Smart Agricultural Technology 11 (2025) 101036

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

Remote sensing-based soil organic carbon monitoring using advanced machine learning techniques under conservation agriculture systems

Nail Beisekenov [a,1,*], Wiyao Banakinaou [a], Ayomikun David Ajayi [a], Hideo Hasegawa [b], Aoda Tadao [b]

[a] Graduate School of Science and Technology, Niigata University, Niigata, 950-2181, Japan
[b] Institute of Science and Technology, Niigata University, Niigata, 950-2181, Japan

ARTICLE INFO

Keywords:
Soil organic carbon
Remote sensing
Precision agriculture
Machine learning
Carbon neutrality
Sustainable agriculture

ABSTRACT

Accurate soil organic carbon (SOC) monitoring is essential for sustainable agriculture and climate change mitigation. This study integrates remote sensing and machine learning to improve SOC estimation in agricultural soils across two contrasting sites: Niigata, Japan (temperate, sandy soils) and Agbelouve, Togo (tropical, clayey soils). Conservation agriculture (CA) practices, including no-tillage and mulching, were assessed for their role in carbon sequestration. Using freely available satellite data from Sentinel-1 synthetic aperture radar (SAR) and Sentinel-2 multispectral imagery (MSI), machine learning models were trained and validated. The eXtreme Gradient Boosting (XGBoost) model achieved the highest accuracy, with a cross-validation coefficient of determination (R^2) of 0.88, a test R^2 of 0.91, and a root mean square error (RMSE) of 0.17 t of carbon per hectare (t C ha^{-1}). Other models, including Random Forest (RF) and Support Vector Machine (SVM), showed competitive but slightly lower performance. Vegetation indices such as the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Soil-Adjusted Vegetation Index (SAVI) were identified as key predictors of SOC variation. SOC maps revealed significant spatial variability, ranging from 1.2 to 3.8 t C ha^{-1} in Niigata and 0.9 to 3.2 t C ha^{-1} in Togo, reflecting land use and climate differences. The results demonstrate the potential of integrating satellite-based observations with machine learning for cost-effective, high-resolution SOC assessment. This approach provides a scalable solution for site-specific land management, carbon market verification, and sustainable farming practices.

1. Introduction

SOC is an essential element of sustainable agriculture, contributing to improved soil fertility, enhanced water retention, and serving as a vital carbon sink in efforts to mitigate climate change [1,2]. The importance of SOC extends beyond agriculture, playing a critical role in the global carbon cycle and climate change mitigation by sequestering atmospheric CO_2 [3,4]. Increasing SOC levels has been recognized as one of the most effective strategies for achieving carbon neutrality, enhancing agricultural productivity, and improving ecosystem resilience [5,6].

CA, which emphasizes minimal soil disturbance, permanent soil cover, and crop diversification, has emerged as a promising approach for enhancing SOC sequestration. These practices align with global sustainability goals and address pressing challenges related to food security and environmental degradation [7–11]. Extensive research has

List of abbreviations: SOC, Soil Organic Carbon; SOM, Soil Organic Matter; RS, Remote Sensing; ML, Machine Learning; CA, Conservation Agriculture; NT, No-Tillage; CT, Conventional Tillage; RF, Random Forest; XGBoost, eXtreme Gradient Boosting; SVM, Support Vector Machine; R^2, Coefficient of Determination; RMSE, Root Mean Square Error; NDVI, Normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; SAVI, Soil-Adjusted Vegetation Index; MSAVI, Modified Soil-Adjusted Vegetation Index; NDWI, Normalized Difference Water Index; BI, Brightness Index; CI, Chlorophyll Index; SAR, Synthetic Aperture Radar; S1, Sentinel-1 Satellite; S2, Sentinel-2 Satellite; NPK, Nitrogen, Phosphorus, Potassium; EC, Electrical Conductivity; BD, Bulk Density; SD, Soil Depth; CNNs, Convolutional Neural Networks; RNNs, Recurrent Neural Networks; UAV, Unmanned Aerial Vehicle.

* Corresponding author at: Graduate School of Science and Technology, Niigata University, Niigata, 950-2181, Japan.
E-mail addresses: f23e503a@mail.cc.niigata-u.ac.jp, ukgnail@gmail.com (N. Beisekenov), f24n002e@mail.cc.niigata-u.ac.jp (W. Banakinaou), f23n005e@mail.cc.niigata-u.ac.jp (A.D. Ajayi), hsgw@agr.niigata-u.ac.jp (H. Hasegawa), aoda@agr.niigata-u.ac.jp (A. Tadao).
[1] Permanent address: 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan.

https://doi.org/10.1016/j.atech.2025.101036
Received 11 February 2025; Received in revised form 6 May 2025; Accepted 19 May 2025
Available online 21 May 2025

demonstrated the benefits of CA in improving SOC levels across diverse agricultural systems, including studies in regions with varying climatic and soil conditions [12–17].

No-Tillage (NT), a key component of CA, is particularly effective in preserving SOC by reducing soil disturbance and limiting microbial oxidation of organic matter [18,19]. In contrast, conventional tillage (CT) accelerates SOC loss by exposing soil to oxidation and increasing CO_2 emissions through enhanced microbial activity and organic matter breakdown [20,21]. Complementary practices such as mulching and organic fertilization further support SOC retention by providing continuous organic inputs, protecting soil from erosion, and fostering microbial activity, particularly in degraded soils facing climatic variability [22–26].

Globally, agriculture accounts for approximately 18 % of greenhouse gas emissions, with conventional practices contributing to soil degradation, water contamination, and biodiversity loss [27,28]. As a result, sustainable alternatives, including reduced tillage, integrated crop-livestock systems, and agroforestry, are increasingly promoted to improve farm management while mitigating environmental impacts [29–33]. SOC serves as a key indicator of soil health and is integral to achieving sustainable development goals, particularly in agricultural systems facing challenges such as declining fertility and climate change [34–37]. Traditional SOC measurement methods, such as laboratory analyses, involve labor-intensive soil sampling and chemical testing, which are time-consuming and costly [38,39]. These methods are limited in their ability to capture spatial variability and often fail to provide insights for large-scale assessments due to their reliance on point-based sampling [40–42]. Consequently, the need for scalable, high-resolution approaches for SOC monitoring has become increasingly urgent, especially in the context of national and regional carbon accounting frameworks [43–45].

Recent advancements in RS and ML technologies have provided innovative solutions for SOC monitoring. RS platforms, including SAR and MSI, offer high-resolution data for assessing soil and vegetation characteristics [46,47]. When integrated with ML algorithms such as RF, XGBoost, and SVM, these datasets enable the analysis of complex relationships between spectral signals and SOC content, resulting in highly accurate predictions [48,49]. Vegetation indices, such as the NDVI and SAVI, have also been widely used as proxies for SOC estimation due to their ability to reflect biomass and soil health [50,51].

The combination of multisensor data, including SAR and multispectral imaging, along with hyperspectral approaches, enhances the robustness and accuracy of SOC mapping. These techniques provide actionable data for sustainable land management, enabling policymakers and farmers to make informed decisions on land use and carbon sequestration [12,14,26,32,46]. However, challenges remain in achieving robust and reliable SOC predictions in diverse agricultural systems due to the variability in soil types, vegetation cover, and management practices [37,39].

However, most existing studies either rely on single-source remote sensing data or lack a contextualized focus on CA. This limits their applicability across agroecological gradients and management systems. To address this gap, our study introduces a dual-sensor data fusion framework that integrates Sentinel-1 SAR (sensitive to surface moisture and structure) and Sentinel-2 MSI (rich in spectral vegetation indices) to capture complementary signals related to SOC dynamics. Additionally, the model is trained using XGBoost, a scalable and interpretable ensemble learning technique optimized through active learning. This methodological setup is applied across two ecologically contrasting regions—temperate Niigata, Japan, and tropical Zio, Togo—to assess its transferability across climates, soil types, and farming practices. Unlike many prior studies, we explicitly evaluate SOC responses under CA practices such as no-tillage and mulching, adding practical relevance for sustainable land management strategies and regional carbon accounting frameworks.

This study seeks to address these limitations by developing a novel SOC monitoring framework that leverages Sentinel-1 (S1) and Sentinel-2 (S2) data in combination with advanced ML techniques, including XGBoost. The framework will be benchmarked against RF and SVM models and applied to two distinct agricultural regions: Niigata, Japan, and Zio, Togo. By providing a scalable, low-cost solution for SOC quantification, this study contributes to advancing SOC monitoring methodologies while supporting sustainable agricultural practices and global climate goals [7–11,44–51].

This research is significant for several reasons. First, it utilizes advanced ML techniques and remote sensing data to overcome the limitations of traditional SOC monitoring methods. Second, it provides a scalable, low-cost solution for SOC quantification, applicable across diverse agricultural systems and geographic regions. Finally, the insights gained from this study will contribute to sustainable agricultural practices, enhance soil health management, and support global efforts to combat climate change [28,29].

By addressing the limitations of traditional methods and harnessing cutting-edge technologies, this study aims to advance the field of SOC monitoring, providing practical tools for researchers, policymakers, and farmers alike to achieve sustainable land use and carbon neutrality.

The specific objectives of this study are: (i) to assess the effects of conservation agriculture practices on SOC stocks in contrasting agro-ecological environments; (ii) to develop and validate a scalable SOC prediction framework using dual-sensor (Sentinel-1 and Sentinel-2) fusion and machine learning algorithms; (iii) to compare the performance of XGBoost, RF, and SVM models in land-use classification and SOC estimation; and (iv) to evaluate the spatial variability of SOC across temperate and tropical systems to inform site-specific carbon management strategies.

