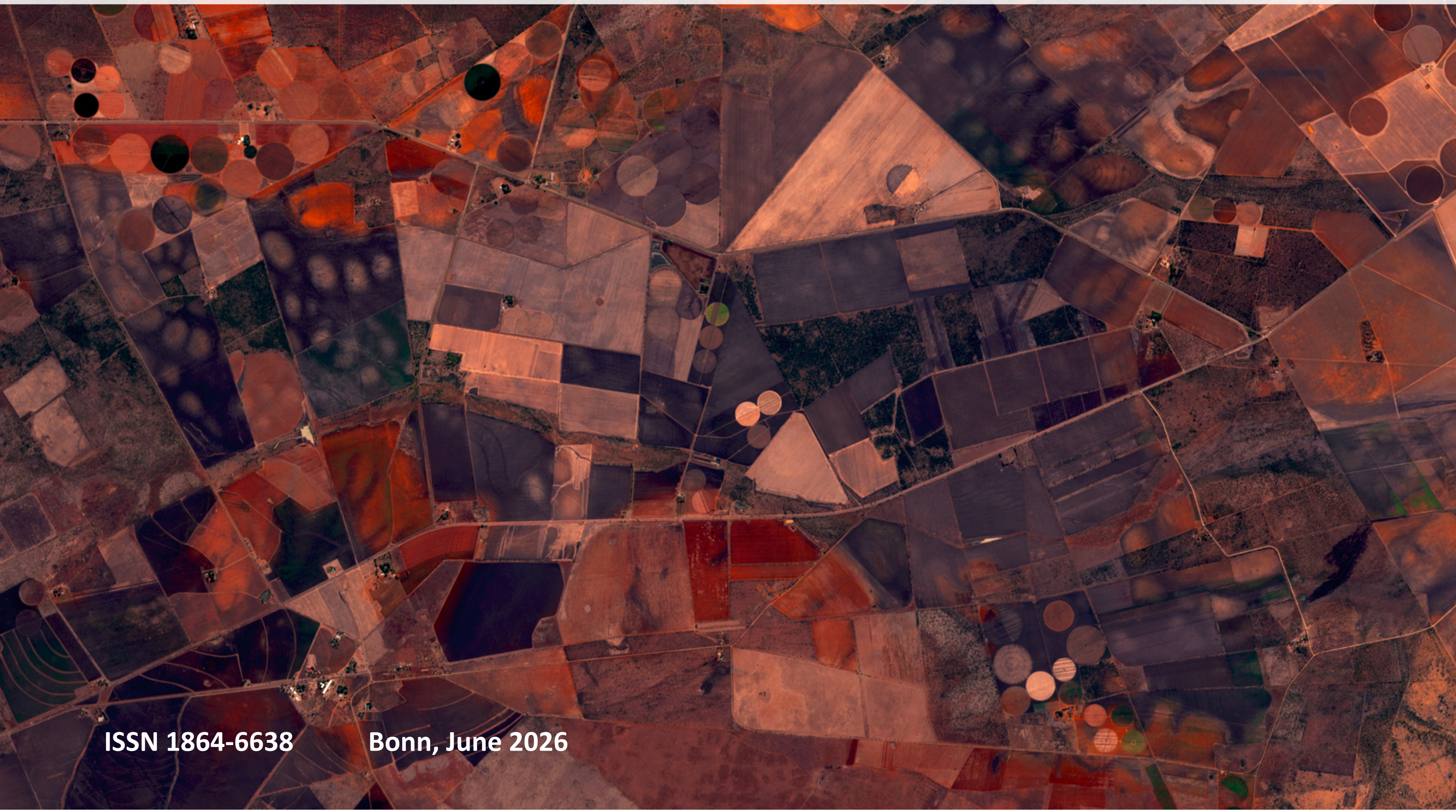
Working Paper 251

Friederike Schilling

Monitoring carbon farming from space – potentials and limitations



ISSN 1864-6638 Bonn, June 2026

ZEF Working Paper Series, ISSN 1864-6638
Center for Development Research, University of Bonn

Author's address

Friederike Schilling
Center for Development Research (ZEF)
Email: fschilling@uni-bonn.de

Monitoring carbon farming from space – potentials and limitations

Friederike Schilling

Abstract

Carbon farming initiatives require robust and scalable monitoring, reporting and verification (MRV) systems to quantify land management practices and their climate mitigation potential. Satellite-based remote sensing is often proposed as a cost-effective solution, yet its performance in heterogeneous smallholder farming systems is not well understood. This study investigates the ability of satellite-derived (Sentinel-2) bare soil indicators (Normalized Difference Vegetation Index (NDVI) and Bare Soil Index (BI)) to discriminate surface conditions in small agricultural plots (median=0.36 acres). The findings suggest: (1) Small plot sizes substantially constrain the possible use of Sentinel-2 data, as many plots are represented by few pixels and nearly one third are too small to apply an inward buffer zone of 10m to avoid mixed signals from neighbouring fields. (2) At the aggregate level, NDVI and BI correlate consistently with increasing bare soil intensity, with a statistically significant difference between fully bare and non-bare plots. (3) Substantial within-category variability and overlap in NDVI and BI ranges across bare soil classes indicate that plot-level discrimination between categories is unfeasible. (4) Restricting the analysis to larger plots does not improve the separability of NDVI and BI across bare soil categories. Instead, distributional overlap persists across all thresholds while statistical uncertainty increases due to reduced sample sizes. The findings highlight both the potential and current limitations of Sentinel-2 for smallholder carbon farming applications, emphasizing the need for higher-resolution data and improved analytical approaches. Future remote sensing-based MRV systems will likely rely on integrating freely available satellite data, field measurements, and advanced modelling techniques to generate reliable indicators of carbon farming monitoring.

Keywords: carbon farming, geospatial data, MRV systems

JEL codes: Q54, Q56, Q57, O13

Acknowledgments

This study was developed in the context of the Program of Accompanying Research for Agricultural Innovation (PARI), supported by the Federal German Ministry for Economic Cooperation and Development (BMZ).

Table of Contents

1 Introduction

As soils gain global recognition for their role as natural carbon sinks, interest has grown in sustainable agricultural practices aimed at enhancing carbon sequestration in agricultural soils, referred to as carbon farming (Griscom et al., 2017; Lal, 2004a). Practices that are commonly sought to contribute to soil organic carbon (SOC) accumulation include agroforestry, integrated soil fertility management, minimum tillage, cover crops, crop rotation, intercropping, and mulching (Lal, 2004b; Lal et al., 2018). The potential to monetize SOC through carbon credit schemes has further amplified this interest. It potentially provides a mechanism to attract climate finance by leveraging both private and public investments, while compensating farmers for the provision of ecosystem services generated by switching to sustainable practices (Lal, 2013; von Braun et al., 2021).

However, operationalizing SOC as a tradable commodity requires robust quantification of changes in carbon stocks, underscoring the need for cost-efficient measurement and monitoring approaches, particularly in smallholder systems (Conant et al., 2011; Raina et al., 2024). At the same time, carbon markets exhibit a growing demand for high-integrity carbon credits based on rigorous, results-based verification - a shift that is increasingly reshaping monitoring approaches while simultaneously affecting the cost structure and feasibility of carbon farming initiatives (Schilling et al., 2026a, b). The increasing emphasis on high-integrity carbon markets has led to a shift from activity-based monitoring (no soil samples) toward results-based approaches in carbon farming projects (verifying sequestration through soil samples) (Schilling et al., 2026a), a shift that is similarly visible in impact reporting in food systems projects operating outside carbon farming schemes (Deconinck et al., 2023). While this transition is motivated by the need for robust and verifiable climate impact claims, results-based monitoring systems are costly, labour, and time-intensive (Chatterjee et al., 2009).

