collected at different acquisition dates. Typically, a threshold value of a characteristic spectral index is used to select each bare soil pixel for generating such BSCs [18], [30], [31], [32]. Construction of BSCs introduces an important methodological consideration such as what should be the length of the temporal window over which bare soil pixels are to be aggregated. The temporal window directly controls the balance between spatial coverage and spectral representativeness. Short temporal windows are more likely to be constrained by limited bare soil availability because of persistent vegetation cover or frequent cloud contamination [33], [34]. Conversely, long temporal windows substantially improve spatial coverage and data availability [19] but may introduce spectral inconsistencies resulting from temporal variability in soil moisture, tillage operations, organic amendments, residue management, and gradual land-use changes [35]. Moreover, most BSC-based studies are reported for soil organic carbon (SOC) contents and soil texture [19], [29],[36], while other soil properties have received comparatively little attention. Consequently, how different soil attributes respond to variations in temporal window length remains unclear, particularly in semi-arid tropical agroecosystems where soil surface conditions markedly vary over short time scale.

In addition to temporal effects, soil surfaces typically show gradual spatial transitions resulting from continuous variation in parent material, topography, and management practices [37]. Analytical approaches that can accommodate such continuity are, therefore, advantageous when interpreting soil spectral variability derived from satellite observations [18]. Clustering techniques such as fuzzy *c*-means (FCM)[38] provide a means of grouping spectrally-similar soils while allowing partial membership across clusters, thereby capturing transitional soil conditions without imposing rigid class boundaries [39], [40]. When applied to multi-temporal bare soil reflectance data, such approaches can assist in characterizing dominant patterns of soil spectral variability across heterogeneous agricultural landscapes [18] and may improve the prediction accuracy [41], [42]. Although previous studies have demonstrated the utility of Bare Soil Composites for digital soil mapping [17], [19], [27], [28], [29] and investigated the influence of compositing strategies or temporal windows for selected soil properties [19], [29], the extent to which temporal compositing window length influences the prediction of a broad range of soil properties remains poorly understood. Moreover, the FCM approach has been widely applied for clustering proximally sensed soil spectral dataset [43] and developing local models within each cluster [37]. However, it is unclear whether incorporating spectral heterogeneity information through FCM-derived membership values consistently improves prediction accuracy across different soil properties. Building on this literature, the present study makes two specific contributions. First, we extend the compositing-window comparison beyond single-property or small-subset studies to a panel of 15 soil properties spanning texture, macronutrients, secondary nutrients, and micronutrients, in the semi-arid tropical, basaltic setting of the Deccan Plateau, thereby evaluating whether the influence of temporal window length differs among soil properties with contrasting pedogenic behavior. Second, we compute continuous FCM membership directly from the same multi-temporal S2 BSC reflectance bands used as regression predictors and incorporate the resulting membership degree as an additional continuous covariate within a single unified regression model, rather than using it to stratify the dataset into local sub-models as in prior FCM-regression soil spectroscopy studies. Accordingly, the major objective of this study is to assess the effects of temporal window on the estimation accuracies of soil properties in the BSC-based remote sensing approach. The second objective is to evaluate whether integrating *fuzzy c*-means membership values with S2 spectral data improves soil property prediction.

II. MATERIAL AND METHODS

A. Study area

This study was conducted in the Indian state of Maharashtra with soil sampling sites distributed in five villages each selected from five different districts (Fig. 1). Geographically, the study area forms a part of the Deccan Plateau, a large and elevated region characterized by basaltic rock formations and gently undulating terrain [44]. The agricultural landscape is predominantly composed of smallholder farming systems, where farmers cultivate a range of crops depending on seasonal crop suitability and water access. During the monsoon season, major crops cultivated include black gram (*Vigna mungo*), pigeon pea (*Cajanus cajan*), cotton (*Gossypium spp.*) onion (*Allium cepa*), maize (*Zea mays*), and sorghum (*Sorghum bicolor*) while chickpea (*Cicer arietinum*) and wheat (*Triticum aestivum*) are typically grown in the post-monsoon season. In addition to field crops, farmers in the region also cultivate orchard crops such as mango (*Mangifera indica*), guava (*Psidium guajava*), and custard apple (*Annona squamosa*). According to the Köppen climate classification, the region falls under Group B, indicating a predominantly semi-arid climate. The area experiences hot summers (average maximum temperature: ~40.1 °C) and relatively mild winters (average minimum temperature: ~16.1 °C). Rainfall in the region is primarily influenced by the southwest monsoon (June–September) and, to a lesser extent, by the northeast monsoon (November–December). Annual precipitation ranges from 550 mm to 950 mm, which is distributed over roughly 20 rainy days per year. Around 40% of the total annual rainfall is concentrated over the month of September, often leading to short periods of water abundance followed by extended dry spells. Agricultural practices in the study area are highly dependent on monsoonal rains, with limited access to irrigation in many parts of the state.

B. Soil sampling and estimation of soil properties

Surface soil samples (depth: 0-15 cm) were collected from 443 ground reference locations (GRL) distributed over selected villages having diverse soil conditions and land-use practices typically seen across the region. Soil sampling campaign was conducted during the month of June 2023. Collected soil samples were used to analyse soil texture (sand, silt, and clay fractions), pH, electrical conductivity (EC), SOC content,

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

macronutrients, secondary nutrients, and micronutrients.

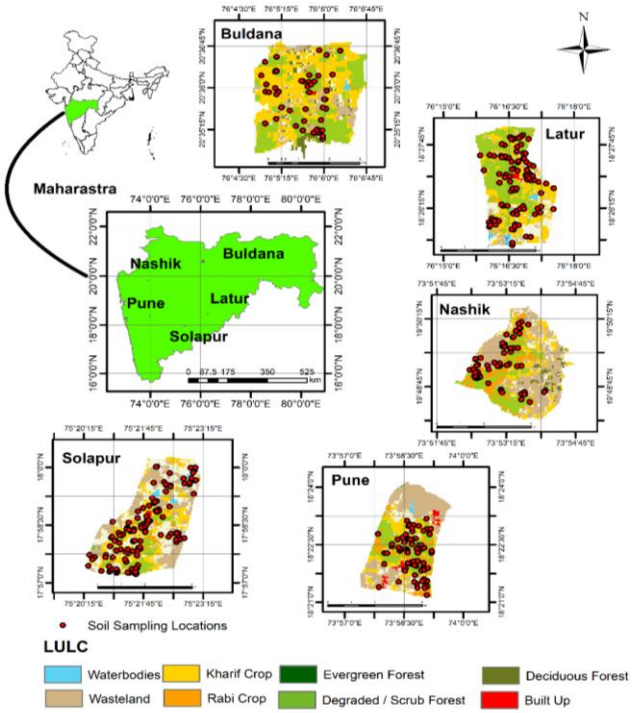


Fig.1. Map of the study area showing soil sampling locations in selected villages across five districts of Maharashtra

Briefly, soil texture was determined using the hydrometer method [45], SOC content was determined through the Walkley-Black method [46], and pH and EC were measured using soil-water slurry in the ratio of 1:2 and 1:2.5, respectively. Available P was estimated using the Olsen's and Bray's method using a continuous auto-analyser [47]. Exchangeable cations such as Ca, Mg, Na, and K were extracted using 1 N ammonium acetate [48] and the analytes were measured using a flame photometer (Systronics India Ltd., Ahmedabad, India). Selected micronutrients (Fe, Mn, Cu, and Zn) were extracted using the diethylenetriaminepentaacetic acid (DTPA) method, and their concentrations were quantified with an inductively coupled plasma (ICP) spectrometer (Model: HD prodigy; Teledyne Leeman Labs, USA). The hot-water extraction method was followed for available B [49].

C. Spectral data acquisition and pre-processing

Two sources of spectral data were used in this study. Spectral reflectance values of processed soil samples collected at each GRLs were measured under laboratory condition (hereinafter, referred to as lab spectra) for use as reference spectra for selecting appropriate S2 spectra (second spectral data source) for each BSC. A portable spectroradiometer (Model: Field spec®4 Hi-Res NG; Malvern Panalytical, UK) and a turntable fitted with halogen bulb as the light source were used for collecting lab spectra over the visible to near-infrared and shortwave-infrared (VNIR) region (350–2,500 nm) inside a dark room [50]. 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. Before collecting soil spectra, the spectroradiometer was warmed for one hour. A Spectralon® white reference panel (Lab sphere, USA) was then used to optimize the sensors and collect reference reflectance spectra. For collecting soil spectra, the petri dish containing levelled soil was placed on the turntable and thirty scans were averaged for obtaining each soil spectra. Spectroradiometer was optimized and calibrated after collecting spectra for every 30 samples. Smoothing of individual spectra was done by a third-order Savitzky-Golay smoothing method with the span length of 9 nm [51].

The S2 spectra corresponding to ten MSI bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12) were extracted using the Google Earth Engine (GEE) platform. These bands were selected because of their frequent use in estimating soil properties [19] while B1, B9, and B10 were excluded with their primary function in atmospheric correction and cloud detection. The GEE platform was used because surface reflectance products are easily extracted from L2 MSI data from S2 using an inbuilt atmospheric correction algorithm without the need for additional corrections [18], [52].

D. Construction of BSCs

For developing Bare Soil Composites (BSCs), S2 L2 images were first filtered based on scene-level cloud cover and further processed using the QA60 bitmask band to mask cloud- and cirrus-contaminated pixels. Bare soil pixels within each acquisition were then identified using thresholds of NDVI $\leq$ 0.25 and BSI $\geq$ 0.0. Three temporal windows were considered: short-term window of one year (Jan.-Dec., 2023), medium-term windows of three years (Jan., 2021 to Dec., 2023), and long-term window of six years (Jan., 2019 to Dec., 2024). Subsequently, all valid bare soil observations within each temporal window were composited at the pixel level using a median compositing algorithm, whereby the median reflectance value of each S2 MSI band was calculated across the temporal stack [19]. This approach reduced the influence of residual noise and outliers while preserving representative bare soil spectral characteristics and ensuring spatial continuity of the bare soil reflectance.

Spectral indices such as normalized difference vegetation index [53] and bare soil index [54] were derived [55] using the S2 bands:

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$BSI = \frac{(SWIR1 + Red) - (NIR + Blue)}{(SWIR1 + Red) + (NIR + Blue)} \quad (2)$$

The S2 band reflectance values from B2, B4, B8, and B11 cover the blue, red, near-infrared (NIR), and shortwave infrared one (SWIR1) regions, respectively.. Moreover, each S2 band was resampled to 10 m spatial resolution to achieve uniform spatial resolution.