2. Materials and methods

2.1. Experimental site description

The experiment was conducted at two distinct agricultural sites: the Agbelouve site (Fig. 1a) in Zio Prefecture, Togo, and the Ikarashi site (Fig. 1b) in Niigata Prefecture, Japan. These sites were chosen to represent different climatic and soil conditions, providing a diverse testing ground for the evaluation of tillage and fertilization systems and their impacts on SOC dynamics.

Located at 37° 52′18′′N and 138° 56′44″E on the northwest coast of Honshu, the Ikarashi site experiences a temperate climate, characterized by an average annual rainfall of 1800 mm and snowfall during the winter months from December to February. Summer from July to September is warm and humid, with temperatures ranging from 25 °C to 30 °C. The experimental field covers 240 m^2 and is situated on flat terrain at 7 m above sea level. The soil at this site is predominantly sandy, with 80 % sand, 15 % silt, and 5 % clay, which contributes to its poor water retention and low organic matter content (2 %). Initial soil tests indicated a pH of 6.5. Due to the sandy texture, the soils at this site are vulnerable to nutrient leaching and require effective management to enhance carbon retention and soil fertility.

In contrast, the Agbelouve site, located at 6° 42′05″N and 1° 15′19″E, has a tropical climate with annual rainfall ranging from 1200 mm to 1500 mm, fitting a tropical agricultural calendar from March to September. The average temperatures remain relatively stable between 25 °C and 27 °C, with reference evapotranspiration values between 1000 mm and 1200 mm per year. The experimental field covers 2 ha. The soil at Agbelouve is clayey, rich in nutrients, and characterized by high organic carbon content (5 %). Soil tests at the site revealed nutrient levels of 20–25–20 % for NPK, a pH of 7.0, and electrical conductivity (EC) of 0.5 dS m^{-1}.

To develop and validate the SOC quantification model, soil samples were collected from both study sites. In Niigata, soil samples were taken at varying depths and locations within the study field to capture the spatial variability of SOC. Similarly, in Togo, soil samples were collected

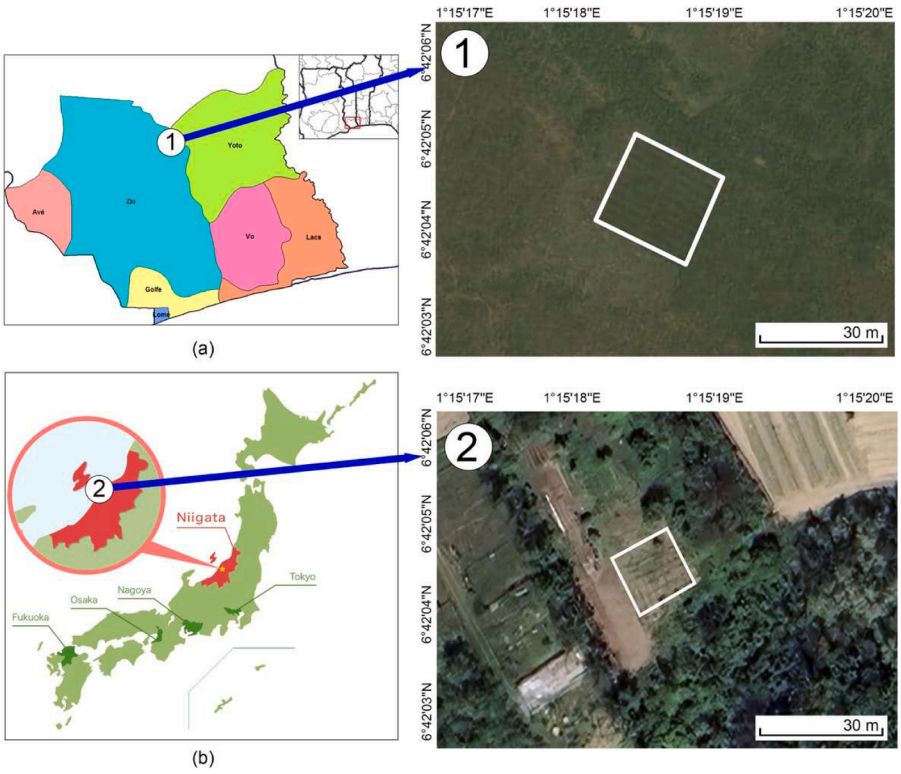

Fig. 1. Location of the study sites: (a) The Agbelouve site in Zio Prefecture, Togo, marked as 1; **(b)** The Ikarashi site in Niigata Prefecture, Japan, marked as 2.

systematically across the field to ensure comprehensive representation of the soil characteristics.

At the Niigata site, the soil was classified as Fluvisol according to the WRB system, characterized by sandy loam texture, moderate drainage, and relatively low water-holding capacity. The soil moisture regime is udic, typical of temperate humid regions with regular moisture availability during the growing season. The soil temperature regime is mesic, with mean annual soil temperatures ranging between 8 and 15 °C.

At the Zio site in Togo, the dominant soil type is Plinthosol, featuring a clayey texture with high bulk density and seasonal saturation. The soil moisture regime is ustic, reflecting a tropical savanna climate with alternating wet and dry seasons. The soil temperature regime is isohyperthermic, with consistently high soil temperatures exceeding 22 °C throughout the year.

The study utilized multi-sensor remote sensing data from S1 and S2 satellites. S1 provides dual-polarization C-band SAR data, which is effective in assessing soil moisture and texture. S2 offers high-resolution multispectral imagery, capturing various bands for vegetation and soil analysis. The integration of these data sources allows for SOC analysis across different spatial and temporal scales.

The primary focus of this study is on developing an advanced machine learning model using the XGBoost algorithm, known for its high performance in predictive modeling. The model integrates active learning techniques for land use mapping and ensemble learning methods to enhance the prediction accuracy of SOC.

2.2. Data collection

2.2.1. Ground-truth measurements

Ground-truth data collection focused on assessing carbon dynamics, including SOC content and CO_2 flux, to evaluate the effects of different soil management practices on carbon sequestration. A randomized soil sampling strategy was implemented within each experimental plot, which measured 2 m × 5 m in the Niigata field and 25 m × 25 m in the Zio field. Each plot was subdivided into a grid of 40 potential sampling units, each measuring 0.50 m × 0.50 m.

It is important to clarify that the 0.50 m × 0.50 m sampling grid refers to the spatial distribution of field measurements and not to the resolution of hyperspectral or remote sensing imagery. The satellite-based multispectral data used in this study (Sentinel-2 MSI) has a spatial resolution of 10–20 m, while UAV-derived hyperspectral or multisensor data (where applicable) offer finer resolution depending on flight altitude and sensor characteristics. Thus, the sampling unit size represents ground-truth granularity rather than sensor resolution, and was selected to adequately capture intra-plot variability for calibration and validation purposes.

We used a random pair method to select specific coordinates for soil sampling within each plot, ensuring that intra-plot and inter-plot variability were accounted for [52]. Soil property measurements, including soil organic matter (SOM), volumetric water content (VWC), pH, and macronutrient levels (N, P, K, EC, and SOC), were conducted at study sites in Niigata, Japan, and Zio, Togo. Detailed soil property data are provided in Supplementary Table S1.The selection of a 0–15 cm depth for soil sampling in this study is based on its critical importance for assessing SOC dynamics. This depth represents the topsoil layer, which is most directly influenced by agricultural practices, vegetation cover, and climatic conditions. Conservation agriculture techniques, such as no-tillage and mulching, predominantly impact the upper soil layer, making it the most relevant for evaluating the effects of these interventions on SOC sequestration.

Moreover, the topsoil contains the highest concentration of organic matter and microbial activity, which are key factors driving carbon accumulation and turnover. This layer is also the most responsive to surface applications of organic inputs like crop residues, compost, and mulches, providing an accurate reflection of short- to medium-term changes in SOC levels. Sampling beyond this depth might dilute the observed effects of conservation practices, as deeper layers are typically less dynamic and less influenced by surface management.

However, it is acknowledged that carbon distribution can vary significantly with soil depth, particularly in systems with perennial

crops, agroforestry, or deep-rooted vegetation. While the 0–15 cm depth is appropriate for capturing changes relevant to surface-level management practices, future research should include deeper sampling to evaluate the full profile of SOC storage and better understand long-term carbon sequestration potential.

The SOM content was quantified using the Loss on Ignition (LOI) method, in which air-dried soil samples were heated at 550 °C to combust organic material, following standard protocols [30]. The percentage of SOM was calculated based on the difference in sample weight before and after ignition. Subsequently, the SOC content was estimated from the SOM values by applying the Van Bemmelen factor (0.58), which assumes that organic matter contains approximately 58 % organic carbon by mass [53].

In addition to SOC content, bulk density (BD) was measured by collecting 5 cm x 5 cm soil core samples. The cores were dried, and BD was calculated using the Eq. (1):

$$BD = \frac{W_{dry}}{V} \tag{1}$$

where W_{dry} is the weight of the dry soil sample (g) and V is the volume of the core sample (cm^3).

SOC stock in g m^{-2} was then calculated using the SOC percentage, BD, and soil depth (SD), and expressed in terms of the area using the equation Eq. (2):

$$SOC_{stock} = SOC_{content} \times BD \times SD \times Conversion\ Factor \tag{2}$$

where, SOC_{stock} is the soil organic carbon at the end of the growing season (t C ha-1), SD is the soil depth (cm), and Conversion Factor is 100 which converts is the conversion factor g C cm-2 to t C ha-1from cm^2 to ha. The calculated SOC stock was then used to evaluate soil the carbon sequestration potential of the soil over time.