Across all monitoring approaches, accurate estimates for single fields are deemed largely unfeasible, especially in the context of smallholder farmers. As a consequence, many carbon finance schemes are faced with a decision: focus on very large farms or aggregate monitoring at larger group or project scales. However, such aggregation introduces its own challenges, most notably the risk of free-riding, where some participants benefit from carbon payments without making proportional contributions to soil carbon sequestration, thereby weakening incentives for sustainable land management practices (Baumber et al., 2024; Mantey et al., 2024). This raises a critical but largely unexplored question (e.g., by Bradford et al., 2023): If direct monitoring of SOC changes at the individual field level is not feasible, how can participating farmers be compensated for their individual contributions?

This study explores the feasibility of using remote sensing–based proxies for allocating individual-level shares of overall carbon sequestration benefits within a hybrid framework that uses project-level measurement and periodic remeasurement of soil carbon stocks for verified emission reductions. In this approach, remotely sensed indicators of soil surface characteristics are used to verify the implementation of carbon farming practices and to approximate each farmer's proportional/ confirmed contribution to total project-level sequestration outcomes. Such an approach requires that remote sensing data can reliably identify plot-level surface characteristics associated with the implementation of carbon farming practices. Therefore, this study uses ground-truth data from Western Kenya to assess the potential of remote sensing for activity-based monitoring in smallholder farming systems. Specifically, the study addresses the following research questions:

(RQ 1) How accurately can satellite imagery identify different levels of bare soil?

(RQ 2) How does classification accuracy vary across plot size?

2 Research background

2.1 A hybrid framework for monitoring carbon farming

Early approaches applied by carbon farming projects under Verra's VCS methodology VM0017, relied exclusively on modelling without the need for soil sampling (Schilling et al, 2026a). The current predominant approach in carbon farming projects is a so-called measure-and-model approach (Potash et al., 2025; Oldfield et al., 2022). In this approach, traditional measurements of SOC are taken before a project starts. These measurements are then used to initialize a process-based biogeochemical model that estimates changes in SOC at each measurement location under the implemented practice as well as under a control (Potash et al., 2025). While process-based biogeochemical models rely on numerous assumptions, an alternative to it, the so-called measure-and-remeasure approach, offers a more direct solution that if well designed, can state with greater confidence real reductions in CO2 emissions (Potash et al., 2025).

In the measure-and-remeasure approach, soil samples are taken at the beginning of a project (baseline measurement) and then remeasured after a certain period of time. It is, however, oftentimes argued that measure-and-remeasure approaches to SOC monitoring are prohibitively expensive. Recent evidence, however, suggests that it can be feasible at larger population-scale levels (Bradford et al., 2023; Potash et al., 2025; Schilling et al., 2026a). Potash et al. (2025) show that economically feasible soil carbon monitoring can be achieved through spatially and temporally optimized sampling strategies, where intensive measurement of a limited number of fields is sufficient to estimate project-level average treatment effects rather than field-specific outcomes. Bradford et al. (2023) demonstrate that single-field SOC inventories remain highly unreliable due to substantial random sampling error, even at high sampling densities, but that accuracy improves substantially when estimating aggregated project-level mean changes in SOC stocks rather than individual field-level changes. Further, even the currently predominant process-based models generally do not provide accurate estimates for a single field, especially without detailed site-specific data (Ogle et al., 2010; Oldfield et al., 2022; Tonitto et al., 2018). While both approaches can provide robust estimates at project-level, their implementation at the individual smallholder plot level remains economically and logistically challenging due to high sampling costs, spatial heterogeneity, and fragmented landholdings.

One potential solution, hybrid monitoring approaches, have so far received little attention in the literature and have not yet been systematically developed, but they could potentially combine results-based verification at the project level with activity-based monitoring (through remote sensing) at the individual level. In addition, hybrid approaches may also be advantageous in larger-scale farming systems where plot-level payments based solely on repeated soil sampling would be financially feasible, as soil-sampling-based approaches typically generate carbon credits only over longer time horizons (e.g., every three to five years), resulting in delayed payments to farmers (Oldfield et al. 2022). Potash et al. (2025) (briefly) mention the potential of hybrid monitoring approaches that combine direct measurement with modeling techniques. In such a framework, modeled estimates could be used to support annual carbon payments to farmers, while periodic remeasurement based on an optimized sampling strategy (e.g., every five years) could serve for verification and the issuance of carbon credits. Model-based approaches, however, typically require reliable field-level information on changes in land management practices as key input variables. Historically, such activity-based monitoring has depended on farm surveys and field visits, which are often labour-intensive, costly, and difficult to implement at scale. Recent advances in the availability and spatial resolution of satellite-based earth observation data offer a potential alternative. These developments raise an important question: whether remotely sensed data are sufficiently accurate to enable scalable, activity-based monitoring of agricultural practices. If feasible, such an approach could substantially reduce monitoring costs and, when combined with periodic soil sampling, provide a more practical and scalable foundation for carbon accounting systems in smallholder agriculture.

The central premise of the hybrid framework is that such monitoring is only feasible if satellite imagery can reliably capture plot-level surface cover conditions that reflect underlying land management practices. In this context, carbon farming activities are not directly observable but manifest through biophysical surface states, which can be categorized as (i) bare soil, (ii) dead plant matter (non-photosynthetic vegetation), and (iii) living vegetation (photosynthetically active biomass). These categories serve as observable proxies for practices such as mulching, residue retention, and cover cropping. The interpretation relies on systematic linkages between management and surface conditions. However, carbon farming implementation is inherently temporal—particularly in systems such as conservation agriculture—where combinations of practices may occur and become visible at different points in the year. Against this background, the starting point of this study is a more fundamental question: whether satellite imagery can capture, at a single point in time, the surface cover states observed on the ground. This first step does not yet aim to quantify the full extent or temporal dynamics of carbon farming implementation, but rather assesses the feasibility of distinguishing different surface status classes in very small plots as a foundational requirement for any scaling of the approach. While this limits the scope of inference, it provides an essential entry point for understanding the potential of remote sensing in this context, particularly given the still limited literature on detecting agricultural management practices on smallholder farms.

2.2 Literature review

The optimal application of remote sensing for monitoring carbon farming would involve the development of a classification framework capable of identifying whether carbon farming practices are implemented at the field level. Such an approach would require the separate detection of key management components over time and their subsequent integration into a broader carbon farming classification (e.g., on the intensity of carbon farming implementation, number of practices).

While most studies look at individual practices, a recent study by Zhou et al. (2025) developed such a framework for the implementation of conservation agriculture. Using Sentinel-based time series data from Belgium, where average field sizes exceed 5 ha (minimum 1 ha), the study reported an overall classification accuracy of 92% for conservation agriculture systems, including accuracies of 86% for cover crop detection and 66–76% for tillage practices.

Studies that aim at classifying individual agricultural management practices are extensive, but generally focus on applications in the USA or Europe with larger plot sizes. For example, Azzari et al. (2019) estimated tillage intensity (binary indicator) in Belgium based on crop residue cover >30% and achieved an overall model accuracy of 79%, with particularly strong performance for detecting low-intensity tillage systems. Similarly, Zheng et al. (2012) demonstrated with data from the US a strong linear relationship between crop residue cover and the minimum normalized difference tillage index (minNDTI), enabling the classification of residue cover into three categories (0–30%, 30–70%, and >70%) with accuracies exceeding 90%.