The spectral angle mapper (SAM) was also used to compare the S2-based BSC spectra with their corresponding lab spectra measured for the processed soil samples of respective GRLs:

$$\alpha = cos^{-1}\left(\frac{\sum_{i=1}^{n} R_i^l . R_i^s}{\sqrt{\sum_{i=1}^{n}(R_i^l)^2}\sqrt{\sum_{i=1}^{n}(R_i^s)^2}}\right) \quad (3)$$

where α is the spectral angle in radians, n is number of spectral bands, R_i^l is the reflectance of the i^{th} band for the lab reference sample, and R_i^s is the reflectance of the i^{th} band for the S2 BSCs

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

sample. The SAM was used as an independent, pre-modelling diagnostic to quantify the spectral agreement between each temporal-window Bare Soil Composite (BSC) spectrum and the corresponding laboratory reference spectrum

Prior to SAM analysis, the laboratory reflectance spectra were spectrally resampled to S2 MSI bands by convolving the spectra with the published S2 relative spectral response functions (SRFs) [73] [74]. The S2-equivalent reflectance for each band was calculated as:

$$R_{s2,b} = \frac{\int_{350}^{2500} R(\lambda) SRF_b(\lambda)\, d\lambda}{\int_{350}^{2500} (\lambda) SRF_b(\lambda)\, d\lambda} \quad (4)$$

Where $R(\lambda)$ is the laboratory-measured reflectance at wavelength λ, $SRF_b(\lambda)$ is the relative spectral response function of Sentinel-2 band b and $R_{s2,b}$ the S2-equivalent reflectance for band b.

E. Fuzzy c-means clustering of bare soil spectra

To represent the continuous nature of soil variability and avoid imposing rigid class boundaries on spectrally-transitional soils, fuzzy *c*-means (FCM) clustering [38] was applied to the multi-band bare soil reflectance data derived from S2 MSI products. The FCM membership values were derived solely from the BSC spectral information, with no terrain, soil texture, or other ancillary spatial variables used in the clustering process. Unlike hard clustering approaches, FCM assigns partial membership of each pixel to multiple clusters, allowing gradual transitions between spectrally-similar soil surfaces to be preserved [56]. For each pixel i, a spectral feature vector was constructed as:

$$x_i = [R_{i,1}, R_{i,2}, \ldots\ldots R_{i,p}] \quad (5)$$

where $R_{i,p}$denotes surface reflectance in the p^{th} spectral band and p is the total number of spectral bands used in the analysis, including visible, red-edge, NIR, and SWIR bands. Prior to clustering, all spectral features were standardized to zero mean and unit variance to ensure equal contribution of all bands to the distance calculations. The FCM algorithm partitions the dataset into Kclusters by minimizing the following objective function [38]:

$$J_m = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik}^{m} \|x_i - c_k\|^2 \quad (6)$$

where N is the number of pixels, u_{ik} is the membership degree of pixel i in cluster k, m is the fuzziness exponent ($m > 1$), c_k is the centroid of cluster k, and $\|\cdot\|$ represents the Euclidean distance in spectral feature space. The fuzziness exponent controls the degree of overlap among clusters, with higher values producing more diffuse memberships [57]. Membership values were constrained such that:

$$\sum_{k=1}^{K} u_{ik} = 1 \text{ and } 0 \leq u_{ik} \leq 1 \; \forall \, i, k \quad (7)$$

ensuring that the total membership of each pixel across all clusters equals unity. The optimization of the objective function was achieved through an iterative alternating update of cluster centroids and membership values [38]. Cluster centroids were updated as:

$$c_k = \frac{\sum_{i=1}^{N} u_{ik}^{m} x_i}{\sum_{i=1}^{N} u_{ik}^{m}} \quad (8)$$

and membership values were recalculated using:

$$u_{ik} = \left\{ \sum_{j=1}^{K} \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}} \right\}^{-1} \quad (9)$$

The resulting fuzzy membership values provided a continuous measure of the spectral affinity of each pixel to the identified spectral clusters. A higher membership value indicated greater similarity between the pixel's BSC spectral signature and the corresponding cluster centroid, whereas comparable membership values across clusters indicated transitional spectral characteristics. Such fuzzy memberships have been shown to effectively capture environmental gradients and transitional landscape features in DSM and remote sensing applications [58], [59]. These continuous membership values were subsequently used as predictor variables in soil property modeling along with the S2 spectral variables, enabling the incorporation of sub-pixel soil variability and transitional soil conditions into the estimation framework.

F. Estimation of soil properties using S2 band reflectance data

Figure 2 shows a schematic of modeling framework followed for estimating soil properties using chemometric models where S2 band reflectance data derived from three selected temporal windows were used as predictor variables. The S2 spectra were extracted for each GRL and the resulting S2 spectral library was then used to evaluate different regression models for (a) identifying the optimum width of the temporal window and (b) evaluating the effects of incorporating the spectral information captured by the fuzzy membership functions in estimating different soil properties. Four regression models of support vector regression (SVR) [60], partial-least-square regression (PLSR) [61], random forest (RF) [62] and feature selection-based PLSR ($PLSR_{FS}$) [63] were examined to estimate surface soil properties. The details of these models are provided in Supplementary Methods. These models were evaluated using two modeling scenarios. In Scenario 1, a baseline model was developed using the S2 spectral library corresponding to each temporal window, with the 10 S2 band reflectance values used as predictor variables. In Scenario 2, FCM membership values derived from the same S2 spectral data were incorporated as auxiliary predictors along with the 10 S2 band reflectance values. Thus, each of the four chemometric models was developed and evaluated separately with and without FCM membership variables. The prediction performance obtained under the two scenarios was then compared for each soil property to quantify the contribution of FCM membership information to soil property estimation.

Before building chemometric models, each property was tested for normality using the Kolmogorov-Smirnov test at 5% significance level. When soil properties did not follow normal distribution, it was either log-transformed or transformed using the Box-Cox transformation [64]. Data partitioning into calibration (80%) and validation (20%) subsets were done using a stratified sampling approach [65]. The prediction performance of each model was assessed using the root-mean-squared error (RMSE) and the coefficient of determination (R^2):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i \right)^2} \quad (10)$$

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited, content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \quad (11)$$

where Y_i and $\hat{Y}_i$ are the measured and estimated soil parameter at the i^{th} location, respectively; $\bar{Y}$ is the mean value of Y_i and N is the number of locations.

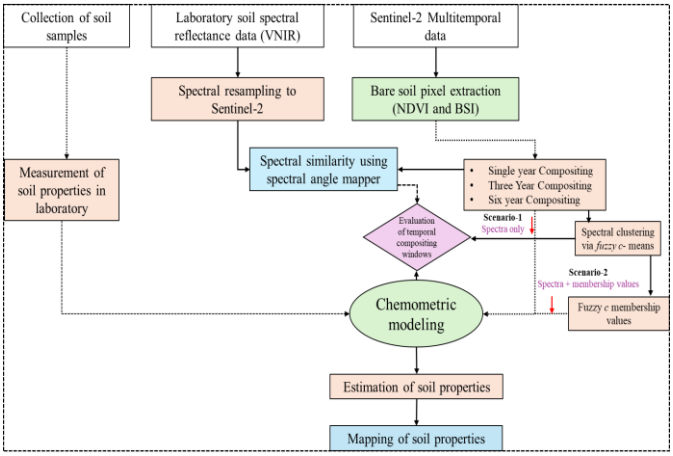


Fig. 2. Flowchart for estimating surface soil properties using spectral reflectance data extracted from bare soil composites (BSCs) based on multi-temporal Sentinel-2 imagery

III. RESULTS

A. Descriptive statistics and distribution of soil properties across study area

Figure 3 shows the frequency distribution of 15 different soil parameters considered in this study along with key descriptive statistics parameters. Each of these soil parameters show substantial soil variability, which is a characteristic feature of semi-arid tropical agroecosystem [66], [67]. With sand contents ranging from 9-86% and clay contents in the range of 3-67% (bi-modal distribution), soils of the study area show the dominance of both coarse-textured and fine-textured soils (Fig. 4).

Existence of a broad spectrum of soil textural classes in the study site may also be seen with a wide range of coefficient of variation (CV) values starting from 30% for sand to 51% for clay contents. Similar to soil clays, pH also shows a distinct bi-modal distribution (Fig. 3) with neutral to moderately alkaline (mean value of 7.97) soil reaction. In contrast, EC values range from 0.02–3.76 dS m^{-1} although most of the samples are within the non-saline range (< 2 dS m^{-1}). The SOC contents ranged from 0.06% to 1.25% with the mean value of 0.48% and low CV of 36% with only a few samples exceeding SOC contents of 1%. Macronutrient availability showed pronounced heterogeneity as farmers primarily apply these nutrients in varying amounts. For instance, available P contents ranged from 0.07 to 93.36 mg kg^{-1} (CV = 104%) reflecting strong variability. Similarly, exchangeable K ranged from 4.54 to 853.72 mg kg^{-1} and showed 93% CV. In contrast, exchangeable Ca and Mg contents showed narrow ranges and low variability (CVs of 23% and 25%, respectively), suggesting more uniform distribution controlled largely by underlying parent material and soil mineralogy. Micronutrients showed the highest degree of variability. Available S showed a wide range (2.77–438.46 mg kg^{-1}) with a CV of 143% and strong positive skewness. Available Zn also showed high CV value (= 174%) with positive skewness suggesting widespread Zn deficiency punctuated by a few highly enriched sites. Similar patterns were observed for available Fe, Mn, Cu, and B, all of which showed positively skewed distributions.