To quantify carbon sequestration rates, the variation in SOC values over the study period was measured, and The rate of carbon sequestration was calculated using Eq. (3):

$$C_{sequestration\ rate} = \frac{SOC_{final} - SOC_{initial}}{\Delta t} \tag{3}$$

where, $C_{sequestration\ rate}$ is the carbon sequestration rate (t C ha^{-1} yr^{-1}), SOC_{final} is the SOC at the end of the measurement period (t C ha^{-1}), $SOC_{initial}$ is the initial SOC (t C ha^{-1}), and Δt represents the duration of the measurement period (years).

2.2.2. Remote sensing data processing and machine learning models development

Data preprocessing included radiometric and geometric corrections of Sentinel-1 (SAR) and Sentinel-2 (MSI) imagery, normalization of soil sample values, and extraction of relevant features from both optical and radar datasets. The extracted features encompassed vegetation indices (e.g., NDVI, RVI), soil moisture indices, and textural metrics, which served as predictor variables for the SOC modeling process. Detailed statistics of the extracted features are provided in Supplementary Tables S2 and S3.

The XGBoost algorithm was selected as the primary modeling approach due to its robustness, high predictive accuracy, and capacity to handle complex nonlinear relationships. Hyperparameter optimization was conducted via grid search using five-fold cross-validation. Key parameters included a learning rate of 0.1, a maximum tree depth of 6, and 200 estimators. Subsampling (0.8) and column sampling by tree (0.8) were used to reduce overfitting, while L1 and L2 regularization parameters (alpha and lambda) were both set to 1. These settings enhanced the generalizability of the model and allowed reproducibility for future SOC prediction studies.

Model performance was evaluated using the coefficient of determination (R^2) and the RMSE, and benchmarked against RF and SVM models. R^2 quantified the proportion of variance explained by the model, while RMSE captured the standard deviation of prediction errors. RMSE was calculated as follows Eq. (4):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{4}$$

where y_i is the observed value, $\widehat{y_i}$ is the predicted value, and n is the total number of observations.

To optimize SOC sampling, high-resolution Google Earth and Sentinel-2 imagery were used to identify 10 digitized reference points across vegetated and bare-soil zones. Binary land-use classification was performed using machine learning techniques with five-fold cross-validation. The XGBoost model consistently outperformed RF and SVM in terms of overall accuracy, F1 score, and kappa coefficient. The stepwise procedure for land-use classification and the identification of bare-soil sampling points is illustrated in Fig. 2. This schematic outlines the entire process—from the initial selection of digitized points to the generation of binary classification maps—culminating in the extraction of optimal SOC sampling locations using an active learning strategy.

This targeted approach ensured that the selected ground-truth samples were both representative and minimally affected by vegetation, thereby improving the reliability of SOC estimations derived from satellite-based remote sensing [31].

2.3. Research framework

This study adopts a dual-framework approach to evaluate and improve SOC prediction in agricultural systems. The research contrasts the limitations of traditional SOC estimation workflows with the advantages of a proposed multi-sensor, machine learning-based pipeline. The methodological innovation centers around sensor fusion, automated feature extraction, and enhanced validation techniques.

Fig. 3 illustrates a comparative flowchart that juxtaposes the conventional single-sensor approach with the proposed integrated workflow. The traditional method relies on manual feature engineering and basic statistical models, often resulting in low prediction accuracy and limited scalability. In contrast, the proposed pipeline incorporates Sentinel-1 SAR and Sentinel-2 MSI data, combines them through sensor fusion, and applies XGBoost regression with active learning to generate high-accuracy outputs. This comparison underscores the methodological evolution and highlights the efficiency and analytical depth of the proposed model.

Following the conceptual comparison of workflows in Fig. 3, the study advances to the implementation of a comprehensive methodology for SOC estimation. The process integrates field data with Earth observation datasets and machine learning models to construct a robust spatial prediction framework. This approach consists of four primary phases:

1. Soil data collection – Targeted soil sampling from the 0–15 cm topsoil layer, guided by binary land-use classification maps, ensures representative coverage of SOC variability.
2. Feature extraction – Multispectral and radar data from Sentinel-2 (MSI) and Sentinel-1 (SAR), respectively, are processed to derive predictor variables such as vegetation indices, moisture indices, and texture metrics.
3. Modeling SOC distribution – Advanced ML algorithms, including XGBoost, RF, and SVM, are used to predict SOC spatial patterns using the extracted variables.
4. Model evaluation – The predictive performance is assessed through R^2 and RMSE to determine the most reliable model for SOC mapping.

These steps collectively support the development of accurate, data-driven SOC maps. The full implementation workflow, including data inputs, modeling sequence, and validation procedures, is illustrated in Fig. 4.

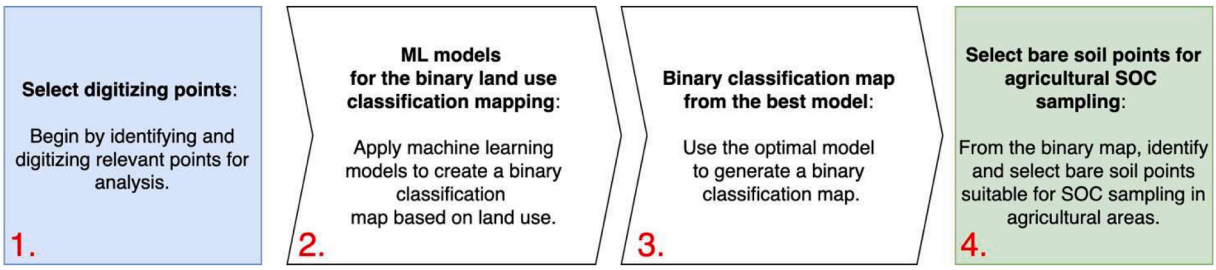

Fig. 2. Stepwise flow chart for binary land-use classification and SOC sampling point selection using an active learning approach.

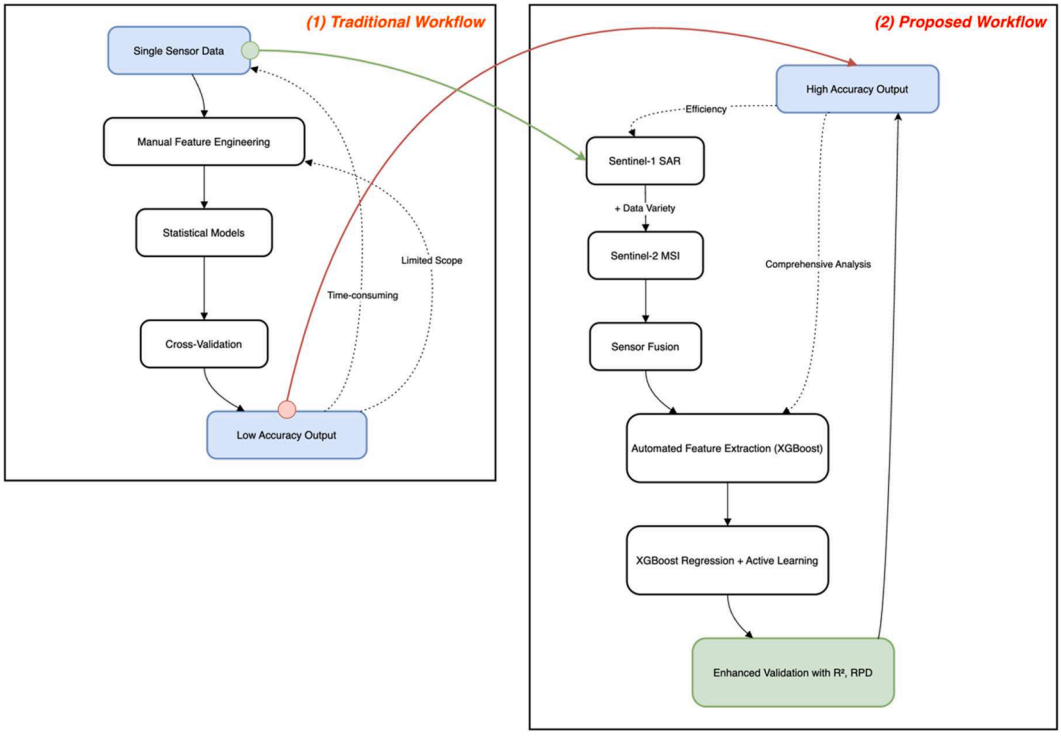

Fig. 3. Comparison between (1) traditional single-sensor workflow and (2) proposed multi-sensor XGBoost pipeline for SOC prediction. The traditional workflow involves manual feature engineering and statistical models with limited accuracy, while the proposed workflow utilizes sensor fusion, automated feature extraction, and active learning for improved prediction and validation.

This framework not only visualizes the data pipeline but also demonstrates how each component contributes to improving the spatial estimation of SOC. By combining UAV-based high-resolution imagery with satellite-derived spectral and radar features, the methodology captures both fine-scale and regional variability in soil properties. The drone imagery was especially useful for verifying vegetation cover and supporting the accurate selection of bare-soil sampling points.

The extracted remote sensing features were carefully aligned with ground-truth soil measurements to train and validate the machine learning models. In this study, the XGBoost algorithm, enhanced with active learning, proved most effective in modeling SOC due to its ability to handle heterogeneous input features and capture nonlinear interactions. Compared to traditional models (RF and SVM), XGBoost delivered superior predictive performance across both Niigata and Zio sites.