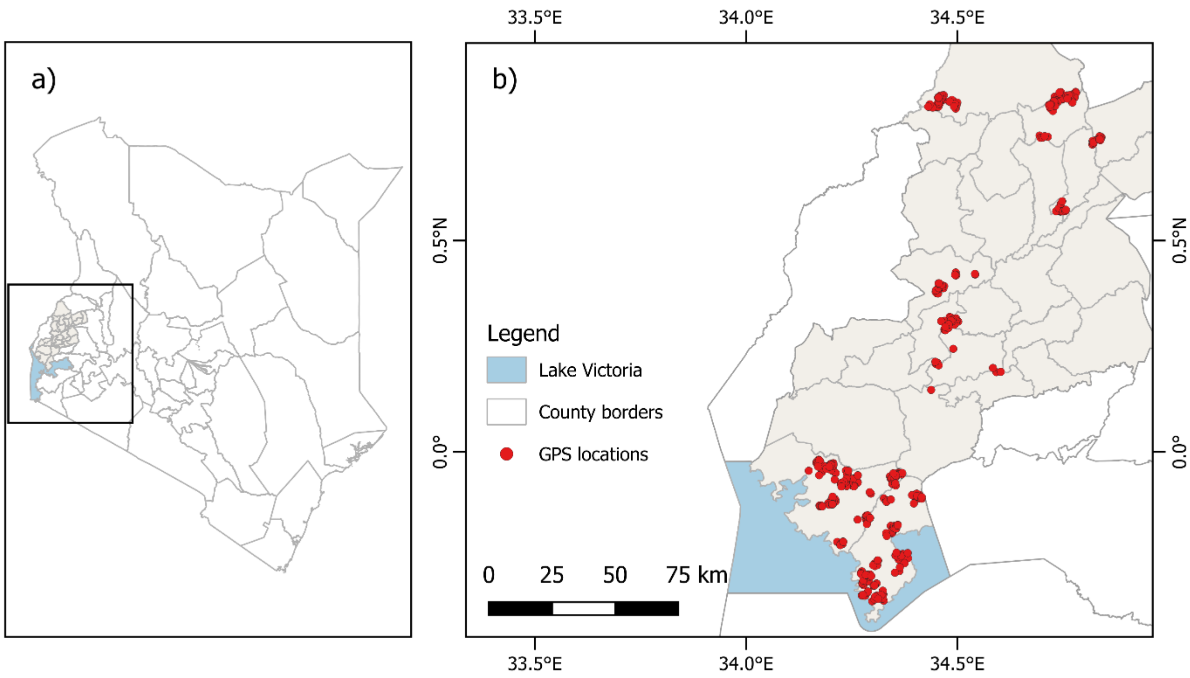
However, literature addressing remote sensing applications in smallholder agricultural systems remains limited. In India, Liu et al. (2022) used Sentinel-2 imagery to distinguish zero tillage from conventional tillage in plots smaller than 0.3 ha and reported a classification accuracy of 83.24% for the binary classification task. Likewise, Kosmowski et al. used single Landsat 8 images (30 m resolution) to classify crop residue cover in smallholder systems with average plot sizes of 1.2 ha. While the study achieved a classification accuracy of 83% for a binary 30% residue-cover threshold, performance was substantially weaker for continuous quantitative estimation and for multi-category classification schemes.

3 Methods and data

3.1 Description of study site

The study is based on a sample of 494 smallholder plots, located across three counties in Western Kenya, namely Bungoma (n=150), Kakamega (n=71), and Siaya (n=273) (Figure 1).

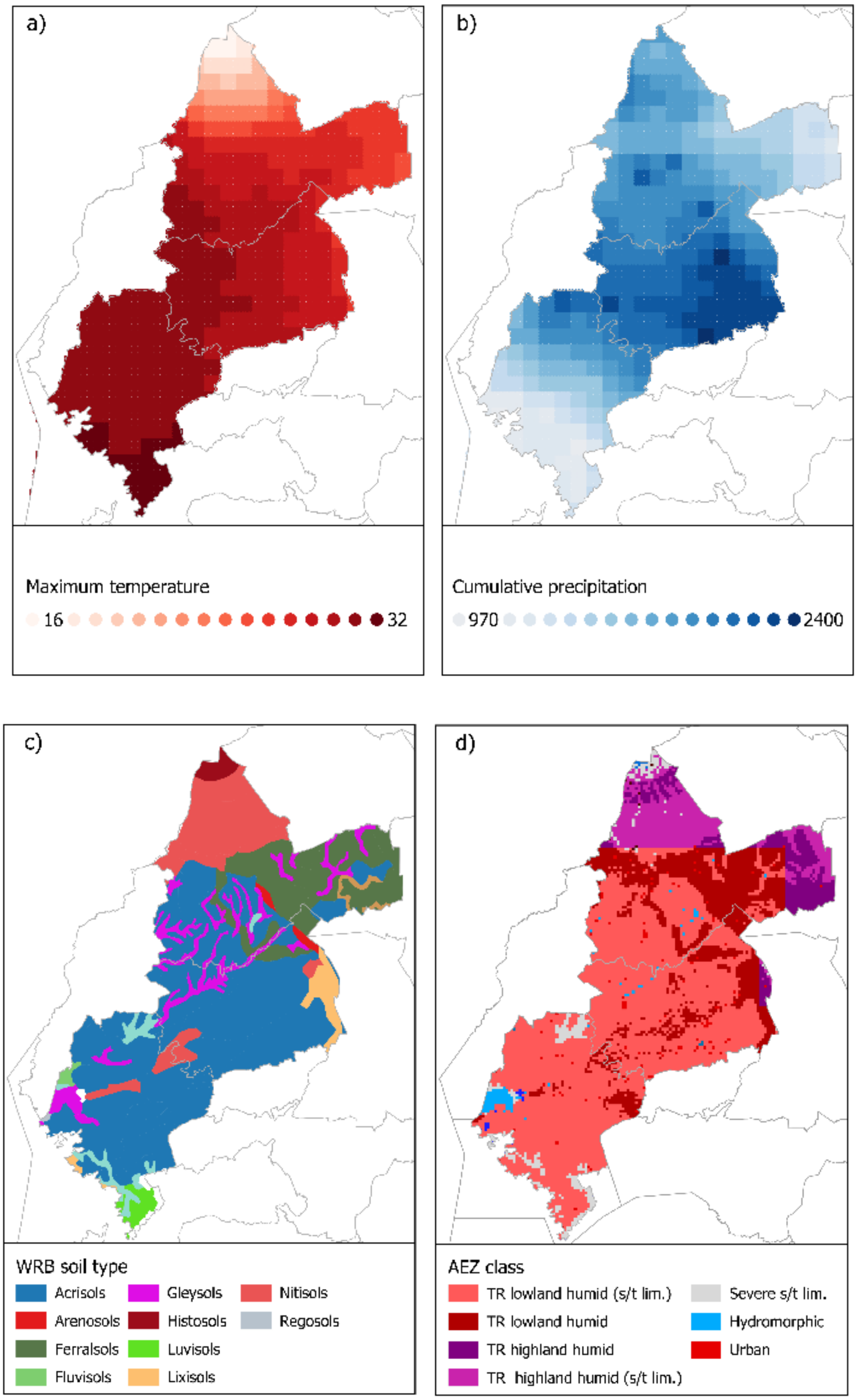
Figure 1: Distribution of sample plots across Bungoma, Kakamega and Siaya



Source: author's own illustration

The study site represents a range of biophysical characteristics varying in climate, topography, and soil properties (Figure 2). The study area is predominantly located in the lowland humid tropics. The only exception is Mount Elgon in north ern Bungoma, which is given higher altitudes classified as highland humid tropics (FAO, 2021). Annual cumulative rainfall is highest in Kakamega, particularly near Kakamega rainforest and lowest in southern Siaya near the Lake Victoria coastline. The study area experiences two rainy seasons: the long rains from March to May and the short rains from November to December. Temperature patterns mirror rainfall distributions, with the lowest temperatures recorded near Kakamega rainforest and high altitudes along Mount Elgon, while the highest temperatures are observed near Lake Victoria. A wide range of soil types characterizes the study area. Acrisols are the predominant soil type across the study area. Nitisols are found in the mountainous regions of Mount Elgon, Ferralsols in Bungome, and Luvisols in southern Siaya (KSS & ISRIC, 2007).