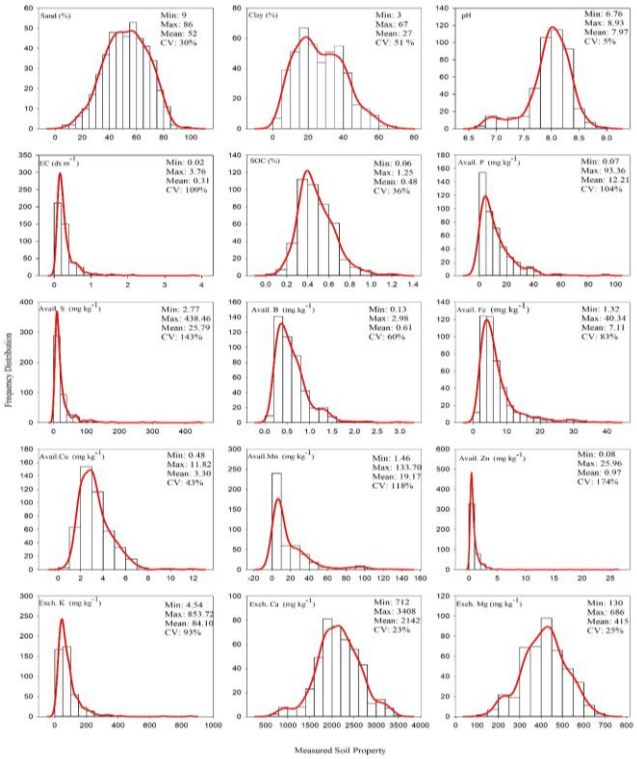


Fig. 3. Frequency distribution and kernel density of soil physical and chemical properties across the study area. EC: electrical conductivity; SOC (soil organic carbon); Avail.: available; Exch.: exchangeable

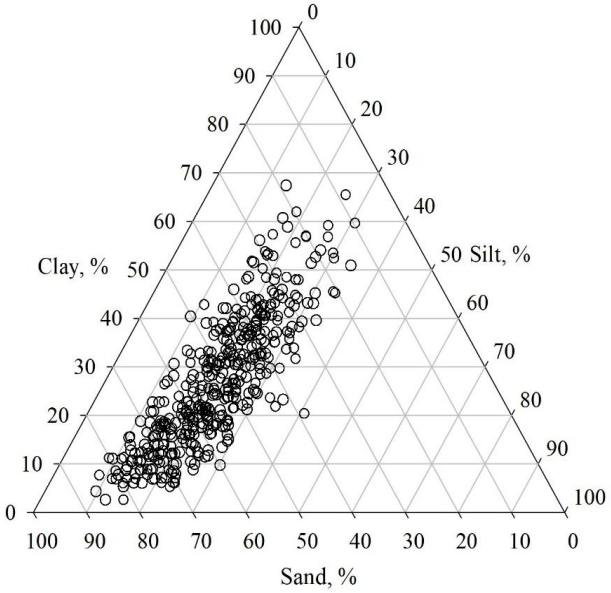


Fig.4. Textural triangle showing the distribution of soil textural classes in the study area

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

B. Selection of optimum temporal window for constructing BSC

Comparison of lab spectra (resampled to S2 MSI bands) with S2 MSI spectra using SAM values reveals a clear improvement in agreement between proximally- and remotely-sensed spectra when the width of temporal window increases during compositing (Fig. 5). The single-day observation shows the largest dispersion and highest mean SAM (0.14 radians), indicating substantial spectral variability likely driven by transient surface conditions. One-year and three-year composites show a modest reduction in both mean SAM (≈0.13 radians) and standard deviation. The six-year composite demonstrates the strongest agreement with lab spectra, characterized by the lowest mean SAM (0.09 radians), smallest standard deviation (0.04 radians), and the highest proportion of samples below the 0.1 radian threshold (50.1%) as shown in Figure 5. This progressive shift of the SAM distribution toward lower values indicates that longer temporal compositing enhances spectral stability and minimizes angular deviation from lab reference measurements. These results suggest that multi-year compositing substantially improves the consistency between S2-derived and lab spectra.

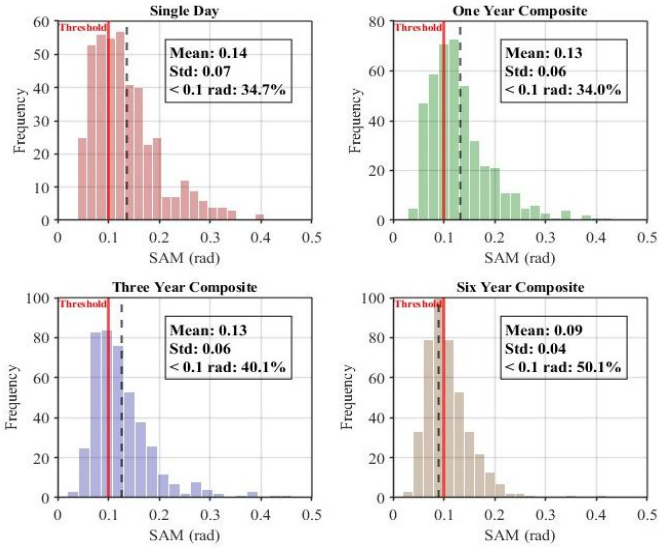


Fig. 5. Comparison of spectral angle mapper (SAM) values of spectral reflectance curves for soil samples measured in the laboratory (resampled to 10 Sentinel-2 bands) and single-year, three-year, and six-year bare soil composites (BSCs). Dashed and solid vertical lines denote the mean SAM and the 0.1 radian threshold, respectively

C. Estimation of soil properties and influence of temporal window length

Table 1 shows the comparative predictive performance of different temporal windows. Predictive performance varied across temporal aggregation windows, with the response differing among soil properties. Rather than showing a uniform trend, the results indicate that the influence of temporal aggregation depends on the property being estimated. The reliability of temporal-window effects varied among soil properties. R^2, RMSE, and RPIQ were therefore considered jointly when evaluating changes in predictive performance across the three temporal windows. Several properties, including sand, clay, SOC, pH, EC, available B, Fe, Cu, Mn, Ca, and Na, showed relatively consistent patterns across these metrics, allowing temporal-window responses to be evaluated. In contrast, available P, available S, available Zn, exchangeable K, and exchangeable Mg showed relatively weak and/or inconsistent predictive performance, with limited changes in RPIQ across the temporal windows. Accordingly, the temporal responses of these properties were interpreted cautiously, while the main discussion of temporal aggregation focused on properties showing comparatively stronger and more consistent predictive performance. Predictive performance varied across temporal aggregation windows, with responses differing according to the dominant soil processes governing each property. Rather than showing a uniform trend, the results show that the optimum length of temporal window may change with the parameter estimated in the MSI approach. For inherent properties such as soil texture, longer temporal aggregation consistently improved model performance. Sand content showed an increase in R^2 from 0.22 in the single-year composite to 0.25 and 0.33 in the three- and six-year composites, respectively, accompanied by a reduction in RMSE from 13.0 to 12.1 and an increase in RPIQ from 1.69 to 1.88. Clay content showed a similar pattern, with R^2 increasing from 0.24 to 0.37, RMSE decreasing from 11.1 to 10.2, and RPIQ increasing from 1.82 to 1.96 with increasing temporal aggregation. These improvements suggest that particle-size related properties benefit from long-term spectral stabilization, as their reflectance characteristics are relatively stable over time and less affected by short-term surface disturbances. SOC also showed improved predictive performance with increasing temporal aggregation, with R^2 increasing from 0.15 to 0.28, RMSE decreasing from 0.13 to 0.12%, and RPIQ increasing from 1.56 to 1.72. Although the predictive accuracy was lower than that reported by Zhu et al. [19], Shi et al. [34], and Silvero et al. [33], who achieved R^2 values of 0.64, 0.62, and 0.38, respectively, using multi-temporal S2 MSI BSCs, our results show a consistent improvement in SOC prediction with increasing temporal window length. This trend aligns with previous findings demonstrating that multi-temporal compositing enhances spectral stability and improves SOC estimation compared to single-date imagery [19]. In contrast, pH showed its highest predictive performance at the intermediate three-year window. R^2 increased from 0.34 for the single-year composite to 0.51 for the three-year composite before declining to 0.43 for the six-year composite. The corresponding RMSE values were 0.26, 0.24, and 0.25, while RPIQ increased from 1.38 to 1.48 at the three-year window before returning to 1.38. These results exceed the range reported by Vaudour et al. [3], who observed R^2 values between 0.08 and 0.51 for pH prediction using S2 MSI data. Soil EC followed a similar pattern, with R^2 increasing from 0.21 to 0.35 at the three-year composite and remaining similar at 0.33 for the six-year composite. RMSE decreased from 0.16 to 0.15, while RPIQ increased from 1.02 to 1.15 and 1.18 for the three- and six-year composites, respectively. Several nutrient-related properties benefited from spectral data aggregation over longer temporal windows. For example, Available Fe showed strong predictive performance across all three temporal windows, with R^2 increasing from 0.56 for the single-year composite to 0.66 for the three-year composite before declining slightly to 0.61

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

for the six-year composite. RMSE followed the same pattern, decreasing from 3.02 to 2.52 before increasing to 2.74, while RPIQ increased from 1.39 to 1.65 before declining to 1.41. Thus, the three-year composite provided the strongest overall predictive performance for available Fe. This value exceeds the upper range reported by Vaudour et al. [3] for Fe (R^2 = 0.05–0.45), highlighting the benefit of extended temporal compositing for properties strongly linked to iron oxide spectral features.

Table I
Validation performance metrics including the coefficient of determination (R^2), root mean square Error (RMSE), and ratio of performance to interquartile range (RPIQ) for soil selected soil properties

Composite type	Single-year composite			Three-year composite			Six-year composite		
Property	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ
Sand, %	0.22	13.0	1.69	0.25	12.5	1.77	0.33	12.1	1.88
Clay, %	0.24	11.1	1.82	0.26	11.05	1.86	0.37	10.2	1.96
pH	0.34	0.26	1.38	0.51	0.24	1.48	0.43	0.25	1.38
EC, dS m^{-1}	0.21	0.16	1.02	0.35	0.15	1.15	0.33	0.15	1.18
SOC, %	0.15	0.13	1.56	0.17	0.12	1.64	0.28	0.12	1.72
Avail. P	0.08	9.60	1.27	0.13	9.40	1.31	0.08	9.89	1.18
Avail. S	0.22	18.81	0.85	0.22	17.81	0.90	0.34	17.10	0.95
Avail. B	0.30	0.25	1.54	0.39	0.22	1.70	0.39	0.23	1.65
Avail. Fe	0.56	3.02	1.39	0.66	2.52	1.65	0.61	2.74	1.41
Avail. Cu	0.28	1.04	1.51	0.32	0.96	1.62	0.44	0.88	1.80
Avail. Mn	0.75	8.24	2.26	0.68	8.50	2.03	0.76	8.86	2.13
Avail. Zn	0.16	0.51	1.12	0.09	0.52	1.12	0.30	0.51	1.17
Exch. K	0.12	46.96	1.15	0.04	43.48	1.20	0.32	40.42	1.30
Exch. Ca	0.23	353.50	1.65	0.20	336.74	1.67	0.29	330.74	1.76
Exch. Mg	0.12	80.53	1.48	0.15	79.41	1.50	0.18	78.51	1.56
Exch. Na	0.19	34.54	1.32	0.46	32.05	1.47	0.45	30.22	1.51