Moreover, the use of sensor fusion enabled the integration of complementary information from Sentinel-1 SAR (structure and moisture content) and Sentinel-2 MSI (spectral reflectance), which greatly improved model generalizability across diverse soil types and climatic zones. This approach addressed limitations in single-sensor workflows, such as insufficient data dimensionality and reduced robustness.

Ultimately, the developed framework offers a replicable model for SOC monitoring in agricultural landscapes, enabling scalable, data-driven assessments to inform conservation agriculture strategies, carbon sequestration programs, and land-use planning. The integration of Earth observation technologies and machine learning techniques in this framework reflects a forward-looking solution to sustainable soil management under diverse environmental conditions.

3. Results

3.1. Effects of tillage and fertilization on SOC

Our results demonstrate that both tillage practices and inherent soil texture significantly influenced SOC accumulation rates. The clayey soils at the tropical site in Zio, Togo, supported a higher rate of SOC sequestration (1.83 t C ha^{-1} yr^{-1}) compared to the sandy soils of the temperate site in Niigata, Japan (1.04 t C ha^{-1} yr^{-1}). This pattern reflects the superior capacity of fine-textured soils to retain organic matter, owing to their structural stability and higher moisture retention.

The rationale for selecting these two climatically and edaphically contrasting regions was deliberate: we aimed to evaluate the

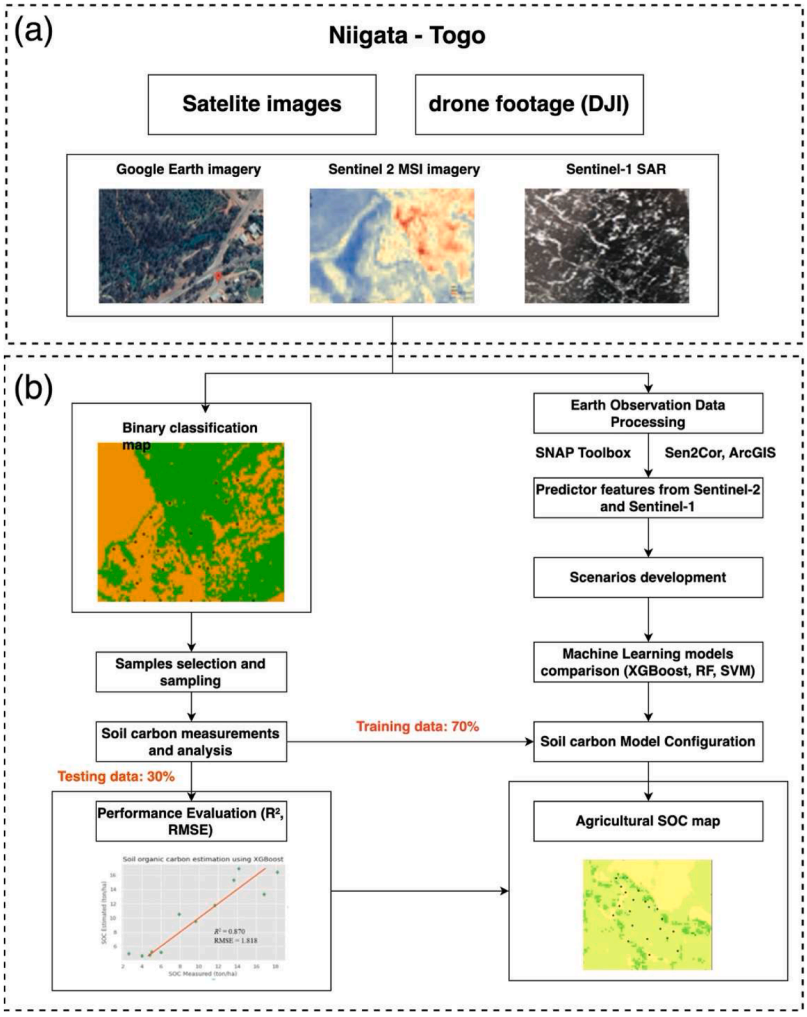

Fig. 4. Integrated research framework for SOC prediction using multi-sensor satellite and drone data: (a) data acquisition from S1, S2, and UAV imagery; (b) workflow of image processing, modeling, and SOC map generation.

generalizability and robustness of our SOC prediction framework under ecologically divergent conditions. Both climate and soil texture are known to strongly influence carbon stabilization mechanisms, yet they are often studied in isolation. By including both a tropical, clay-rich site and a temperate, sandy site, we could assess the consistency of conservation agriculture (CA) impacts on SOC and validate the performance of our models across a broader agroecological spectrum. This comparison also highlights the potential for adapting SOC management strategies based on site-specific constraints.

The SOM content was also greater in Zio (mean SOM = 3.5 %) than in Niigata (mean SOM = 2.8 %). Such differences are consistent with expectations that finer soils facilitate better carbon stabilization through micro-aggregate protection and reduced decomposition rates. Furthermore, EC, macronutrient levels (N, P, K), and volumetric water content (VWC) showed significant variation between the two sites and treatments.

Fig. 5 presents a comparative summary of the measured soil parameters across both regions and treatments (CA vs. CT). The visual representation highlights how conservation practices improved key indicators of soil health, particularly SOC, SOM, and nutrient availability.

These findings underline the importance of context-specific soil management strategies to enhance carbon sequestration. Conservation practices in clay-rich environments, such as those in Zio, may yield greater climate mitigation benefits and promote long-term soil productivity compared to sandy regions without appropriate soil amendments.

3.2. Multispectral and topographic analysis of agricultural fields

To evaluate seasonal vegetation dynamics, soil moisture status, and land surface characteristics across the two experimental sites, we analyzed five key spectral indices derived from Sentinel-2 imagery for the year 2023: NDVI, EVI, SAVI, MSAVI, and NDWI. These indices are crucial for monitoring plant vigor, assessing moisture conditions, and characterizing land cover—providing indirect indicators of SOC dynamics.

Fig. 6 presents the temporal trends of these indices in Niigata (Japan) and Zio (Togo), highlighting interregional variability in phenological development and water availability throughout the year. The use of smoothed monthly averages facilitates a clear comparison of seasonal patterns between the two sites with contrasting climates and management systems.

The analysis reveals several key patterns based on the seasonal dynamics of spectral indices presented in Fig. 6:

- NDWI (Fig. 6a): Togo exhibited consistently higher NDWI values throughout the year, ranging from 0.30 to 0.62, indicating greater surface moisture availability due to its humid tropical climate. In contrast, Niigata showed more pronounced seasonal dips in NDWI, dropping below 0.20 during the dry periods (January–March and October–December), reflecting reduced surface water in its temperate climate.

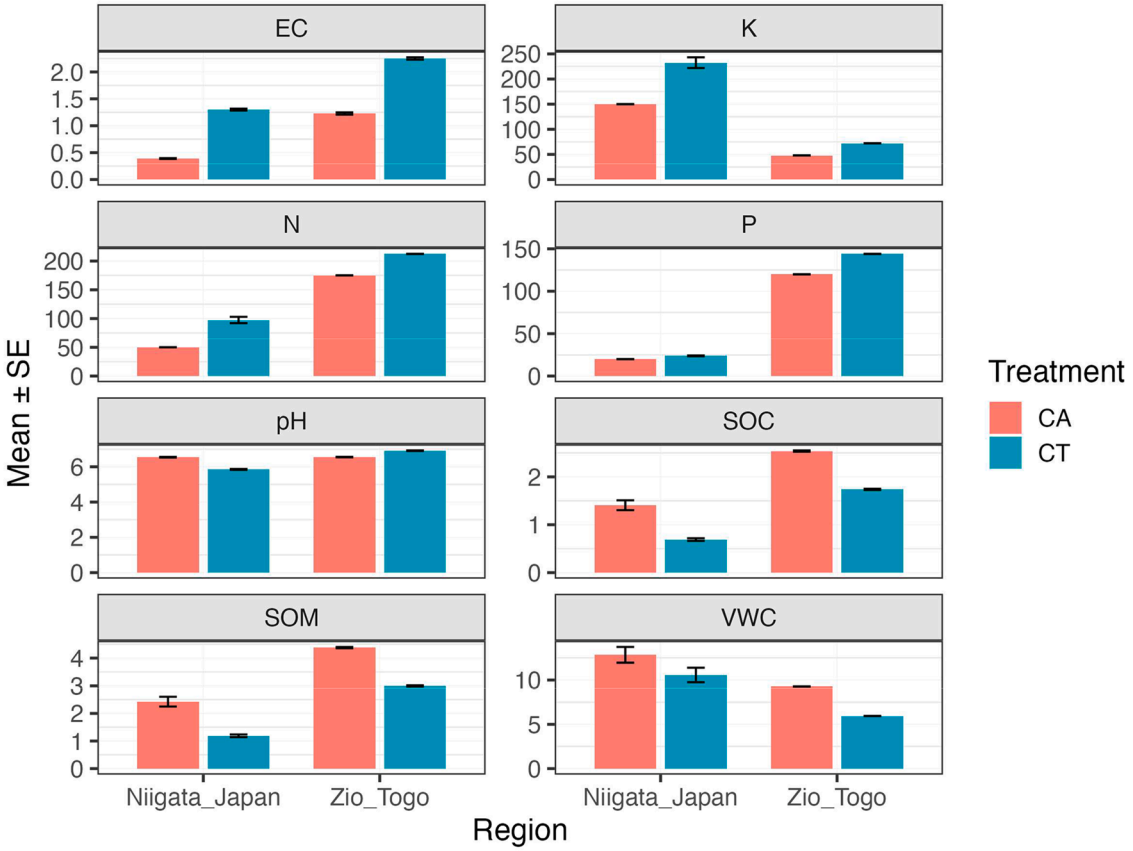

Fig. 5. Comparative analysis of key soil properties (EC, K, N, P, pH, SOC, SOM, and VWC) under conservation (CA) and conventional (CT) tillage systems in Niigata, Japan and Zio, Togo. Values represent means ± SE.