Figure 2: Maps of the biophysical characteristics of the study site: a) maximum average temperatures, b) average cumulative annual precipitation, c) WRB soil type classification, and d) agro-ecological zoning



Source: author's own illustration based on data from a) CHIRTS, b) CHIRPS v2, c) KENSOTER v2 and d) FAO GAEZ v4

3.2 Data availability and preparation: plot polygons and photographs

A total of 494 agricultural plots were visited during field data collection. Data collection took place over a three-week period between 5 February and 23 February. In week 1, 28% of plots were recorded. In week 2, 57% of plots were recorded. In week 3, 15% of plots were recorded.

For each plot, enumerators delineated plot boundaries by walking the perimeter and recording geospatial polygons using KoboCollect. In addition to polygon data, a single GPS point was recorded per plot to enable consistency checks. Plot polygons were subsequently visually validated by overlaying them on Google Earth imagery and by using the recorded GPS point as a reference. Polygons that did not match the visible field boundaries in Google Earth, were geometrically implausible (e.g. self-intersecting shapes or incorrectly ordered vertices resulting in distorted geometries such as triangles), or for which the GPS point did not fall within the recorded polygon were excluded from the dataset (26 observations).

Photographs were taken for all visited plots to document field conditions. The plot-level photographs are used to construct ground-truth data for surface cover classification. All images were manually inspected and classified. Unsuitable photographs (taken at night or at excessively close zoom to the ground) were excluded from the dataset (9 observations). Each plot was assessed based on the proportion of bare soil cover. Bare soil was first classified as a categorical variable reflecting different levels of surface exposure (zero, intermediate, and full bare soil cover) (Figure 3).

After performing the quality control (plausible plot polygon + suitable photo), 462 plots remained with valid polygons and corresponding usable photographs.

Figure 3: Bare soil classification

barecat=0: 0% bare soil or minimal	barecat=0.5: intermediate bare soil	barecat=1: 100% bare soil or close to

3.3 Remote sensing data: extraction and pre-processing

Satellite imagery is obtained from Sentinel-2, which provides multispectral imagery at 10-meter spatial resolution for visible and near-infrared bands and 20-meter resolution for additional spectral bands. This resolution is particularly relevant in smallholder contexts, where plot sizes are often comparable to or smaller than individual pixels, increasing the risk of mixed-pixel effects. Satellite data are extracted using Google Earth Engine, enabling efficient processing of multi-temporal and multi-band imagery. For each plot polygon, the means of six different spectral bands are extracted, namely band 2 (blue), 3 (green), 4 (red), 8 (NIR), 11 (SWIR1) and 12 (SWIR 2).

Sentinel-2 imagery used in this study was obtained from three adjacent Sentinel-2 tiles (36MXE, 36NXF, and 36NYF), covering the full spatial extent of the sampled plots. The majority of plots were located in tile 36NXF (61.5%), followed by 36MXE (approximately 35%), while a smaller share fell within tile 36NYF (approximately 15%). Image selection was guided primarily by cloud conditions within a pre-determined observation window (1 February to 15 March 2024). Given persistently cloudy conditions during this period, the initial cloud cover threshold of 10% was relaxed to 15% to ensure sufficient temporal coverage. This adjustment resulted in the selection of one suitable Sentinel-2 acquisition date for two tiles, and two acquisition dates for one tile (36NXF). Overall, four Sentinel-2 scenes were retained, with tile-level cloud cover ranging between 5.5% and 13.9%. Specifically, the selected scenes exhibited cloud cover of 5.5%, 8.5%, 9.4%, and 13.9%, respectively. To ensure data availability and quality at the plot level, all images were additionally subjected to manual visual inspection to confirm that individual plots were not affected by clouds or cloud shadows.

In total, 15 plot observations (3.25%) were matched with imagery from 5 February 2024, 207 observations (44.81%) with imagery from 20 February 2024, and 240 observations (51.95%) with imagery from 11 March 2024. The temporal distance between field photograph acquisition and the corresponding Sentinel-2 observation was, on average, 16 days (median: 18 days). In 38.3% of cases, field photographs were taken within ±10 days of the respective satellite overpass date.

To reduce boundary-related contamination from mixed pixels, buffer zones are applied by excluding pixels along plot edges (-10 m), ensuring that only interior pixels are included in the analysis. This is also done by Zhou et al. (2025), who used a 15m inward buffer zone to remove mixed signals from neighbouring fields or adjacent roads. In the subsequent analysis, both the full sample and the restricted (buffered) sample are retained to assess the robustness of the results to edge-related pixel contamination.

3.4 Bare soil indicators

The Normalized Difference Vegetation Index (NDVI) is a widely used indicator of vegetation greenness, defined as the normalized ratio between near-infrared (NIR) and red reflectance. NDVI is calculated using the following formula:

$$\frac{NIR - RED}{NIR + RED} \quad (1)$$

, where RED refers to Sentinel-2 band 4 and NIR to band 8. The NDVI values range from -1 to 1: negative values indicate water bodies, values close to zero represent barren areas such as rock, sand, or snow, low positive values correspond to shrub and grassland, and high values nearing 1 indicate dense temperate and tropical rainforests (Huang et al., 2021; Pettorelli et al., 2005). Values between 0 and 0.25 are commonly considered bare soil, while values above 0.25 may be considered vegetation (Demattê et al., 2018; Xue et al., 2024).

NDVI captures vegetation presence and density due to the strong contrast between red absorption by chlorophyll and high NIR reflectance of healthy vegetation. While NDVI tends to saturate in dense vegetation and is less effective at distinguishing bare soil from other low-vegetation surfaces, these limitations may be partly outweighed by the advantage of deriving the index only from bands available at a 10 m spatial resolution, which may improve the representation of vegetation variability at the level of small agricultural plots and may reduce mixed-pixel effects in fragmented farming landscapes.

A bare soil index is an index used to estimate the proportion of bare soil in an area. The study follows Rikimaru et al. (2002) and defines the bare soil index (BI) using the formula:

$$\frac{(SWIR1 + RED) - (NIR + BLUE)}{(SWIR1 + RED) + (NIR + BLUE)} * 100 + 100 \qquad (2)$$

, where RED refers to Sentinel-2 band 4, NIR to band 8, SWIR to band 11, and BLUE to band 2. The index ranges between 0 and 200, with increasing values corresponding to increased bare soil exposure (Rikimaru et al., 2002). Commonly, values above 102 are considered the threshold for bare soil identification (Diek et al., 2017; Xue et al., 2024). SWIR 1 is the only band in this formula available at a resolution of 20m only.

3.5 Empirical approach

The plot polygons form the spatial basis for linking ground observations with remotely sensed data. To assess the suitability of remotely sensed bare soil indicators, continuous bare soil indices derived from Sentinel-2 imagery (introduced above) are compared with a categorical bare soil classification based on manual plot inspection. First, the relationship between each bare soil index and the categorical classification is explored graphically using boxplots and density distributions. Second, summary statistics of the indices are calculated by bare soil category, including means, medians, and standard deviations, and scatter plots are generated to assess the extent to which the indices differentiate within- and between categories.

To address the second research question, which examines heterogeneity in classification performance across plot sizes, the analysis incorporates plot size as a key stratification variable. Plot areas are computed from the georeferenced polygons, and the sample is divided into smaller and larger plots using four different cut-off values (0.35 acres, 0.5 acres, 0.7 acres, and 1 acre). The analyses are then rerun separately for different plot size groups, and classification accuracy is compared across these subsamples to assess how plot size affects the reliability of satellite-based surface cover classification.