EC: electrical conductivity; SOC: soil organic carbon; Avail.: available; Exch.: exchangeable; all nutrient contents are in mg kg^{-1}

These responses suggest that aggregation of spectral data over Available P showed consistently low predictive performance across the three temporal windows, with R^2 ranging from 0.08 to 0.13 and RPIQ from 1.18 to 1.31. Although R^2 increased slightly for the three-year composite, it declined again for the six-year composite, while RMSE remained relatively high (9.40–9.89). These results indicate a weak and inconsistent temporal response for available P. Available S showed an increase in R^2 from 0.22 for the single- and three-year composites to 0.34 for the six-year composite, accompanied by a reduction in RMSE from 18.81 to 17.10. However, RPIQ remained below 1 across all three temporal windows (0.85–0.95), indicating limited predictive performance. Therefore, the increase in R^2 at the six-year window was not considered sufficient to indicate a clear improvement in temporal-window suitability. Available B also improved progressively with wider temporal windows. These responses suggest that aggregation of spectral data over wide data acquisition windows may better capture broader soil chemical status, which evolves over longer timescales as a part of pedogenesis. Micronutrients such as available Mn demonstrated consistently high predictability (R^2 = 0.75, 0.68, and 0.76 and RPIQ values of 2.26, 2.03, and 2.13 for single-, three-, and six-year composites, respectively), indicating strong indirect spectral relationships, likely mediated by iron oxide-related spectral features in the VIS–SWIR region. Exchangeable bases generally benefited from compositing over wide data acquisition windows. Exchangeable K improved substantially from R^2 = 0.12 to 0.32 with a marked reduction in RMSE from 46.96 to 40.42 and increase in RPIQ from 1.15 to 1.30, while Ca, Mg, and Na showed moderate but consistent gains under extended temporal window. Exchangeable Na increased sharply from R^2 = 0.19 to 0.46, RMSE decreased from 34.54 to 30.22, while RPIQ increased from 1.32 to 1.51 under the three-year composite and maintained similar performance at six-year windows, highlighting the effectiveness of multi-temporal stabilization for salinity-related properties. The observed patterns suggest that temporal compositing may be beneficial for some soil properties, whereas others show limited or inconsistent responses to changes in the temporal window. Physical and mineral-controlled soil parameters continued to benefit from longer aggregation, whereas dynamic soil parameters peaked at intermediate windows. These results indicate a process-dependent trade-off between spectral stabilization and temporal window length, which does not uniformly translate into improved soil property estimation, and that temporal compositing strategies should be tailored to the specific spectral behaviour of the target variable, with appropriate caution for properties where absolute prediction accuracy remains low.

D. Fuzzy c-means clustering of soil spectra

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited; content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

The FCM method was applied to both lab spectra resampled to S2 and the multi-temporal S2 BSC spectra, with membership distribution plotted in ternary space (Fig. 6a). For the laboratory spectra (Fig. 6a), membership values were distributed across all three pairwise edges of the ternary plot, with few samples occupying the corners indicating that the soil variability exists as a continuous spectral continuum rather than discrete classes. When transitioning to satellite-derived spectra, distinct and systematic effects of temporal window length on FCM membership structure emerged. The single-year BSC (Fig. 6b) produced a sparser and more scattered membership distribution compared to the laboratory reference, possibly because of the transient surface noise (e.g., crop residues, soil moisture variability). The three-year BSC (Fig. 6c) showed a pronounced concentration of membership values along the Cluster 1–Cluster 3 edge, with minimal contribution from Cluster 2, suggesting a reduced expression of spectral diversity in the composite. In contrast, the six-year BSC (Fig. 6d) recovered a membership distribution spanning all three ternary edges, most closely resembling the laboratory reference among all satellite-derived composites. The majority of samples occupied transition zones between cluster pairs, with fewer samples exhibiting pure affiliation with a single class. The convergence of the six-year BSC membership structure toward the laboratory reference, therefore, explains its superior prediction performance across the broadest range of soil properties.

The resulting clusters show consistent differences in their mean spectral signatures across the entire spectral range using the six-year BSC (Fig. 6e). The identified spectral clusters corresponded with measured soil physicochemical properties, confirming that the grouping reflects meaningful pedological gradients rather than arbitrary partitions. Cluster 2 is characterized by the highest sand content (58%) and relatively lower clay content (22%), along with lower concentrations of available Fe (5.97 mg kg^{-1}) and available Mn (12.98 mg kg^{-1}). These combined physical and chemical attributes are consistent with the comparatively higher spectral reflectance observed for this group across S2 MSI bands (Fig. 6b). In contrast, Cluster 3 represents finer-textured soils, with 35% clay. This cluster also shows higher concentrations of available Fe (9.73 mg kg^{-1}) and available Mn (37.79 mg kg^{-1}), elements commonly occurring as oxide coatings that contribute to absorption processes in the visible region through electronic transition mechanisms. Cluster 1 occupies an intermediate position in textural composition (53% sand; 27% clay), along with moderate concentrations of available iron (6.55 mg kg^{-1}) and manganese (15.35 mg kg^{-1}). The multi-temporal S2 MSI composites therefore capture soil variability structured not only by particle size distribution but also by micronutrient-associated mineral phases and cationic properties that influence surface reflectance behaviour (Nocita et al., 2015). The resulting soil groups therefore represent physically meaningful entities structured by soil properties rather than arbitrary statistical divisions.

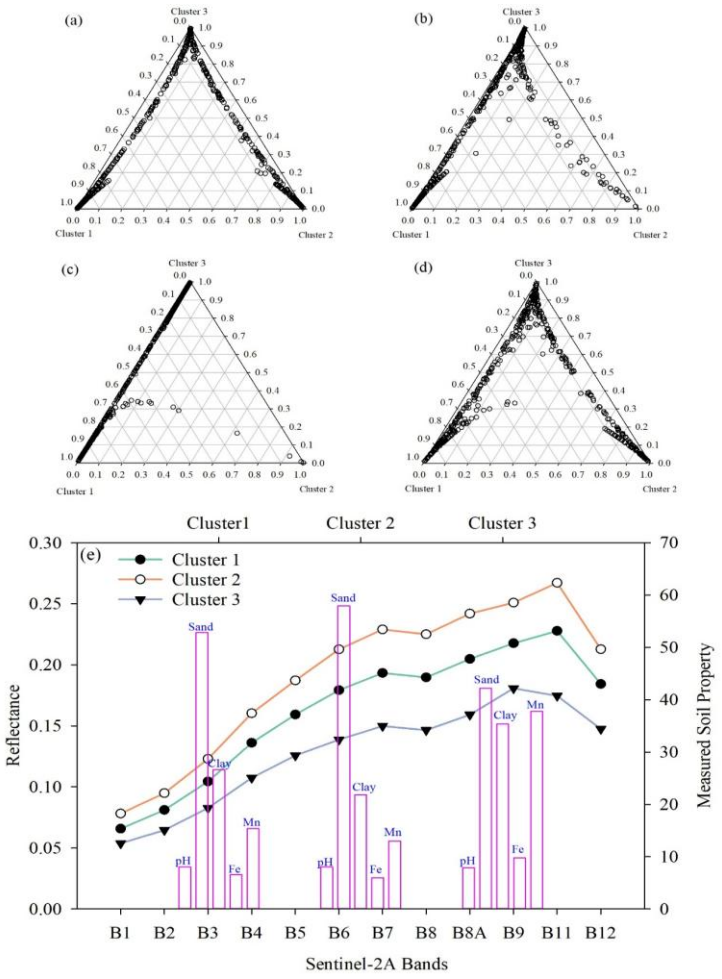


Fig. 6. Ternary plot of fuzzy *c*-means (FCM) membership values derived from Sentinel-2 (S2) bare soil composites (BSCs). (a) Laboratory spectra; (b) Single year BSCs; (c) Three-year BSCs; (d) Six-year BSCs (e) Mean spectra of the three identified FCM clusters alongside their respective soil properties. Sand and clay contents are expressed in percentage; micronutrients are expressed in mg kg^{-1}

E. Impact of Fuzzy c means spectral derived membership on soil property prediction

Figure 7 presents the comparative performance of spectra only models (scenario 1; Fig. 2) with those incorporating FCM membership values (scenario 2; Fig. 2) across selected soil properties. Fig. 8 shows the measured versus predicted values for six representative soil properties under Scenario 1 and Scenario 2 using the six-year composite, illustrating the effect of incorporating FCM membership variables on prediction agreement. Results show that the incorporation of FCM membership values generally enhances predictive performance across the majority of variables. Improvements through clustering have been reported in laboratory reflectance-based soil studies [43], [68]. For several properties, particularly those influenced by mineralogical composition, incorporating FCM membership values as predictors along with S2 MSI spectra resulted in noticeable improvement in R^2 values.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited, content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

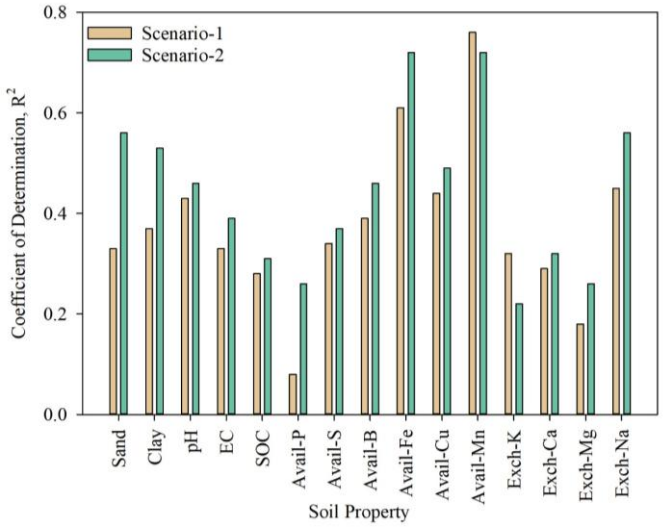


Fig. 7. Comparative coefficients of determination (R^2) for selected soil properties between scenario-1 (spectra only) and scenario-2 (spectra+ FCM membership values) modeling. Sand and clay contents are in percentage; FCM: fuzzy *c*-means; EC: electrical conductivity; SOC: soil organic carbon; Avail.: available soil nutrient contents (mg kg^{-1}); Exch.: exchangeable soil nutrient contents (mg kg^{-1})

For instance, the prediction accuracy for sand contents improved from R^2 = 0.33 in scenario 1 to 0.56 in scenario 2 (FCM membership values used as additional predictor), while R^2 values clay contents increased from 0.37 to 0.53. Micronutrients also benefited with incorporating FCM membership values, with R^2 for available Fe increasing from 0.61 to 0.72, suggesting that FCM membership-guided modeling better captures iron- and manganese-oxide related spectral features in the visible and red-edge regions. The improvement aligns with studies demonstrating that iron-bearing minerals strongly influence soil reflectance and that spectral grouping prior to modeling can enhance their prediction [18], [69]. Exchangeable Na and Mg similarly showed notable increases (R^2 improving from 0.45 to 0.56 for Na and from 0.18 to 0.26 for Mg), indicating that salinity and cation exchange-related variability is more effectively represented when spectral heterogeneity is accounted for. Moderate but consistent improvements were observed for pH, EC, SOC, and available S. Th R^2 values increased from 0.43 to 0.46 for pH, from 0.33 to 0.39 for EC, from 0.28 to 0.31 for SOC, and from 0.34 to 0.37 for available S, suggesting that FCM membership-based modelling enhances the predictability of spectrally inactive soil properties. Although a few properties showed limited change or marginal decline, the overall increasing trend in R^2 values indicates that the incorporation of FCM membership values enhances model performance by accounting for landscape-level spectral heterogeneity. These findings support the premise that soil spectral relationships are property dependent and that integrating unsupervised spectral classification (membership values) prior to regression modeling can improve predictive accuracy [42] in characteristically heterogeneous semi-arid agricultural systems.