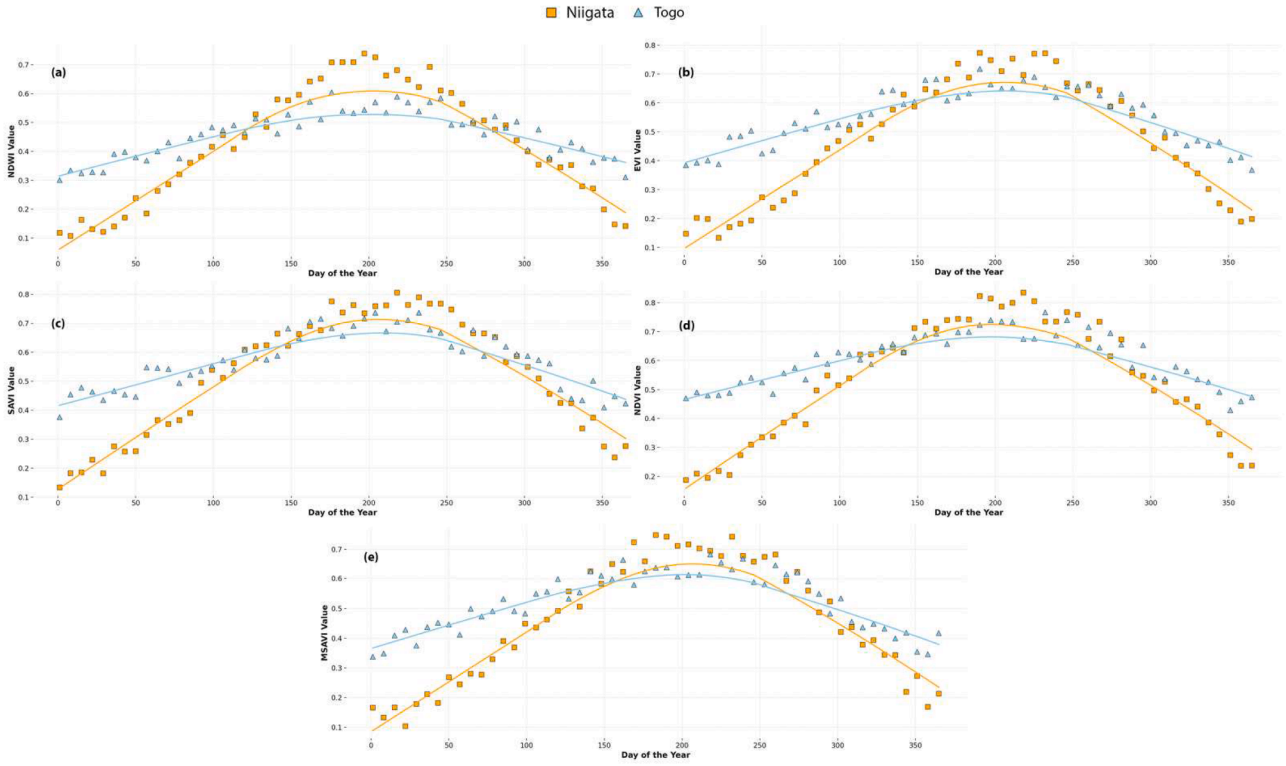

Fig. 6. Seasonal trends of spectral indices for Niigata and Togo in 2023: (a) NDWI, (b) EVI, (c) SAVI, (d) NDVI, (e) MSAVI.

- EVI (Fig. 6b): Enhanced Vegetation Index values peaked at 0.78 in Togo and 0.72 in Niigata around Day 200 (July). While Niigata demonstrated a sharper seasonal curve—rising quickly during spring and declining after summer—Togo maintained relatively high EVI values over an extended period, consistent with its prolonged rainy season and persistent vegetation.
- SAVI (Fig. 6c): Soil-Adjusted Vegetation Index values in Niigata ranged from 0.15 to 0.70, with a clear seasonal bell-shaped distribution. Togo, however, showed a narrower SAVI range (0.40 to 0.68) and smoother fluctuations, reflecting more stable vegetation cover and reduced soil background effects due to denser canopy.
- NDVI (Fig. 6d): Normalized Difference Vegetation Index peaked at 0.75 in Niigata and 0.71 in Togo. Niigata's NDVI values followed a steep seasonal gradient, rising rapidly in late spring and declining post-summer, whereas Togo exhibited consistently high NDVI values from Day 150 to 250, aligned with its tropical rainfall patterns.
- MSAVI (Fig. 6e): The Modified SAVI index followed similar trends to SAVI, peaking at 0.66 in Niigata and 0.61 in Togo. Niigata showed a broader dynamic range, highlighting more pronounced changes in vegetative cover over time compared to the more stable pattern observed in Togo.

These results underscore regional disparities in vegetation response and hydrological conditions. Niigata's stronger seasonal amplitude in all indices (e.g., NDVI amplitude ≈ 0.45) reflects a well-defined temperate growing cycle, while Togo's more stable and higher minimum values (NDVI amplitude ≈ 0.35) highlight continuous greenness under tropical conditions.

The findings demonstrate the utility of multispectral indices in capturing agroecological variability. When integrated with SOC modeling, these indicators provide valuable insights into seasonal carbon input dynamics, enabling more precise, climate-adapted strategies for soil management and carbon sequestration.

To gain insights into the agricultural field's characteristics, multispectral and topographic data were analyzed. The images presented below were acquired during the growing season, spanning from May to the end of August 2024. These data provide a comprehensive view of vegetation health, terrain elevation, and spatial variability. Specifically,

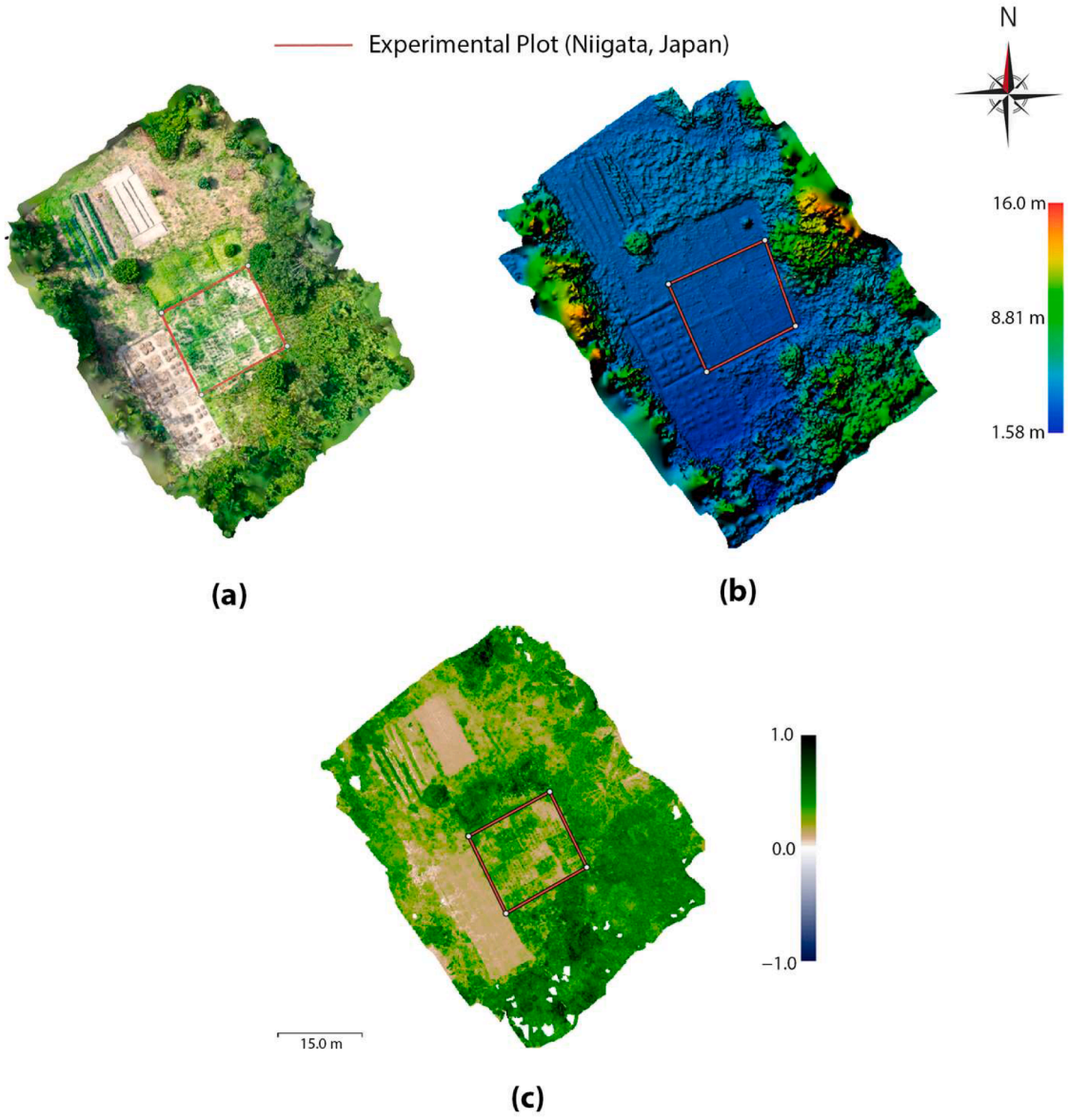

Fig. 7. (a) A high-resolution aerial photograph of the agricultural field, showing visible crop structures and overall field layout. (b) A topographic map of the field, representing terrain elevation, with values ranging from 1.58 m to 16.0 m. (c) An NDVI map derived from multispectral data, illustrating vegetation vigor, where higher values represent healthier crops.