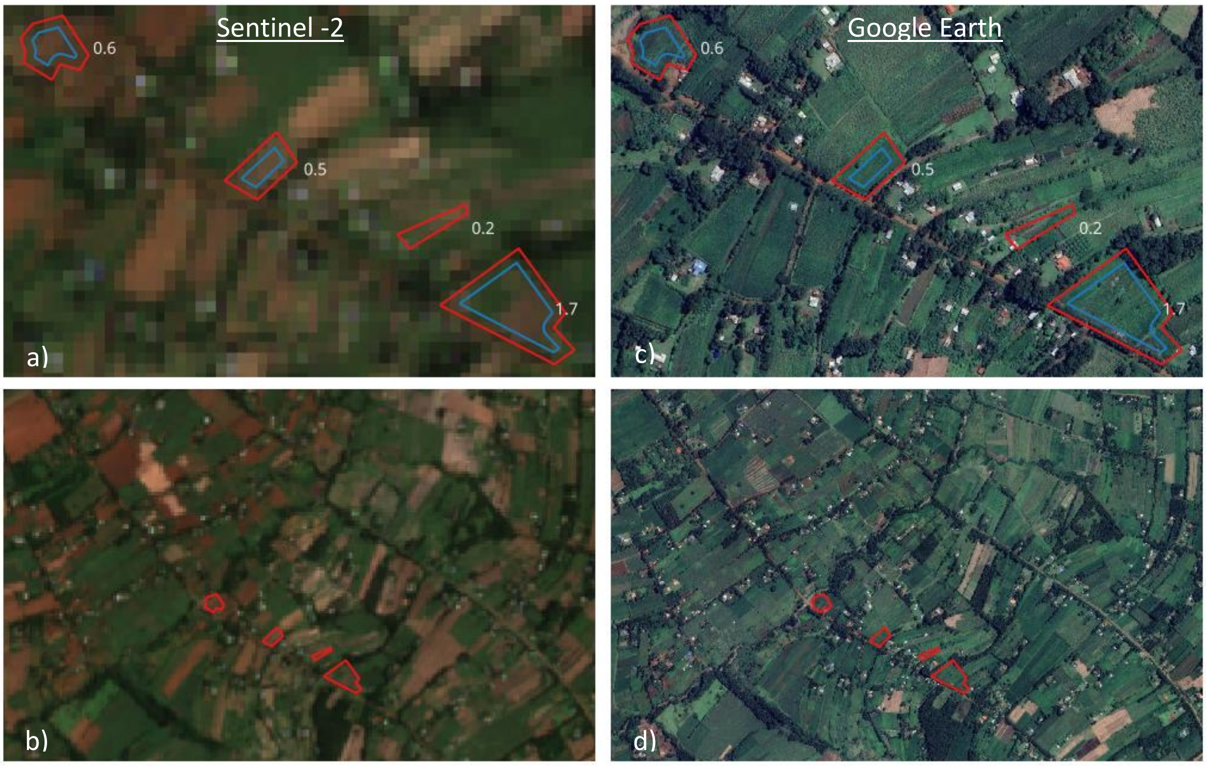
4 Results

4.1 Plot size, granularity and availability of remote sensing data

Study plots have an average size of 0.51 acres and a median size of 0.36 acres. Plot sizes range from 0.01 to 10.31 acres. At the spatial resolution of Sentinel-2 imagery, an average of 21 pixels is available per plot, with a median of 15 pixels. In total, 10.4% of plots are represented by five pixels or fewer. After applying a 10-m inward buffer to plot boundaries to avoid mixed signals from plot borders or neighbouring fields, 29% of plots (134 plots of 462 quality-controlled plots) no longer retained a valid polygon shape. For the remaining plots, the average number of available Sentinel-2 pixels decreased to 11.1 pixels per plot, with a median of 6 pixels. Given this substantial reduction in usable observations, the results presented in Sections 4.2 and 4.3 are based on the full sample without applying the buffer, while the buffered sample is used as a robustness check to assess the sensitivity of the findings to potential edge effects and mixed-pixel contamination.

Figure 4 illustrates plot boundaries of four plots at different spatial resolutions and zoom levels using Sentinel-2 imagery (panels a and b) and Google Earth imagery (panels c and d). The red outline represents the original plot boundary, while the blue outline represents the plot boundary after applying a 10-m inward buffer. In the Sentinel-2 imagery, plot boundaries become less visually distinguishable at closer zoom levels (panel a), whereas at a more distant zoom level (panel b), individual plots remain visible within the landscape. Plot sizes are reported in acres for each example. The smallest plot in this example, measuring 0.2 acres, does not retain a valid polygon shape after applying the inward buffer, resulting in the disappearance of the blue boundary.

Figure 4: Exploration of Sentinel-2 resolution compared to Google Earth imagery



4.2 Correlation between Sentinel-2 indicators and manual bare soil classification

Descriptive statistics indicate systematic differences in NDVI across bare soil categories. Mean NDVI declines from 0.572 in plots with no bare soil (barecat = 0, n = 165) to 0.552 in intermediate conditions (barecat = 0.5, n = 254), and further to 0.443 in plots classified as fully bare (barecat = 1, n = 43) (Figure 5 a). Overall, these patterns suggest a clear negative association between observed bare soil intensity and NDVI, consistent with NDVI capturing decreasing vegetation cover across categories. Pairwise comparisons indicate that NDVI differences are statistically significant between bare plots (barecat = 1) and the other two categories, whereas the confidence intervals for the intermediate (barecat = 0.5) and non-bare category (barecat = 1) overlap (Figure 5 a). This suggests that, at the aggregate level, NDVI can separate vegetated or non-bare plots from strongly bare conditions, but provides less clear discrimination between intermediate and fully bare classes. Descriptive statistics for the Bare Soil Index (BI) similarly indicate systematic differences across bare soil categories. Mean BI increases from 96.8 in plots with no bare soil (barecat = 0, n = 165) to 98.8 in intermediate conditions (barecat = 0.5, n = 254), and further to 106.9 in plots classified as fully bare (barecat = 1, n = 43). Pairwise comparisons further indicate that BI more clearly separates the extreme bare soil/ non-bare categories than intermediate conditions (Figure 5 b).

Figure 5: Mean NDVI and BI values across the three categories of bare soil

a) Normalized Difference Vegetation Index (NDVI)	b) Bare Soil Index (BI)

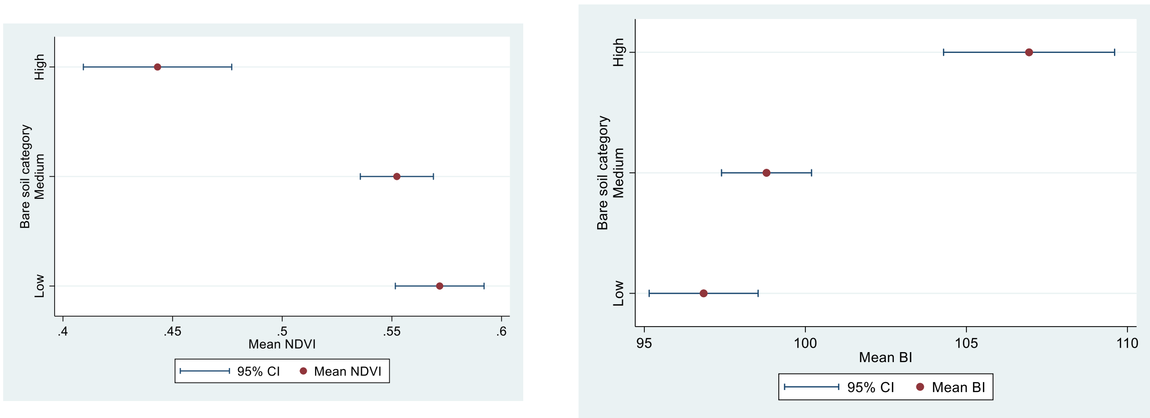
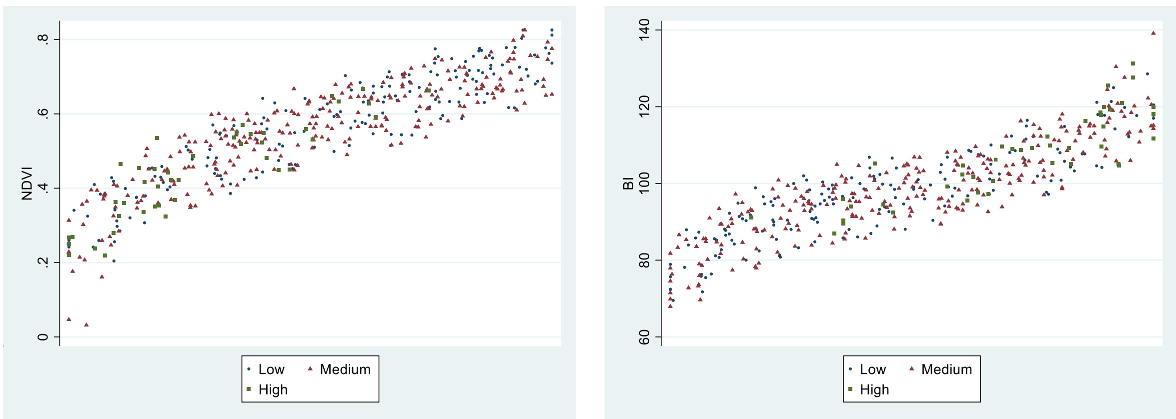