The Taylor diagrams (Fig. 9) illustrate the predictive performance of PLSR, RF, SVR, and $PLSR_{FS}$ models across 15 soil properties. The RF, SVR and $PLSR_{FS}$ algorithms consistently achieved closer proximity to the observed reference point compared to standard PLSR.

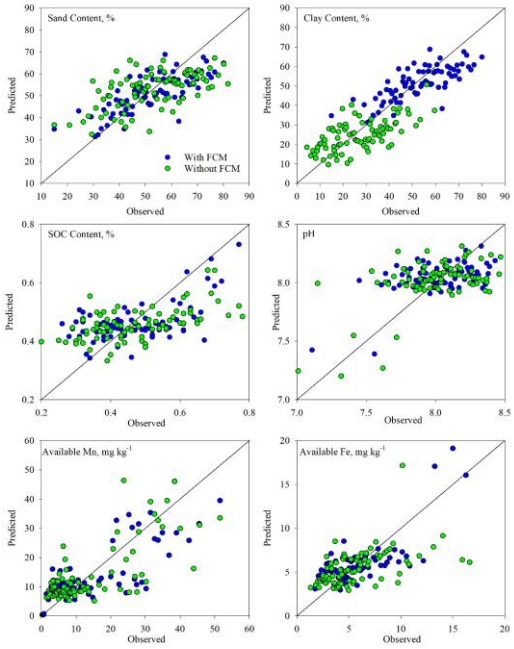


Fig. 8. Measured versus predicted values for six representative soil properties, comparing Scenario 1 (spectra only, green) and Scenario 2 (spectra + FCM membership, blue) predictions against the 1:1 line.

This strategic feature selection approach effectively minimizes centered RMSE and improves the alignment of predicted variances with ground-truth observations, even when the overall correlation remains low by the inherent signal-to-noise ratio of the input data. This pattern is further supported by the model-wise comparison presented in Tables S1–S4 of the Supplementary Material. The inclusion of FCM membership variables generally increased R^2 and reduced RMSE across the four modelling approaches, although the magnitude and direction of the changes varied among soil properties. Under Scenario 2, $PLSR_{FS}$ achieved the highest predictive performance for 7 of the 16 soil properties, followed by RF for 5 properties, while the remaining properties were best predicted by PLSR or SVR. For example, SVR achieved the highest R^2 for sand (0.56) and exchangeable Na (0.56), whereas RF achieved the highest R^2 for available Fe (0.72). Thus, the model showing the greatest benefit from FCM integration was not necessarily the model providing the highest overall predictive performance, indicating that the effect of FCM membership variables was both property- and model-dependent.

To further examine the contribution of FCM membership beyond its effect on overall prediction accuracy, random forest permutation importance was evaluated for the S2 spectral bands and FCM membership variables (Fig. 10). The importance of rankings varied among soil properties. For selected soil properties, FCM membership variables ranked among the most influential predictors, whereas for others their contribution was comparatively small relative to individual S2 bands. These results indicate that FCM membership variables provide additional predictive information for selected soil

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

properties, but their contribution is not uniform across properties or modelling approaches.

Figure 11 shows the spatial distribution of key soil properties such as pH, SOC, sand content, and clay content within a representative clipped portion of the study area. The selected subset highlights localized spatial variability in soil texture and associated soil chemical properties. Within the clipped area, clear contrasts between sand dominated and clay-dominated zones are visible. Regions with relatively higher sand content correspond closely with areas of lower clay content, reflecting the inverse relationship between coarse and fine soil fractions. The predicted soil pH within the clipped area ranges approximately from 7.14 to 8.25, indicating predominantly neutral to moderately alkaline conditions as indicated by the ground truth data (Fig. 3). Areas with relatively higher clay content tend to show slightly higher pH values, while regions dominated by sandy soils display comparatively lower pH [70]. This localized pattern reflects the buffering capacity of clay-rich soils, which retain base cations more effectively than sandy soils.

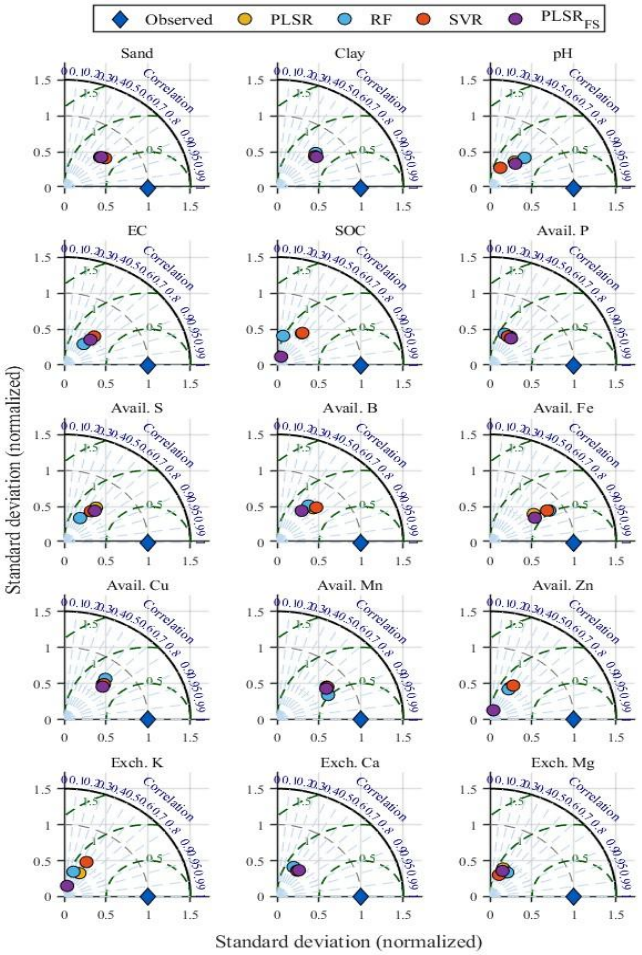


Fig. 9. Taylor diagrams comparing the performance of chemometric models for predicting soil properties using Sentinel 2 (S2) data. Sand, clay and SOC content are expressed in %, while macro and micronutrients are in mg kg^{-1}

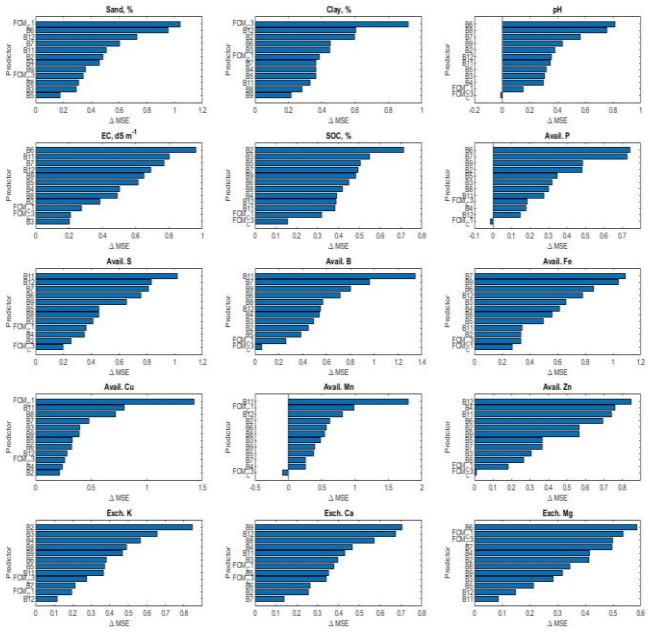


Fig. 10. Random Forest variable importance (OOB permuted ΔMSE) for the six-year bare soil composite, shown separately for each of the 15 soil properties. Predictors comprise 10 Sentinel-2 bands (B2–B9, B11, B12) and two fuzzy c-means (FCM) membership values (Cluster 1, Cluster 3)

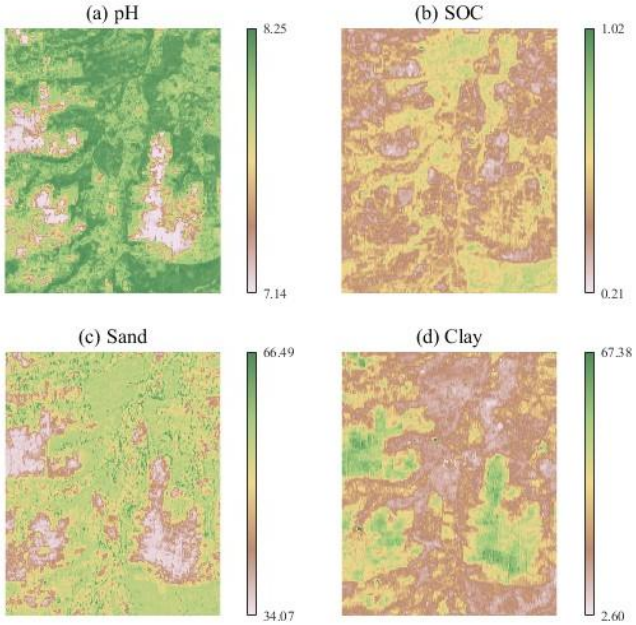


Fig.11. Spatial prediction of selected soil properties using surface Sentinel 2 (S2) spectra. Soil organic carbon (SOC), sand and clay content are expressed in percentage