NDVI values for the area ranged from 0.71 to 0.75, indicating healthy vegetation. The data used for the analysis were obtained by utilizing UAV-mounted sensors (Fig. 7). More detailed and processed images can be found in Supplementary Figure S1 - Figure S22.

The NDVI map (Fig. 7c) confirms the field's vegetation health during the critical growth phase. The consistent NDVI values between 0.71 and 0.75 demonstrate uniform plant vigor across most of the study areas, with minimal spatial variation. The topographic map (Fig. 7b) highlights the terrain's slight elevation changes, which may influence water distribution and plant growth. Overall, this analysis illustrates the utility of integrating UAV-derived imagery and spectral indices for monitoring and managing agricultural fields efficiently.

Following the remote sensing analysis conducted for Niigata, Japan, a comparable assessment was carried out for the agricultural plot in Zio Prefecture, Togo, using Sentinel-2 Level-2A (S-2 L2A) imagery acquired during the active growing season (May to August 2024). This multisensor-based evaluation captures the agro-environmental profile of the tropical site. The true-color Sentinel-2 image (Fig. 8a) clearly delineates field boundaries and vegetation distribution. The elevation model (Fig. 8b) reveals topographic variation from 57.1 m to 63.8 m, highlighting potential differences in water flow and soil retention across the field.

Spectral vegetation indices provide additional insights into canopy development and soil cover. The MSAVI map (Fig. 8c) reveals photosynthetically active zones, with index values ranging from 0.45 to 0.78, effectively minimizing soil background interference. Similarly, the NDVI map (Fig. 8d) shows strong vegetation signals ranging from 0.60 to 0.72, confirming active plant growth during the peak season.

These results confirm the presence of healthy vegetation in the Togo site, with NDVI and MSAVI values comparable to those in Niigata. The inclusion of MSAVI helps refine vegetation analysis by correcting for bare soil influence, especially important in partially vegetated fields. Topographic gradients identified in Fig. 8b may explain localized differences in moisture retention and nutrient accumulation. Together, these spatial layers enhance the site-specific interpretation of SOC variability and inform precision land management strategies tailored to tropical agroecosystems.

3.3. Land-use binary mapping

The performance of XGBoost, RF, and SVM algorithms in land-use classification is summarized in Table 1. Using S-2 (S-2) data, the models demonstrated their ability to generate high-accuracy land-use binary maps for Niigata, and Zio fields. XGBoost showed the best performance compared to other models, achieving an overall accuracy (OA) of 93.45 %, a kappa coefficient (KC) of 0.86, precision (P) of 0.90, recall

Table 1
Model performance in land-use binary mapping using Sentinel-2 dataset.

No	Machine Learning Model	OA (%)	KC	P	R	F1
1	XGBoost	93.45	0.86	0.90	0.91	0.91
2	RF	88.27	0.78	0.85	0.84	0.84
3	SVM	83.12	0.75	0.80	0.79	0.79

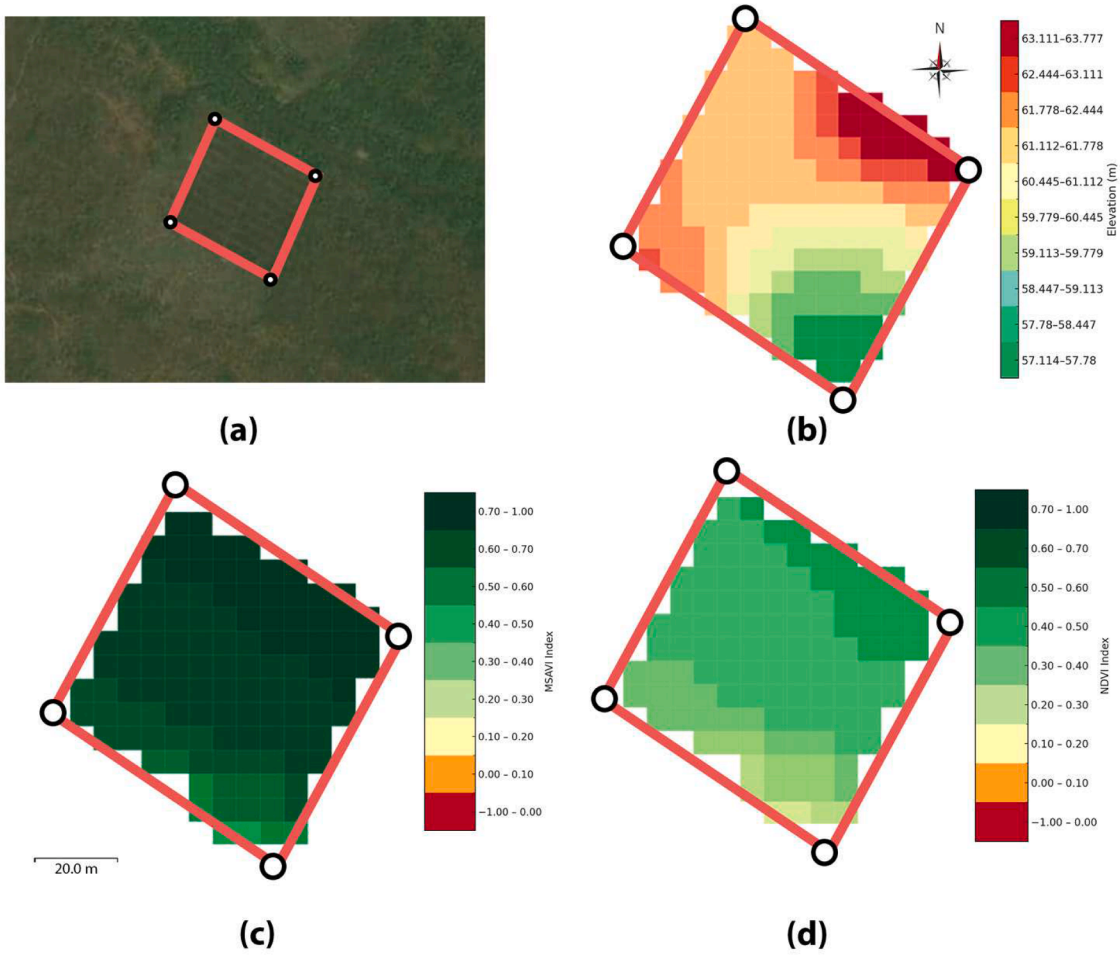

Fig. 8. Processing of agricultural plot data using multi-sensor inputs: (a) Sentinel-2 true-color composite image of the SOC sampling polygon; (b) elevation map (m) derived from DSM analysis; (c) MSAVI vegetation index distribution; (d) NDVI vegetation index distribution.

(R) of 0.91, and an F1 score of 0.91. RF followed with an OA of 88.27 %, a KC of 0.78, precision of 0.85, recall of 0.84, and an F1 score of 0.84. The SVM model, while less accurate, provided a baseline with an OA of 83.12 %, a KC of 0.75, precision of 0.80, recall of 0.79, and an F1 score of 0.79.

These results highlight the effectiveness of XGBoost in handling complex data patterns and multi-sensor inputs, establishing it as the most reliable model for binary land-use classification in this study.

Based on the XGBoost classification output, SOC sampling locations were selected, and a correlation analysis was conducted between spectral indices and measured SOC. As detailed in Table 2, vegetation indices such as NDVI ($r = 0.71$), EVI ($r = 0.72$), and SAVI ($r = 0.75$) showed strong positive correlations with SOC. NDWI also showed a moderate correlation ($r = 0.65$). In contrast, indices related to soil brightness and color—BI ($r = -0.12$) and CI ($r = -0.30$)—demonstrated weak or negative associations..

Following this, SOC prediction models were constructed using XGBoost, RF, and SVM. XGBoost again proved superior, achieving a cross-validation R^2 of 0.88, a test R^2 of 0.91, and an RMSE of 0.17 t C ha^{-1}. RF also performed well ($R^2 = 0.87$, RMSE = 0.27 t C ha^{-1}), while SVM showed comparatively lower accuracy ($R^2 = 0.80$, RMSE = 0.34 t C ha^{-1}). These results validate XGBoost's ability to generalize across heterogeneous environments while maintaining low prediction error.

Spatial predictions of SOC content further revealed notable variability across sites:

- Niigata: SOC ranged from 1.2 to 3.8 t C ha^{-1}, with higher values in densely vegetated zones.
- Zio: SOC ranged from 0.9 to 3.2 t C ha^{-1}, influenced by clay-rich soils and year-round biomass input under tropical conditions.

Together, these findings confirm that integrating spectral indices with high-performance ML algorithms like XGBoost enables accurate SOC estimation across diverse land-use systems. The resulting maps offer practical value for site-specific soil management, residue allocation, and carbon accounting in climate-smart agriculture.