Figure 6: Scatter plot of NDVI and BI

a) Normalized Difference Vegetation Index (NDVI)	b) Bare Soil Index (BI)



While the preceding results demonstrate systematic differences in mean NDVI and BI across bare soil categories at the aggregate level, the following analysis examines the distribution of observations at the individual plot level to assess within-category variability and the extent of overlap between classes. Within-category distributional characteristics highlight substantial heterogeneity in NDVI across bare soil classes. The observed ranges show considerable overlap between groups: NDVI spans from 0.226 to 0.796 for barecat = 0, from 0.047 to 0.826 for barecat = 0.5, and from 0.159 to 0.684 for barecat = 1, indicating that all categories share broadly similar ranges. This overlap is particularly pronounced for the intermediate class, which covers almost the full observed NDVI range. Consistent with this pattern, coefficients of variation are relatively high across all groups (approximately 0.23–0.26), indicating substantial dispersion relative to group means. The scatter plots (Figure 6) further corroborate the high degree of within-category variability observed in the summary statistics. However, for the high bare soil category (barecat = 1), a somewhat more pronounced clustering at lower NDVI values can be observed (Figure 6, b).

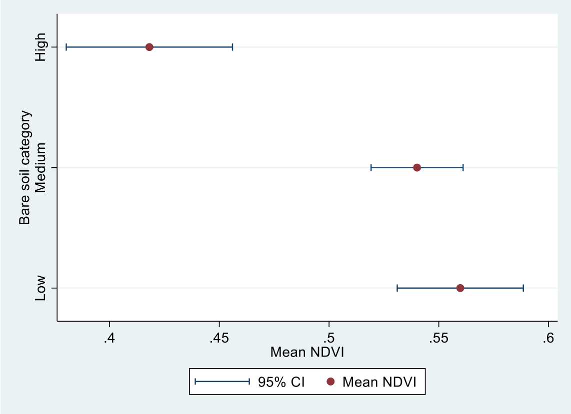
4.3 Correlation between Sentinel-2 indicators and manual bare soil classification by plot size

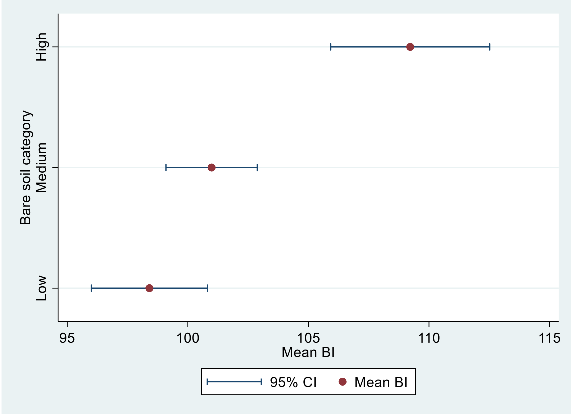
To test whether plot size affects the separability of NDVI and BI across bare soil categories, the analysis is repeated using progressively more restrictive subsamples based on alternative thresholds for the minimum plot size (>0.35 acres, >0.5 acres, >0.7 acres, and =1 acre). This exercise addresses whether alignment between plot extent and the spatial resolution of the remote sensing data improves classification performance. However, the results do not indicate improved separability of NDVI and BI across categories at coarser sample restrictions. Across all thresholds, NDVI and BI distributions remain substantially overlapping, and no clearer differentiation between groups emerges. Instead, confidence intervals widen as sample size decreases, particularly in the most restrictive subsamples, reflecting increased statistical uncertainty rather than enhanced discriminatory power (Figure 7).

Figure 7: Mean NDVI and BI values across the three categories of bare soil by plot size

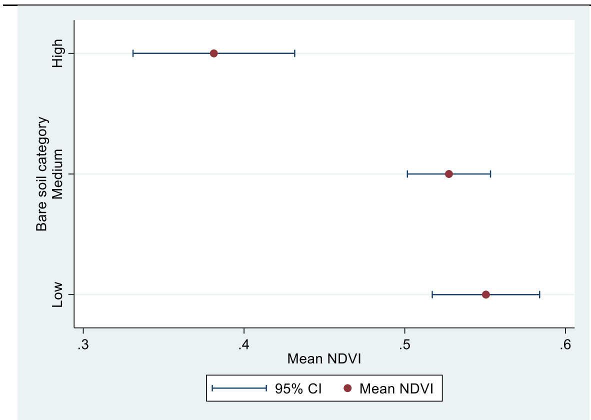
NDVI	BI

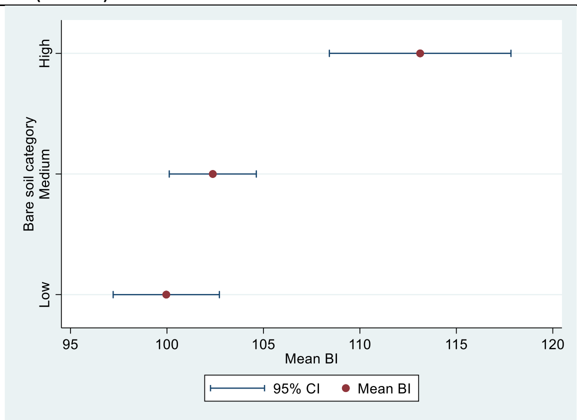
Plots > 0.35 acres (n=223)



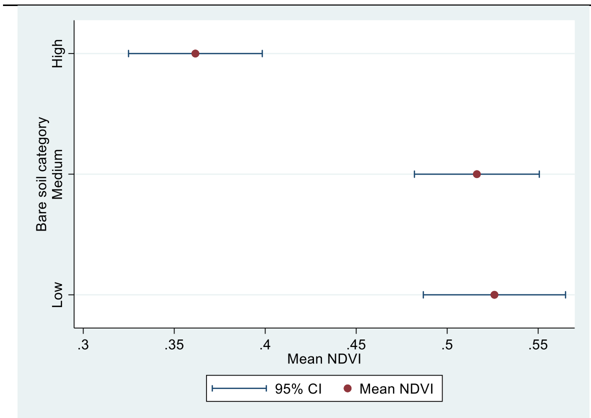


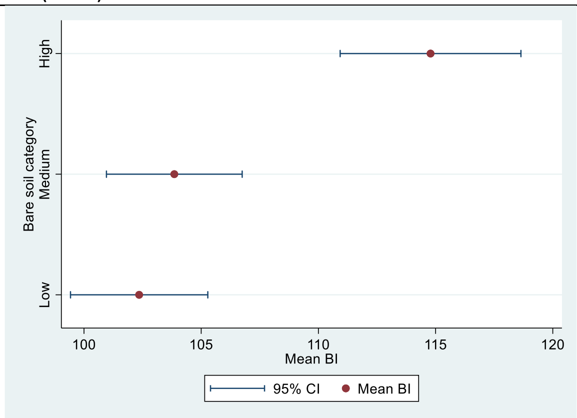
Plots > 0.5 acres (n=153)