IV. DISCUSSION

A. Influence of temporal window length on spectral stability and soil property estimations

Temporal window exerted a systematic yet process-dependent influence on both spectral similarity and predictive performance. While increasing temporal window consistently

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

improved spectral similarity toward lab spectra (as indicated by decreasing SAM values), its effect on soil property prediction was not straightforward. In line with these findings, Zhu et al. [19] demonstrated that multitemporal compositing reduces the discrepancy between S2 MSI data and laboratory-measured soil spectra. The temporal dynamics observed in this study had direct influence on soil property modeling. Predictive performance did not follow a uniform trend across all properties; rather, responses were process dependent. The single-year BSCs produced comparatively low R^2 values across most properties (e.g., R^2 of 0.22 for sand content, 0.24 for clay contents, and 0.15 for SOC contents), reflecting the limited number of bare-soil observations available within a 12-month period. Under such short windows, composite reflectance remains sensitive to transient surface conditions, including soil moisture variability, crop residue remnants, and recent tillage [71]. Extending the window to three years substantially improved prediction for several chemically dynamic properties, including pH ($R^2 = 0.51$), EC ($R^2 = 0.35$), exchangeable Na ($R^2 = 0.46$), and available Fe ($R^2 = 0.66$). The three-year composite appears to represent a balance point at which short-term spectral noise is sufficiently suppressed while soil chemical variability linked to recent management and hydrological regimes remains preserved. For properties influenced by equilibrium reactions, and salinity dynamics, this intermediate window captures representative soil conditions without excessive temporal mixing. In contrast, inherent attributes such as sand and clay contents continued to benefit from longer aggregation, reaching R^2 values of 0.33 and 0.37, respectively, at six years. These properties are governed primarily by mineralogical composition and pedogenic processes, which vary slowly over time and are less sensitive to management-driven fluctuations. For such attributes, spectral data aggregation over extended temporal window enhances spectral stabilization without eroding meaningful variability. Micronutrients such as available Fe and Mn maintained relatively high predictability across windows, reflecting strong indirect spectral coupling through mineralogical and moisture-related features. Results also indicate that temporal window selection is a fundamental determinant of model success, but its optimal length depends on the dominant soil processes governing each property.

Previous studies have demonstrated the usefulness of multitemporal BSCs for soil-property mapping, particularly for SOC and topsoil characteristics [19], [30], while Heiden et al. [31] examined approaches for improving the consistency and performance of soil reflectance composites. These studies highlight the advantage of temporal compositing in increasing the availability of bare-soil observations. However, most existing studies have focused on individual soil properties or on developing improved compositing approaches rather than systematically assessing the effect of different temporal window lengths on a broad range of soil properties. In contrast, the present study evaluates three explicitly defined temporal windows (one-, three-, and six-year) across 15 soil properties, demonstrating that the optimal temporal window varies among soil properties. Thus, the present results extend previous BSC and temporal compositing studies by showing that temporal-window optimization should be considered property-dependent rather than assuming a single optimal compositing period.

The finding that BSC SAM improves monotonically with compositing window length, while prediction accuracy for dynamic soil properties peaks at an intermediate window, reflects two distinct timescales operating simultaneously. The contrasting trends between spectral similarity and prediction accuracy suggest that temporal compositing affects spectral consistency and soil-property estimation through different mechanisms. The continuous improvement in SAM with increasing temporal window indicates that longer compositing periods enhance the stability of the BSC by increasing the number of bare-soil observations and reducing transient spectral variability. SAM reflects the internal consistency of the composite itself, which improves as more bare-soil observations are aggregated and transient surface noise is averaged out. Prediction accuracy, by contrast, additionally depends on whether the compositing window is well-matched to the timescale over which the target property varies. For relatively stable properties such as texture, which are not expected to change substantially over the one- to six-year period considered in this study, the improved spectral stability of longer-window composites can contribute to better prediction. For dynamic, management- and moisture-sensitive properties, such as EC, available S, and exchangeable Mg, however, the property's true state itself changes appreciably within a multi-year compositing window. Long-term compositing therefore smooths over genuine temporal variability in these properties that is correlated with their measured spatial pattern at the time of soil sampling, degrading the spectra-property relationship even as generic spectral consistency continues to improve. The three-year window may represent a compromise for several of these properties, long enough to suppress short-term compositing noise, but not so long as to average over substantial true temporal change in the target property itself. These results suggest that the selection of an appropriate compositing window should consider not only spectral similarity but also the temporal behavior of the target soil property.

B. Advantages of FCM membership for soil property prediction

A key methodological implication of this study is the advantage of FCM based approach over categorical stratification schemes such as land use or soil type using S2 MSI data. Spectral clusters capture continuous gradients in soil surface conditions and allow overlapping memberships, which better reflect real-world soil variability [40]. This is particularly important in semi-arid agroecosystems, where management practices and topsoil properties vary gradually rather than abruptly [72]. The results showed the advantage of FCM with properties showing substantial increase in prediction performance, except for few showing limited or marginal declines. By using FCM membership values as predictors, the modeling framework incorporates sub-pixel heterogeneity and transitional soils, improving the robustness of soil property predictions [68]. This

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

approach also avoids the need for externally defined stratification layers, making it more transferable across regions where ancillary data may be limited.

The findings of this study have important implications for operational soil property mapping using MSI satellite data. Specifically, careful temporal window selection is essential for balancing spatial coverage and spectral integrity. Second, results also show that FCM membership-based representations of soil variability can substantially enhance the predictive power of spectral models. The six-year bare soil composite, combined with FCM membership values, offers a scalable and computationally feasible framework for regional soil assessment. This approach is particularly well suited for semi-arid regions, where fully bare soil conditions occur intermittently and soil properties show strong spectral heterogeneity.

C. Limitation of the study

First, temporal window length is partly confounded with the number of valid bare-soil observations available for compositing. Longer windows generally contain more cloud-free acquisitions and may therefore produce more robust composites through increased observation density, in addition to providing a longer temporal representation of soil spectral conditions. Consequently, the improvements observed from single- to multi-year BSCs cannot be attributed exclusively to temporal span. In the present analysis, these two effects could not be fully separated because the composites were generated from the available observations within each temporal window. Future studies should use controlled subsampling of observations, moving-window analyses, or time-series decomposition to disentangle the effects of temporal duration, observation density, and seasonal representativeness.

Second, no explicit soil-moisture normalization was applied during the generation of the BSCs. Because the six-year temporal window combines bare-soil observations acquired under different antecedent rainfall and surface moisture conditions, the resulting median composite may contain residual moisture-related spectral variability that is not fully addressed by NDVI- and BSI-based bare-soil masking. Nevertheless, the six-year BSC exhibited the highest spectral agreement with the laboratory spectra (mean SAM = 0.09 rad; Fig. 5), suggesting that the increased availability of bare-soil observations and improved spatial representation associated with the longer temporal window may have compensated for some of the additional spectral variability introduced by differing soil moisture conditions. However, residual moisture effects cannot be ruled out and may represent an additional source of spectral variability. Future studies should investigate soil-moisture-aware compositing strategies, such as restricting image selection to dry antecedent conditions or incorporating soil-moisture indicators as additional covariates, to further reduce the influence of moisture-related spectral variability.

Third, the present study was conducted across five districts of the Deccan Plateau of Maharashtra, representing a basaltic, semi-arid agroecosystem with specific soil, climatic, and cropping conditions. Therefore, the numerical findings obtained in this study, including the optimal temporal compositing windows, FCM cluster structure, and prediction performance, should be interpreted within the context of this pedo-climatic setting rather than considered universally applicable. Regions with different parent materials, soil types, cropping systems, climatic conditions, or bare-soil availability, including other semi-arid regions with different rainfall seasonality or vegetation cover dynamics, may exhibit different optimal compositing windows and FCM cluster structures. Nevertheless, the proposed framework, which combines multi-year bare-soil compositing, spectral evaluation, FCM-based spectral representation, and soil-property modelling, provides a general methodological approach that can be applied to other agricultural regions. Further evaluation across contrasting soil types, parent materials, climatic zones, and cropping systems is required to establish the robustness and broader transferability of the framework

V. Conclusion

This study demonstrates that both temporal compositing and spectral context play critical roles in determining the accuracy of soil property prediction from S2-derived BSCs. Longer Temporal window improved spectral stability and agreement with laboratory measurements, with the six-year composite showing the closest correspondence to lab spectra (mean SAM = 0.09). However, predictive performance did not automatically increase with wider temporal window. While texture-related properties improved with longer windows (e.g., R^2 of 0.22 to 0.33, RMSE decreased from 13.0 to 12.1, and RPIQ increased from 1.69 to 1.88 for sand content), dynamic soil properties showed optimal performance at intermediate scales, with soil pH increasing from R^2 = 0.34 to 0.51 and RMSE decreasing from 0.26 to 0.24 under the three-year composite before declining at six years. The incorporation of FCM membership values as auxiliary predictors significantly enhanced model performance across most soil properties. Improvements were particularly pronounced for texture (sand and clay) and micronutrients (Fe), as well as exchangeable cations such as Na and Mg. Sand prediction improved from R^2 = 0.33 to 0.56 and clay from 0.37 to 0.53, while Fe increased from 0.61 to 0.72 and exchangeable Na from 0.45 to 0.56. Unlike conventional approaches that utilize clustering for data partitioning or local modeling, the integration of fuzzy membership within a unified modeling framework enables continuous representation of soil spectral variability across heterogenous landscapes.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

[1] A. R. Dexter, "Soil physical quality," *Geoderma*, vol. 120, no. 3–4, pp. 201–214, Jun. 2004, doi: 10.1016/j.geoderma.2003.09.004.

[2] R. Lal, "Soil health and carbon management," *Food Energy Secur.*, vol. 5, no. 4, pp. 212–222, Nov. 2016, doi: 10.1002/fes3.96.

[3] E. Vaudour, C. Gomez, Y. Fouad, and P. Lagacherie, "Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems," *Remote Sens. Environ.*, vol. 223, pp. 21–33, Mar. 2019, doi: 10.1016/j.rse.2019.01.006.

[4] Y. Hao *et al.*, "Soil moisture controls over carbon sequestration and greenhouse gas emissions: a review," *Npj Clim. Atmospheric Sci.*, vol. 8, no. 1, p. 16, Jan. 2025, doi: 10.1038/s41612-024-00888-8.