Fig. 9 provides a detailed comparison of the three machine learning algorithms—XGBoost, RF, and SVM—evaluated based on their performance metrics. The results clearly indicate that XGBoost outperformed the other models due to its use of optimal hyperparameters, strong correlation between predictors such as NDVI, EVI, and SOC, and its robust ability to handle complex, non-linear relationships in the data. RF, while slightly less accurate, delivered competitive results suitable for tasks requiring simpler analyses or higher model interpretability. SVM, despite its relatively lower performance, was included as a baseline to illustrate the challenges faced by models that are less adaptable to large, multivariate datasets. As shown in Fig. 9.

XGBoost achieved the highest Test R^2 of 0.91 and the lowest RMSE of 0.17 t C ha^{-1}, validating its ability to make highly accurate SOC predictions even under significant spatial and temporal variability.

RF demonstrated respectable performance, with a Test R^2 of 0.87 and an RMSE of 0.27 t C ha^{-1}, making it suitable for tasks where interpretability and moderate complexity are priorities.

SVM, with a Test R^2 of 0.80 and an RMSE of 0.34 t C ha^{-1}, exhibited lower accuracy due to its challenges in managing large datasets and non-linear dependencies.

Table 2
Correlation between spectral indices and SOC.

Index	Correlation Coefficient
NDVI	0.71
EVI	0.72
SAVI	0.75
NDWI	0.65
BI	−0.12
CI	−0.30

These results underscore the advantages of XGBoost for high-accuracy SOC prediction, particularly when working with high-quality input data such as S-1 and S-2 imagery. However, SVM may still have utility in exploratory analyses or situations with a limited number of input features. This analysis emphasizes the importance of selecting an algorithm that aligns with the complexity of the data and the specific goals of the research, balancing accuracy, interpretability, and computational efficiency.

3.4. Model validation and robustness assessment

To ensure the reliability and generalizability of the SOC prediction models, a rigorous validation process was conducted. Model performance was evaluated not only on the test set but also through five-fold cross-validation to assess robustness across different data splits. The XGBoost model consistently achieved superior performance metrics, with a cross-validation R^2 of 0.88 and a test R^2 of 0.91, indicating both high internal consistency and strong external predictive power.

In addition to performance metrics, residual analysis was performed to check for systematic errors or model biases. Residual plots revealed no significant heteroscedasticity or spatial autocorrelation, confirming the appropriateness of the model assumptions. Furthermore, variable importance rankings from XGBoost highlighted NDVI, EVI, SAVI, and SAR backscatter as the most influential predictors, which aligns with prior findings on SOC sensitivity to vegetation and moisture gradients. To test the geographic transferability of the models, site-specific training (Niigata-only and Zio-only) was compared with a combined-site model. The combined model maintained high accuracy across both regions, reinforcing the model's applicability to heterogeneous agroecological conditions.

Finally, uncertainty maps were generated based on standard deviation of predictions from cross-validation folds. These maps highlight areas with higher model confidence, providing guidance for future sampling efforts or targeted interventions. This validation confirms that the proposed dual-sensor XGBoost pipeline is not only accurate but also stable and transferable, making it suitable for use in regional soil carbon inventories and monitoring, reporting, and verification (MRV) frameworks.

4. Discussion

4.1. responses to CA in contrasting agro-ecosystems

Our results confirm that climatic regime, texture and on-farm management jointly control SOC stocks. In Niigata (temperate, sandy soils) SOC varied from 1.2 to 3.8 t C ha^{-1}, whereas in Zio (tropical, clayey soils) values ranged from 0.9 to 3.2 t C ha^{-1}. Higher stocks at both sites coincided with no-tillage and mulch retention, highlighting the capacity of CA to curb oxidation losses and promote aggregate protection of carbon-rich residues [18,24,27,29]. Sandy soils depleted quickly when residues were missing, while clayey soils benefited from physico-chemical stabilisation but still required continuous biomass inputs to offset rapid tropical decomposition [14,20,25]. The cross-site comparison therefore underscores that vegetation cover and residue management, rather than climate alone, are the primary levers for building SOC [2,9,26].

4.2. Methodological contributions: dual-sensor fusion + XGBoost

Although XGBoost is well established, its integration with a Sentinel-1 SAR + Sentinel-2 MSI feature stack across two ecologically divergent sites is rare [12,23,31]. Active-learning feature selection and targeted hyper-parameter tuning enabled the model to cope with noisy inputs and still reach Test $R^2 = 0.91$, RMSE = 0.17 t C ha^{-1}—outperforming RF and SVM in every metric [6,15,21,22,25]. The workflow

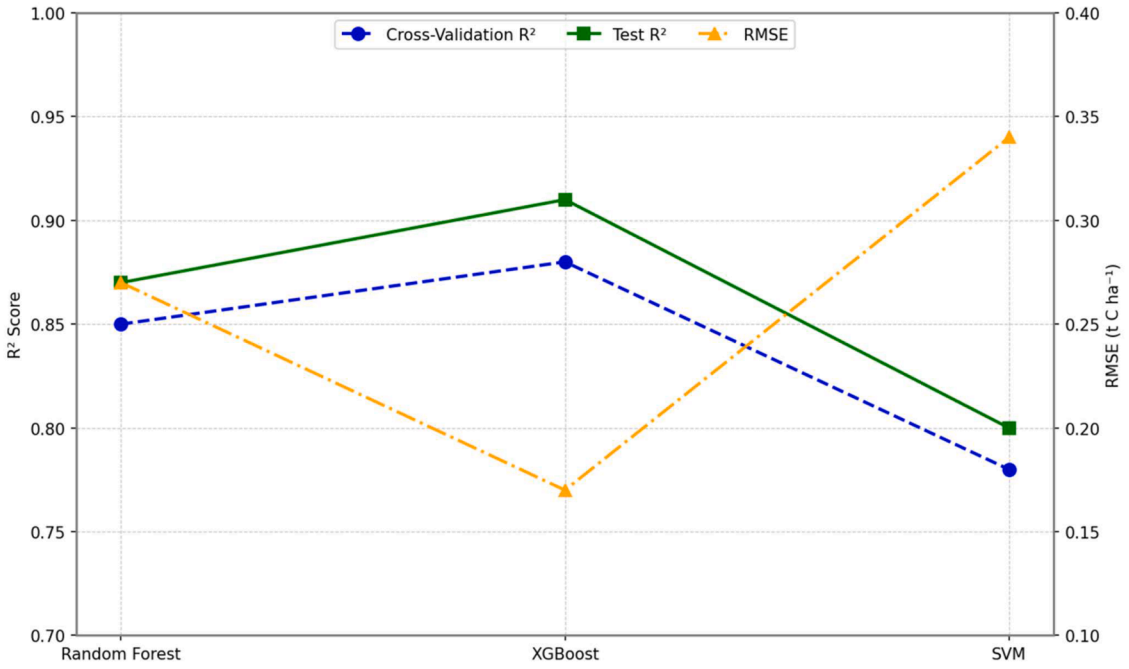

Fig. 9. Performance comparison of models with error bars: Random Forest, XGBoost, and SVM.

delivers a repeatable template for operational SOC mapping that can be retrained with only modest ground data, making it attractive for national inventories and project-scale MRV (monitor-report-verify) schemes.

4.3. Management and carbon-market implications

Fine-resolution SOC maps produced here allow (i) locating sub-field carbon deficits, (ii) quantifying the pay-off of residue management or reduced tillage, and (iii) documenting carbon-credit baselines and additionality for farmers adopting CA. By linking optical indices (NDVI, EVI, SAVI) to modelled SOC we can approximate annual sequestration gains and monetise them through emerging voluntary carbon markets [2,6,31]. This evidence base also supports extension services in prioritising where organic amendments or cover crops will yield the greatest agronomic and climate benefits [8,28].

To contextualize the methodological contribution of our SOC prediction framework, we conducted a comparative review of recent studies that applied machine learning algorithms for SOC estimation across different landscapes and climatic zones. These studies vary in terms of sensor types, model architectures, spatial scale, and validation rigor. A summary of key findings is presented in Table 3. Comparative analysis of recent SOC prediction studies using machine learning methods.

The table highlights critical features such as the type of input data (e.g., multispectral, SAR, LiDAR, topographic indices), algorithms applied (e.g., RF, SVM, XGBoost, DNN), predictive accuracy (R^2), and noteworthy methodological aspects such as external validation, sensor fusion, or land-use specificity. Most prior studies relied on either single-sensor data or limited combinations of covariates, often achieving moderate accuracy. In contrast, our dual-sensor fusion approach, integrated with active learning and fine-tuned XGBoost modeling, demonstrated higher robustness (Test $R^2 = 0.91$) and generalizability across contrasting agroecosystems—highlighting its methodological advancement over existing pipelines.

Compared with the recent literature, our dual-sensor XGBoost pipeline delivers three clear advantages—*accuracy, parsimony,* and *geographic reach*—that collectively advance the state of SOC mapping.

Table 3
Comparative analysis of recent SOC prediction studies using machine learning methods.