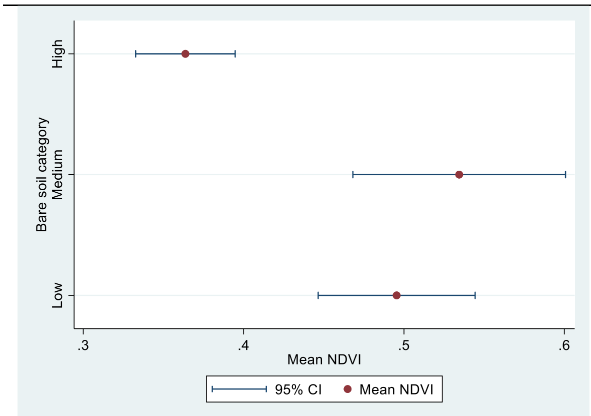


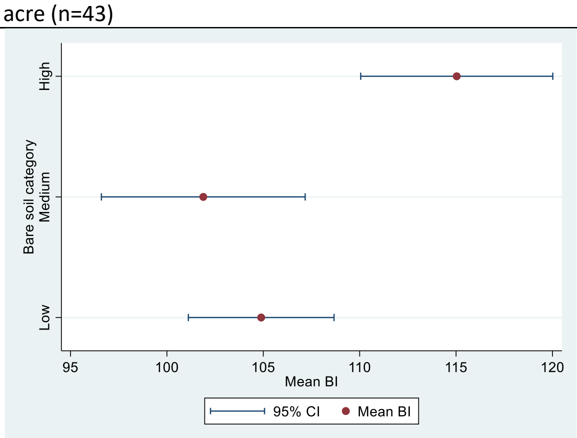
Plots > 0.7 acres (n=98)





Plots > 1 acre (n=43)





5 Opportunities and future directions for remote sensing in carbon farming projects

Remote sensing has increasingly gained attention as a potential tool to support MRV systems for carbon farming projects by providing spatially explicit, scalable, and cost-effective information on land management practices and ecosystem changes. Compared to conventional field-based monitoring approaches, satellite-based observations offer the possibility to continuously track vegetation dynamics, land cover changes, and management-related indicators over large areas. However, this study shows that the applicability of Sentinel-2 data is constrained by small plot sizes, which limit the number of usable pixels and prevent reliable buffering to avoid mixed signals from neighbouring fields. While NDVI and Bare Soil Index values exhibit consistent aggregate-level relationships with increasing bare soil intensity, clear plot-level discrimination between surface condition classes is not possible due to substantial within-class variability and overlapping spectral distributions. Moreover, restricting the analysis to larger plots does not improve classification accuracy, as distributional overlap persists across all thresholds and statistical uncertainty increases with decreasing sample size.

The results demonstrate both the potential and limitations of freely available medium-resolution satellite imagery, particularly Sentinel-2, for smallholder contexts, especially when applied to relatively basic spectral–surface associations as explored in this study. At the same time, they point to opportunities for improved performance through more advanced data integration, higher-resolution imagery, and more sophisticated analytical approaches. With its 10 m spatial resolution and frequent revisit time, Sentinel-2 represents a major advancement for agricultural monitoring and provides valuable information on vegetation dynamics through spectral indices. In this study, Sentinel-2-derived indicators were able to capture meaningful differences in surface conditions at an aggregated level, as shown by the consistent relationship between increasing bare soil intensity, decreasing NDVI, and increasing Bare Soil Index values. This suggests that remote sensing can successfully detect broad patterns related to vegetation cover and soil exposure, which are relevant proxies for several carbon farming practices. However, translating these capabilities into reliable plot-level monitoring remains challenging, particularly in smallholder farming systems characterized by fragmented landscapes and small plot sizes. The limited number of available pixels per plot reduces the ability of medium-resolution imagery to capture within-plot variability and increases the influence of mixed pixels, where spectral information represents a combination of soil, vegetation, residues, and other surface features. The results further show that increasing plot homogeneity through size restrictions alone does not necessarily improve classification performance, indicating that spatial resolution is only one factor influencing the effectiveness of remote sensing-based monitoring.

For carbon farming projects, these findings underline that remote sensing is (currently) unlikely to replace field-based measurements entirely but can increasingly complement them by improving spatial coverage and reducing monitoring costs. Proposed monitoring approaches combine satellite imagery with field observations, soil sampling, farmer surveys, and project-level management data. Such integrated MRV systems allow remote sensing to provide continuous landscape-level information while field data provide the ground truth required to interpret spectral signals and quantify changes in carbon stocks.

Future developments in remote sensing offer several opportunities to overcome current limitations:

1. **Availability and accessibility of sufficiently detailed remote sensing data will determine whether carbon farming can effectively capitalize on remote sensing opportunities.** Freely available, high-resolution, and timely Earth observation data are essential for scalable applications in smallholder contexts. Recent developments highlight both the potential and remaining gaps in data accessibility, including previously available high-resolution products such as Planet imagery at 3 m resolution and emerging satellite missions such as the ESA Biomass satellite. However, the practical uptake of remote sensing depends not only on spatial resolution but also on consistent data availability, timely acquisition, and user-friendly

extraction and processing platforms (e.g., cloud-based tools such as Google Earth Engine) that enable projects to integrate remote sensing into operational MRV systems.

2. **Higher spatial and temporal resolution data can substantially improve monitoring of smallholder carbon farming practices.** Higher-resolution imagery can reduce mixed-pixel effects and provide more reliable information at the plot-level, particularly in fragmented landscapes. Combining different satellite platforms with complementary spatial and temporal characteristics may further improve the detection of management changes relevant for carbon farming, such as changes in vegetation cover, residue management, or crop establishment.
3. **Future applications will increasingly rely on integrated modelling approaches rather than individual spectral indicators.** Machine learning approaches can combine multiple data sources, including spectral information, vegetation trajectories, climate variables, topography, soil characteristics, and management data, to capture complex relationships between remotely sensed signals and agricultural practices. However, their success depends on the availability of high-quality training data, independent validation datasets, and approaches that ensure model transferability across regions and farming systems.
4. **Great potential lies in AI-supported integration of remote sensing with field measurements and soil carbon modeling.** While remote sensing alone cannot directly quantify soil organic carbon (SOC) stocks at the field scale in most smallholder systems, it can provide spatially continuous environmental information and management proxies that improve SOC modelling, support more efficient field sampling, and strengthen MRV systems.

Overall, this study highlights that remote sensing already provides valuable information for carbon farming monitoring, but current satellite products and methods have limitations when applied to highly fragmented smallholder landscapes. The future of remote sensing-based carbon farming MRV will likely depend on integrating multiple data sources, combining satellite observations with field measurements, and applying advanced analytical methods to translate remotely sensed signals into reliable estimates of agricultural practices and carbon outcomes.