[5] W. Kabato, N. Hailegnaw, L. Mutum, and Z. Molnar, "Managing soil health for climate resilience and crop productivity in a changing environment," *Sci. Total Environ.*, vol. 1000, p. 180460, Oct. 2025, doi: 10.1016/j.scitotenv.2025.180460.

[6] I. Majeed *et al.*, "Assessing Soil Degradation in Agricultural Landscapes of Semi-Arid Tropics Using Proximal and Remote Sensing-Based Diffuse Reflectance Spectroscopy," *Soil Use Manag.*, vol. 41, no. 2, p. e70101, Apr. 2025, doi: 10.1111/sum.70101.

[7] A. Bationo, J. Kihara, B. Vanlauwe, B. Waswa, and J. Kimetu, "Soil organic carbon dynamics, functions and management in West African agro-ecosystems," *Agric. Syst.*, vol. 94, no. 1, pp. 13–25, Apr. 2007, doi: 10.1016/j.agsy.2005.08.011.

[8] B. Stenberg, R. A. Viscarra Rossel, A. M. Mouazen, and J. Wetterlind, "Visible and Near Infrared Spectroscopy in Soil Science," in *Advances in Agronomy*, vol. 107, Elsevier, 2010, pp. 163–215. doi: 10.1016/S0065-2113(10)07005-7.

[9] V. L. Mulder, S. De Bruin, M. E. Schaepman, and T. R. Mayr, "The use of remote sensing in soil and terrain mapping — A review," *Geoderma*, vol. 162, no. 1–2, pp. 1–19, Apr. 2011, doi: 10.1016/j.geoderma.2010.12.018.

[10] I. Majeed, N. K. Purushothaman, P. Chakraborty, N. Panigrahi, H. B. Vasava, and B. S. Das, "Estimation of soil and crop residue parameters using AVIRIS-NG hyperspectral data," *Int. J. Remote Sens.*, vol. 44, no. 6, pp. 2005–2038, Mar. 2023, doi: 10.1080/01431161.2023.2195570.

[11] F. Castaldi *et al.*, "Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands," *ISPRS J. Photogramm. Remote Sens.*, vol. 199, pp. 40–60, May 2023, doi: 10.1016/j.isprsjprs.2023.03.016.

[12] S. R. Swain *et al.*, "Estimation of soil texture using Sentinel-2 multispectral imaging data: An ensemble modeling approach," *Soil Tillage Res.*, vol. 213, p. 105134, Sep. 2021, doi: 10.1016/j.still.2021.105134.

[13] F. Castaldi, A. Palombo, F. Santini, S. Pascucci, S. Pignatti, and R. Casa, "Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon," *Remote Sens. Environ.*, vol. 179, pp. 54–65, Jun. 2016, doi: 10.1016/j.rse.2016.03.025.

[14] T. Loiseau *et al.*, "Satellite data integration for soil clay content modelling at a national scale," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 82, p. 101905, Oct. 2019, doi: 10.1016/j.jag.2019.101905.

[15] C. Gomez, S. Dharumarajan, J.-B. Féret, P. Lagacherie, L. Ruiz, and M. Sekhar, "Use of Sentinel-2 Time-Series Images for Classification and Uncertainty Analysis of Inherent Biophysical Property: Case of Soil Texture Mapping," *Remote Sens.*, vol. 11, no. 5, p. 565, Mar. 2019, doi: 10.3390/rs11050565.

[16] F. Castaldi, S. Chabrillat, A. Don, and B. Van Wesemael, "Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects," *Remote Sens.*, vol. 11, no. 18, p. 2121, Sep. 2019, doi: 10.3390/rs11182121.

[17] K. Dvorakova, U. Heiden, K. Pepers, G. Staats, G. Van Os, and B. Van Wesemael, "Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties," *Geoderma*, vol. 429, p. 116128, Jan. 2023, doi: 10.1016/j.geoderma.2022.116128.

[18] Y. Bao, F. Yao, X. Meng, J. Zhang, H. Liu, and A. Mounem Mouazen, "Predicting soil organic carbon in cultivated land across geographical and spatial scales: Integrating Sentinel-2A and laboratory Vis-NIR spectra," *ISPRS J. Photogramm. Remote Sens.*, vol. 203, pp. 1–18, Sep. 2023, doi: 10.1016/j.isprsjprs.2023.07.020.

[19] Y. Zhu, L. Qi, Z. Wu, and P. Shi, "Spectra-based predictive mapping of soil organic carbon in croplands: Single-date versus multitemporal bare soil compositing approaches," *Geoderma*, vol. 449, p. 116987, Sep. 2024, doi: 10.1016/j.geoderma.2024.116987.

[20] M. Sadeghi, E. Babaeian, M. Tuller, and S. B. Jones, "The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations," *Remote Sens. Environ.*, vol. 198, pp. 52–68, Sep. 2017, doi: 10.1016/j.rse.2017.05.041.

[21] E. Ben-Dor *et al.*, "Using Imaging Spectroscopy to study soil properties," *Remote Sens. Environ.*, vol. 113, pp. S38–S55, Sep. 2009, doi: 10.1016/j.rse.2008.09.019.

[22] Y.-H. Tseng, "Spectral unmixing for the classification of hyperspectral images," *Int. Arch. Photogramm. Remote Sens.*, vol. 33, no. B7/4; PART 7, pp. 1532–1538, 2000.

[23] N. Keshava, "A survey of spectral unmixing algorithms," *Linc. Lab. J.*, vol. 14, no. 1, pp. 55–78, 2003.

[24] K. Dvorakova, P. Shi, Q. Limbourg, and B. Van Wesemael, "Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues," *Remote Sens.*, vol. 12, no. 12, p. 1913, Jun. 2020, doi: 10.3390/rs12121913.

[25] D. K. Ray and J. A. Foley, "Increasing global crop harvest frequency: recent trends and future directions," *Environ. Res. Lett.*, vol. 8, no. 4, p. 044041, Dec. 2013, doi: 10.1088/1748-9326/8/4/044041.

[26] M. Shabou *et al.*, "Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands," *Remote Sens.*, vol. 7, no. 5, pp. 6059–6078, May 2015, doi: 10.3390/rs70506059.

[27] J. A. M. Demattê *et al.*, "Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring," *Sci. Rep.*, vol. 10, no. 1, p. 4461, Mar. 2020, doi: 10.1038/s41598-020-61408-1.

[28] B. C. Gallo *et al.*, "Multi-Temporal Satellite Images on Topsoil Attribute Quantification and the Relationship with

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

Soil Classes and Geology," *Remote Sens.*, vol. 10, no. 10, p. 1571, Oct. 2018, doi: 10.3390/rs10101571.

[29] C. Luo *et al.*, "Regional mapping of soil organic matter content using multitemporal synthetic Landsat 8 images in Google Earth Engine," *CATENA*, vol. 209, p. 105842, Feb. 2022, doi: 10.1016/j.catena.2021.105842.

[30] E. Vaudour *et al.*, "Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 96, p. 102277, Apr. 2021, doi: 10.1016/j.jag.2020.102277.

[31] U. Heiden *et al.*, "Soil Reflectance Composites—Improved Thresholding and Performance Evaluation," *Remote Sens.*, vol. 14, no. 18, p. 4526, Sep. 2022, doi: 10.3390/rs14184526.

[32] F. Castaldi, "Sentinel-2 and Landsat-8 Multi-Temporal Series to Estimate Topsoil Properties on Croplands," *Remote Sens.*, vol. 13, no. 17, p. 3345, Aug. 2021, doi: 10.3390/rs13173345.

[33] N. E. Q. Silvero *et al.*, "Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison," *Remote Sens. Environ.*, vol. 252, p. 112117, Jan. 2021, doi: 10.1016/j.rse.2020.112117.

[34] P. Shi, J. Six, A. Sila, B. Vanlauwe, and K. Van Oost, "Towards spatially continuous mapping of soil organic carbon in croplands using multitemporal Sentinel-2 remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 193, pp. 187–199, Nov. 2022, doi: 10.1016/j.isprsjprs.2022.09.013.

[35] S. Chabrillat, E. Ben-Dor, J. Cierniewski, C. Gomez, T. Schmid, and B. Van Wesemael, "Imaging Spectroscopy for Soil Mapping and Monitoring," *Surv. Geophys.*, vol. 40, no. 3, pp. 361–399, May 2019, doi: 10.1007/s10712-019-09524-0.

[36] A. Gasmi, C. Gomez, P. Lagacherie, H. Zouari, A. Laamrani, and A. Chehbouni, "Mean spectral reflectance from bare soil pixels along a Landsat-TM time series to increase both the prediction accuracy of soil clay content and mapping coverage," *Geoderma*, vol. 388, p. 114864, Apr. 2021, doi: 10.1016/j.geoderma.2020.114864.

[37] Z. Shi *et al.*, "Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations," *Sci. China Earth Sci.*, vol. 57, no. 7, pp. 1671–1680, Jul. 2014, doi: 10.1007/s11430-013-4808-x.

[38] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. in Advanced Applications in Pattern Recognition. Boston, MA: Springer, 1981. doi: 10.1007/978-1-4757-0450-1.

[39] R. A. Viscarra Rossel *et al.*, "A global spectral library to characterize the world's soil," *Earth-Sci. Rev.*, vol. 155, pp. 198–230, Apr. 2016, doi: 10.1016/j.earscirev.2016.01.012.

[40] J. Heil, V. Häring, B. Marschner, and B. Stumpe, "Advantages of fuzzy k-means over k-means clustering in the classification of diffuse reflectance soil spectra: A case study with West African soils," *Geoderma*, vol. 337, pp. 11–21, Mar. 2019, doi: 10.1016/j.geoderma.2018.09.004.

[41] Y. Ogen, J. Zaluda, N. Francos, N. Goldshleger, and E. Ben-Dor, "Cluster-based spectral models for a robust assessment of soil properties," *Geoderma*, vol. 340, pp. 175–184, Apr. 2019, doi: 10.1016/j.geoderma.2019.01.022.

[42] Y. Ma, B. Minasny, J. A. M. Demattê, and A. B. McBratney, "Incorporating soil knowledge into machine-learning prediction of soil properties from soil spectra," *Eur. J. Soil Sci.*, vol. 74, no. 6, p. e13438, Nov. 2023, doi: 10.1111/ejss.13438.