6 References

Azzari, G., Grassini, P., Edreira, J. I. R., Conley, S., Mourtzinis, S., & Lobell, D. B. (2019). Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. Remote Sensing of Environment, 221, 417–429. https://doi.org/10.1016/j.rse.2018.11.010

Baumber, A., Cross, R., Ampt, P., Waters, C., Ringbauer, J., Bowdler, I., Scott, A., Gordon, L., Sutton, A., & Metternicht, G. (2024). Soil-based carbon farming: Opportunities for collaboration. Journal of Rural Studies, 108, 103268. https://doi.org/10.1016/j.jrurstud.2024.103268

Bradford, M. A., Eash, L., Polussa, A., Jevon, F. V., Kuebbing, S. E., Hammac, W. A., Rosenzweig, S., & Oldfield, E. E. (2023). Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling. Geoderma, 440, 116719. https://doi.org/10.1016/j.geoderma.2023.116719

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M. Z., & Ebinger, M. H. (2009). Evaluation of different soil carbon determination methods. Critical Reviews in Plant Science, 28(3), 164–178. https://doi.org/10.1080/07352680902776556

Conant, R. T., Ogle, S. M., Paul, E. A., & Paustian, K. (2011). Measuring and monitoring soil organic carbon stocks in agricultural lands for climate mitigation. Frontiers in Ecology and the Environment, 9(3), 169–173. https://doi.org/10.1890/090153

Deconinck, K., Jansen, M., & Barisone, C. (2023). Fast and furious: The rise of environmental impact reporting in food systems. European Review of Agricultural Economics, 50(4), 1310–1337. https://doi.org/10.1093/erae/jbad018

Demattê, J. A. M., Fongaro, C. T., Rizzo, R., & Safanelli, J. L. (2018). Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sensing of Environment, 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047

Diek, S., Fornallaz, F., Schaepman, M. E., & De Jong, R. (2017). Barest pixel composite for agricultural areas using landsat time series. Remote Sensing, 9(12), 1245–1245. https://doi.org/10.3390/rs9121245

FAO. (2021). FAO Global Agro Ecological Zones v4. https://gaez.fao.org/pages/agromaps

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., & Smith, P. (2017). Natural climate solutions. Proceedings of the National Academy of Sciences, 114(44), 11645–11650. https://doi.org/10.1073/pnas.1710465114

Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. Journal of Forestry Research, 32(1), 1–6. https://doi.org/10.1007/s11676-020-01155-1

Kosmowski, F., Stevenson, J., Campbell, J., Ambel, A., & Haile Tsegay, A. (2017). On the Ground or in the Air? A Methodological Experiment on Crop Residue Cover Measurement in Ethiopia. Environmental Management, 60(4), 705–716. https://doi.org/10.1007/s00267-017-0898-0

KSS & ISRIC. (2007). Soil and terrain database for Kenya (KENSOTER), version 2.0. https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/73e27136-9efe-49e4-af35-fd98b841d467

Lal, R. (2004a). Soil carbon sequestration impacts on global climate change and food security. Science, 304(5677), 1623–1627. https://doi.org/10.1126/science.1097396

Lal, R. (2004b). Soil carbon sequestration to mitigate climate change. Geoderma, 123(1), 1–22. https://doi.org/10.1016/j.geoderma.2004.01.032

Lal, R. (2013). Soil carbon management and climate change. Carbon Management, 4(4), 439–462. https://doi.org/10.4155/cmt.13.31

Lal, R., Smith, P., Jungkunst, H. F., Mitsch, W. J., Lehmann, J., Nair, P. K. R., McBratney, A. B., de Moraes Sá, J. C., Schneider, J., & Zinn, Y. L. (2018). The carbon sequestration potential of terrestrial ecosystems. Journal of Soil and Water Conservation, 73(6), 145A-152A. https://doi.org/10.2489/jswc.73.6.145A

Liu, Y., Rao, P., Zhou, W., Singh, B., Srivastava, A. K., Poonia, S. P., Van Berkel, D., & Jain, M. (2022). Using Sentinel-1, Sentinel-2, and Planet satellite data to map field-level tillage practices in smallholder systems. PloS One, 17(11), e0277425. https://doi.org/10.1371/journal.pone.0277425

Mantey, V., Birner, R., Birkenberg, A., Bosch, C., Yameogo, V. G., & Mburu, J. (2024). Governance Challenges of Smallholder Agricultural Carbon Projects and the Potential of Digital Tools: Insights from Kenya. https://dx.doi.org/10.2139/ssrn.5266412

Ogle, S. M., BREIDT, F. J., EASTER, M., WILLIAMS, S., KILLIAN, K., & PAUSTIAN, K. (2010). Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. Global Change Biology, 16(2), 810–822. https://doi.org/10.1111/j.1365-2486.2009.01951.x

Oldfield, E. E., Eagle, A. J., Rubin, R. L., Rudek, J., Sanderman, J., & Gordon, D. R. (2022). Crediting agricultural soil carbon sequestration. Science, 375(6586), 1222–1225. https://doi.org/10.1126/science.abl7991

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., & Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology & Evolution, 20(9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011

Potash, E., Bradford, M. A., Oldfield, E. E., & Guan, K. (2025). Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale. Environmental Research Letters, 20(2), 024025. https://doi.org/10.1088/1748-9326/ada16c

Raina, N., Zavalloni, M., & Viaggi, D. (2024). Incentive mechanisms of carbon farming contracts: A systematic mapping study. Journal of Environmental Management, 352, 120126–120126. https://doi.org/10.1016/j.jenvman.2024.120126

Rikimaru, A., Roy, P. S., & Miyatake, S. (2002). Tropical forest cover density mapping. Tropical Ecology, 43(1), 39–47.

Schilling, F., Beesigamukama, D., & Tanga, C. M. (2026a). Reliability of approaches for measuring soil organic carbon and implications for results-based payments for smallholder carbon farming. Journal of Environmental Management, 398, 128562. https://doi.org/10.1016/j.jenvman.2026.128562

Schilling, F. (2026b). Funding models for scaling soil health: Assessing the economic viability and inclusiveness of carbon farming projects in Kenya. Land Use Policy, 165, 107996. https://doi.org/10.1016/j.landusepol.2026.107996

Tonitto, C., Woodbury, P. B., & McLellan, E. L. (2018). Defining a best practice methodology for modeling the environmental performance of agriculture. Environmental Science & Policy, 87, 64–73. https://doi.org/10.1016/j.envsci.2018.04.009

von Braun, J., Afsana, K., Fresco, L. O., & Hassan, M. (2021). Food systems: Seven priorities to end hunger and protect the planet. Nature, 597, 28–30. https://doi.org/10.1007/978-3-031-15703-5_1

Xue, J., Zhang, X., Huang, Y., Chen, S., Dai, L., Chen, X., Yu, Q., Ye, S., & Shi, Z. (2024). A two-dimensional bare soil separation framework using multi-temporal Sentinel-2 images across China. International Journal of Applied Earth Observation and Geoinformation, 134, 104181. https://doi.org/10.1016/j.jag.2024.104181

Zheng, B., Campbell, J. B., Serbin, G., & Galbraith, J. M. (2014). Remote sensing of crop residue and tillage practices: Present capabilities and future prospects. Soil and Tillage Research, 138, 26–34. https://doi.org/10.1016/j.still.2013.12.009

Zhou, Y., Ferdinand, M. S., van Wesemael, J., Dvorakova, K., Baret, P. V., Van Oost, K., & van Wesemael, B. (2025). A framework for mapping conservation agricultural fields using optical and radar time series imagery. Remote Sensing of Environment, 328, 114858. https://doi.org/10.1016/j.rse.2025.114858

zef
Center for
Development Research
University of Bonn

Working Paper Series

Author: Friederike Schilling

Contact: Friederike Schilling, Center for Development Research (ZEF), fschilling@uni-bonn.de

Image credit: Monja Šebela

Published by:
Zentrum für Entwicklungsforschung (ZEF)
Center for Development Research
Genscherallee 3
D – 53113 Bonn
Germany

Phone: +49-228-73-1861
Fax: +49-228-73-1869
E-Mail: presse.zef@uni-bonn.de

www.zef.de