[43] L. Peng, H. Cheng, L.-J. Wang, and D. Zhu, "Comparisons of the prediction results of soil properties based on fuzzy *c* -means clustering and expert knowledge from laboratory Visible – Near-Infrared reflectance spectroscopy data," *Can. J. Soil Sci.*, vol. 101, no. 1, pp. 33–44, Mar. 2021, doi: 10.1139/cjss-2020-0025.

[44] V. S. Kale, MakarBodas, P. Chatterjee, and K. Pande, "Emplacement history and evolution of the Deccan Volcanic Province, India," *Episodes*, vol. 43, no. 1, pp. 278–299, Mar. 2020, doi: 10.18814/epiiugs/2020/020016.

[45] G. W. Gee and J. W. Bauder, "Particle-size Analysis," in *SSSA Book Series*, 1st ed., vol. 5, A. Klute, Ed., Wiley, 1986, pp. 383–411. doi: 10.2136/sssabookser5.1.2ed.c15.

[46] A. Walkley and I. A. Black, "AN EXAMINATION OF THE DEGTJAREFF METHOD FOR DETERMINING SOIL ORGANIC MATTER, AND A PROPOSED MODIFICATION OF THE CHROMIC ACID TITRATION METHOD:," *Soil Sci.*, vol. 37, no. 1, pp. 29–38, Jan. 1934, doi: 10.1097/00010694-193401000-00003.

[47] S. R. Olsen and L. E. Sommers, "Phosphorus," in *Agronomy Monographs*, 1st ed., vol. 9, A. L. Page, Ed., Wiley, 1982, pp. 403–430. doi: 10.2134/agronmonogr9.2.2ed.c24.

[48] J. Hanway, "Soil analysis methods as used in the iowa state college soil testing laboratory," *Iowa Agric.*, vol. 57, p. 1, 1952.

[49] R. Keren, "Boron," in *SSSA Book Series*, 1st ed., vol. 5, D. L. Sparks, A. L. Page, P. A. Helmke, R. H. Loeppert, P. N. Soltanpour, M. A. Tabatabai, C. T. Johnston, and M. E. Sumner, Eds., Wiley, 1996, pp. 603–626. doi: 10.2136/sssabookser5.3.c21.

[50] I. Majeed *et al.*, "Diffuse reflectance spectroscopy (DRS) for rapid soil testing and soil quality assessment in smallholder farms," *Eur. J. Soil Sci.*, vol. 74, no. 2, p. e13358, Mar. 2023, doi: 10.1111/ejss.13358.

[51] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.

[52] C. Lin, C.-C. Wu, K. Tsogt, Y.-C. Ouyang, and C.-I. Chang, "Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery," *Inf. Process. Agric.*, vol. 2, no. 1, pp. 25–36, May 2015, doi: 10.1016/j.inpa.2015.01.003.

[53] J. W. Rouse Jr, R. H. Haas, D. Deering, J. Schell, and J. C. Harlan, "Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation," 1974.

[54] A. Rikimaru, P. S. Roy, S. Miyatake, and others, "Tropical forest cover density mapping," *Trop. Ecol.*, vol. 43, no. 1, pp. 39–47, 2002.

[55] S. Diek, F. Fornallaz, M. E. Schaepman, and R. De Jong, "Barest Pixel Composite for Agricultural Areas Using Landsat Time Series," *Remote Sens.*, vol. 9, no. 12, p. 1245, Dec. 2017, doi: 10.3390/rs9121245.

[56] M. B. Ferraro, "Fuzzy k -Means: history and applications," *Econom. Stat.*, vol. 30, pp. 110–123, Apr. 2024, doi: 10.1016/j.ecosta.2021.11.008.

[57] T. J. Ross, *Fuzzy logic with engineering applications*. John Wiley & Sons, 2005.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146

[58] P. A. Burrough, P. F. M. Van Gaans, and R. Hootsmans, "Continuous classification in soil survey: spatial correlation, confusion and boundaries," *Geoderma*, vol. 77, no. 2–4, pp. 115–135, Jun. 1997, doi: 10.1016/S0016-7061(97)00018-9.

[59] A. B. McBratney, M. L. Mendonça Santos, and B. Minasny, "On digital soil mapping," *Geoderma*, vol. 117, no. 1–2, pp. 3–52, Nov. 2003, doi: 10.1016/S0016-7061(03)00223-4.

[60] V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[61] H. Wold, "Systems analysis by partial least squares," 1983.

[62] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[63] M. C. Sarathjith, B. S. Das, S. P. Wani, and K. L. Sahrawat, "Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils," *Geoderma*, vol. 267, pp. 1–9, Apr. 2016, doi: 10.1016/j.geoderma.2015.12.031.

[64] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 26, no. 2, pp. 211–243, Jul. 1964, doi: 10.1111/j.2517-6161.1964.tb00553.x.

[65] R. A. Viscarra Rossel and R. M. Lark, "Improved analysis and modelling of soil diffuse reflectance spectra using wavelets," *Eur. J. Soil Sci.*, vol. 60, no. 3, pp. 453–464, Jun. 2009, doi: 10.1111/j.1365-2389.2009.01121.x.

[66] U. K. Mandal *et al.*, "Assessing Soil Quality in a Semiarid Tropical Watershed Using a Geographic Information System," *Soil Sci. Soc. Am. J.*, vol. 75, no. 3, pp. 1144–1160, May 2011, doi: 10.2136/sssaj2009.0361.

[67] J. Durán *et al.*, "Temperature and aridity regulate spatial variability of soil multifunctionality in drylands across the globe," *Ecology*, vol. 99, no. 5, pp. 1184–1193, May 2018, doi: 10.1002/ecy.2199.

[68] S. R. Araújo, J. Wetterlind, J. A. M. Demattê, and B. Stenberg, "Improving the prediction performance of a large tropical vis- NIR spectroscopic soil library from B razil by clustering into smaller subsets or use of data mining calibration techniques," *Eur. J. Soil Sci.*, vol. 65, no. 5, pp. 718–729, Sep. 2014, doi: 10.1111/ejss.12165.

[69] E. Ben-Dor, "Quantitative remote sensing of soil properties," in *Advances in Agronomy*, vol. 75, Elsevier, 2002, pp. 173–243. doi: 10.1016/S0065-2113(02)75005-0.

[70] T. S. Müller, R. Dechow, and H. Flessa, "Inventory and assessment of pH in cropland and grassland soils in Germany[#]," *J. Plant Nutr. Soil Sci.*, vol. 185, no. 1, pp. 145–158, Feb. 2022, doi: 10.1002/jpln.202100063.

[71] E. Vaudour *et al.*, "The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands," *Remote Sens.*, vol. 11, no. 18, p. 2143, Sep. 2019, doi: 10.3390/rs11182143.

[72] I. O. A. Odeh, A. B. McBratney, and D. J. Chittleborough, "Design of optimal sample spacings for mapping soil using fuzzy-k-means and regionalized variable theory," *Geoderma*, vol. 47, no. 1–2, pp. 93–122, Aug. 1990, doi: 10.1016/0016-7061(90)90049-F.

[73] M. D. Steven, T. J. Malthus, F. Baret, H. Xu, and M. J. Chopping, "Intercalibration of vegetation indices from different sensor systems," Remote Sens. Environ., vol. 88, no. 4, pp. 412–422, 2003.

[74] S. Roy, S. R. Ojha, N. N. Reddy, R. N. Samal, and B. S. Das, "Suspended particulate matter and Secchi disk depth in the Chilika Lagoon from in situ and remote sensing data: A modified semi-analytical approach," Int. J. Remote Sens., vol. 43, no. 10, pp. 3628–3654, 2022.



Dr. Israr Majeed received his Ph.D. degree in Agricultural and Food Engineering from IIT Kharagpur, India. Currently, he is working as an Associate Scientist at the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. His research is focused on improving agricultural production using proximal and remote sensing approaches.



Dr. Debi Prasad Sahoo earned his B.Tech in Agricultural Engineering from Odisha University of Agriculture and Technology in 2014, followed by an M.Tech in Water Engineering and Management with distinction from the Indian Institute of Technology (IIT) Kharagpur in 2017. He completed his Ph.D. at IIT Kharagpur in 2023. He has over a year of postdoctoral research experience at CNR-IRPI, Italy, where he contributed to several international projects, including CCI-RD, CRYOTEMPO, and UAWOS. His expertise lies in hydrology and remote sensing, with a focus on the development of multi-functional landscapes that integrate water resource management, ecosystem services, and sustainable land use planning.



Dr Sourav Roy received his PhD from the Agricultural and Food Engineering Department of the Indian Institute of Technology (IIT) Kharagpur. Currently, he is working as an Associate Scientist at the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. His research lies in surface hydrological processes and water resource management, and land resources inventory.



Dr. N. Nagarjuna Reddy received his Ph.D. degree in Agricultural and Food Engineering from IIT Kharagpur, India. Currently, he is working as an Associate Scientist at the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. He works on digital soil mapping, remote sensing and landscape management.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3747146



Dr Venkata Radha is a Senior Scientist-Vadoze Zone Hydrology with more than 10 years of experience in hydrology, micro-meteorology, spatial modelling, water budgeting, and their applications in agriculture, water resources, and natural resource management. Dr Radha has a doctorate in Environmental Engineering from the University of Melbourne, Australia, and master's degrees in Geo-informatics and Mathematics. Dr Radha is the recipient of the prestigious Jhon Al-Wright Fellowship for pursuing his Ph.D.



Mr. Nagaraju Budama is an Associate Manager at the International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India. He holds an M. Tech in Geomatics from the Indian Institute of Surveying and Mapping, Survey of India, JNTU. Proficient in GIS and remote sensing for project analysis, skilled in conducting topographic surveys using Total Station and DGPS, and adept at creating thematic maps with ArcGIS.



Dr. Ramesh Singh is an Indian national who currently holds the position of Principal Scientist & Head-ICRISAT Development Centre under the Global Research Program on Resilient Farm & Food Systems, ICRISAT. He earned his Ph.D from the GB Pant University of Agriculture & Technology, Pantnagar, India, and has 25 years of research and development experience in the areas of landscape resource conservation following land resource inventory and hydrology-based approach.



Dr. Bhabani Sankar Das received the Ph.D. degree in Agronomy from Kansas State University, Manhattan, KS, USA in 1996. He works as a Professor in the Agricultural and Food Engineering Department, IIT Kharagpur, India. Dr. Das uses proximal and remote sensing approaches and geohydrological modeling for agricultural decision making.



Dr. Kaushal Garg is a Principal Scientist - Natural Resource Management with PhD, from the Indian Institute Technology Kharagpur with expertise in Water management, Irrigation Scheduling, and Hydrological modeling across various agro climatic zones.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/