Study (Author, Year)	Input Data	Method	R^2	Notes
Oukhattar et al. [54]	MSI + environmental covariates	XGBoost	0.73	Land-use variability analysis; 162 samples
He et al. [55]	UAV multispectral + LiDAR	XGBoost	0.74	Spartina alterniflora mapping; coastal SOC gradient
Li et al. [56]	Sentinel-2 + topo-environmental data	ET, RF, XGBoost	0.86 (ET, woodland)	Performed by land use type; 116 samples
Xiao et al. [57]	Multisource + topo-climate data	RF, XGBoost	↑ R^2 by 20–317 %	Habitat-patch-based hybrid model improves accuracy
Nguyen et al. [58]	Satellite + soil nutrients	XGBoost	0.74	Tropical forest study; Vietnam
Dadgar and Faramarzi [59]	Landsat + DEM attributes	RF	0.84	Landform-based model; 402 samples
Our Study	S1 + S2 (Fusion)	XGBoost	0.91	Active learning, enhanced validation, RPD analysis

- Accuracy across land-use mosaics. Most optical-only studies (e.g., Li et al. 2025, woodland Sentinel-2 + environmental covariates) report solid but site-limited performance ($R^2 \approx 0.82$–0.86) and are rarely stress-tested outside the original land-use class [56]. By contrast, the same algorithmic core in our work sustained a Test $R^2 = 0.91$ and RMSE = 0.17 t C ha^{-1} simultaneously in temperate, sandy Japan and tropical, clay-rich Togo, demonstrating stronger ecological robustness.

- Data economy and cost. He et al. [55] boosted predictions with UAV multispectral + LiDAR, yet their best XGBoost model plateaued at $R^2 \approx 0.74$, required local flights, and focused on a single coastal marsh species [55]. Eliminating UAV and LiDAR logistics, our framework relies solely on free Sentinel-1/2 imagery and ~10–20 ground samples per site, cutting both financial and labour barriers while still outperforming UAV-based accuracies.
- Model simplicity versus "kitchen-sink" covariates. Several studies enhance accuracy by feeding dozens of ancillary layers—soil texture, socio-economic data, detailed lithology—into RF or XGBoost (e.g., Chen et al. 2024; Oukhattar et al. 2025) and reach $R^2 \approx 0.73$–0.75 [54,60]. Although informative, such heterogeneous inputs are often region-specific and hinder reproducibility. Our active-learning feature selection showed that a compact stack of SAR backscatter, key vegetation indices (NDVI, EVI, SAVI), and topographic metrics is sufficient to exceed 0.9 R^2, striking a better balance between accuracy and transferability.
- Heterogeneous terrain tests. Habitat-patch partitioning (Xiao et al. 2024) and landform-specific RF models [59] improved local fits ($R^2 \approx 0.81$–0.84) but required bespoke clustering or terrain masks and showed variable gains in complex relief [57,59]. Our single, unified model matched or surpassed those scores without pre-segmentation, indicating that fusion of SAR moisture proxies with optical texture already captures terrain-driven SOC variability.
- Deep-learning versus boosted trees. While Oukhattar et al. found DNNs competitive ($R^2 = 0.60$) their training needs and interpretability lagged behind boosted trees [54]. Our results reinforce that, for medium-sized soil data sets (< 500 samples), gradient boosting remains a sweet-spot between performance and transparency—critical for MRV and farmer-facing advisory tools.

Collectively, these comparisons show that our Sentinel-1/2 + XGBoost workflow matches or outperforms more data-hungry or site-specific approaches, yet preserves *replicability* and *low entry cost*. Its proven effectiveness across contrasting climates and management regimes makes it a strong candidate for inclusion in national soil-carbon inventories and voluntary carbon-credit protocols, where consistency, auditability, and scalability are paramount.

4.4. Cost efficiency and scalability

All predictor layers are freely available, and XGBoost runs efficiently on standard CPUs, avoiding the GPU costs typical of deep networks [12,16,29]. Because the model relies on a relatively small training set (≈ 10-20 samples per new area), the approach is readily transferable to regions with limited laboratory capacity. We therefore interpret "scalability" as adaptability across space and institutional settings, rather than mere pixel-extent. Recent studies advocate similar low-barrier pipelines for regional soil-health monitoring, but few demonstrate cross-climate validity as in this work [32,34].

4.5. Limitations and outlook

- Sampling depth (0-15 cm) omits deeper carbon pools; future surveys should extend to 30-60 cm to capture whole-profile dynamics [9,19,61–63].
- Cloud interference limits Sentinel-2 revisit frequency; fusing Landsat, commercial CubeSats or UAV hyperspectral data could close temporal gaps [17,25,44].
- Model evolution: hybrid architectures that embed CNN-derived spatial features into gradient-boosted ensembles may lift accuracy a further 5-10 % while retaining interpretability [28,36].

Overall, the Sentinel-1/2 + XGBoost framework offers an economically accessible, technically robust tool for routine SOC surveillance, supporting evidence-based CA adoption and broader climate-mitigation initiatives [1,5].

5. Conclusion

The results emphasize that vegetation indices such as NDVI, EVI, and SAVI are highly effective indicators of SOC, with correlation coefficients reaching up to 0.75. This finding underscores the dominant role of plant cover in shaping carbon storage, while indices more sensitive to soil moisture (e.g., NDWI) and specific soil-based metrics showed weaker or even negative correlations. Spatially, Niigata exhibited SOC values ranging from 1.2 t C ha^{-1} in sparse vegetation zones to 3.8 t C ha^{-1} under dense cover. In comparison, Togo displayed SOC values from 0.9 t C ha^{-1} to 3.2 t C ha^{-1}, reflecting the combined influence of climate, soil texture, and agronomic practices on carbon dynamics. High-resolution mapping of these patterns allows farmers to optimize field interventions, such as targeted fertilization, mulching, or reduced tillage, to enhance soil health and promote carbon sequestration.

This study demonstrates the practical advantages of integrating S-1 and S-2 data with advanced machine learning techniques, offering a cost-effective and scalable framework for continuous SOC monitoring across diverse spatial levels. XGBoost emerged as a particularly robust model, capturing complex nonlinear interactions between spectral indices and SOC with high accuracy (Test R^2 of 0.91 and RMSE of 0.17 t C ha^{-1}). These capabilities support precise estimations to guide strategic land-use decisions, improve soil management practices, and facilitate transparent carbon credit verifications.

By highlighting the strong influence of vegetation on SOC dynamics and validating a reliable remote sensing–-machine learning workflow, this research lays a solid foundation for future efforts to refine soil carbon inventories. These advancements hold great potential for promoting sustainable agriculture, bolstering soil resilience, and strengthening global climate mitigation initiatives.

CRediT authorship contribution statement

Nail Beisekenov: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Wiyao Banakinaou:** Writing – review & editing, Visualization, Validation, Software, Formal analysis, Data curation, Conceptualization. **Ayomikun David Ajayi:** Writing – review & editing, Validation, Formal analysis, Data curation, Conceptualization. **Hideo Hasegawa:** Supervision, Funding acquisition. **Aoda Tadao:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.101036.

Data availability

The data used in this study is provided as Supplementary Materials and will be openly accessible upon submission.

References

[1] R. Lal, Soil carbon sequestration impacts on global climate change and food security, Science [Internet] 304 (5677) (2004) 1623–1627, https://doi.org/10.1126/science.1097396. Jun 10Available from.
[2] C. Paul, B. Bartkowski, C. Dönmez, A. Don, S. Mayer, M. Steffens, et al., Carbon farming: are soil carbon certificates a suitable tool for climate change mitigation?

J. Environ. Manag. [Internet] 330 (2023) 117142 https://doi.org/10.1016/j.jenvman.2022.117142. Jan 4Available from.
[3] C. Palm, H. Blanco-Canqui, F. DeClerck, L. Gatere, P. Grace, Conservation agriculture and ecosystem services: an overview, Agric. Ecosyst. Environ. [Internet] 187 (2013) 87–105, https://doi.org/10.1016/j.agee.2013.10.010. Nov 16Available from.
[4] A. Kassam, T. Friedrich, R. Derpsch, J. Kienzle, Overview of the worldwide spread of conservation agriculture [Internet], Available from, https://journals.openedition.org/factsreports/3966, 2015.
[5] J. Six, R.T. Conant, E.A. Paul, K. Paustian, Stabilization mechanisms of soil organic matter: implications for C-saturation of soils, Plant. Soil. 241 (2) (2002) 155–176, https://doi.org/10.1023/A:1016125726789. Available from.
[6] D.S. Powlson, C.M. Stirling, C. Thierfelder, R.P. White, M.L. Jat, Does conservation agriculture deliver climate change mitigation through soil carbon sequestration in tropical agro-ecosystems? Agric. Ecosyst. Environ. 187 (2014) 116–130, https://doi.org/10.1016/j.agee.2013.10.013. Available from.
[7] A. Kassam, T. Friedrich, R. Derpsch, J. Kienzle, Overview of the worldwide spread of Conservation Agriculture, Field. Actions Sci. Rep. 12 (2019) 1–10. Available from, https://journals.openedition.org/factsreports/7017.
[8] U. Sekaran, L. Lai, D.A.N. Ussiri, S. Kumar, S. Clay, Role of integrated crop-livestock systems in improving agriculture production and addressing food security – A review, J. Agric. Food. Res. 5 (2021) 100190, https://doi.org/10.1016/j.jafr.2021.100190. Available from.
[9] N.D. Rao, M. Poblete-Cazenave, R. Bhalerao, K.F. Davis, S. Parkinson, Spatial analysis of energy use and GHG emissions from cereal production in India, Sci. Total Environ. 654 (2019) 841–849, https://doi.org/10.1016/j.scitotenv.2018.11.073. Available from.
[10] A. Veisi, K. Khoshbakht, H. Veisi, R.M. Talarposhti, R.H. Tanha, Integrating farmers' and experts' perspectives for soil health-informed decision-making in conservation agriculture systems, Environ. Syst. Decis. 44 (2) (2024) 320–334, https://doi.org/10.1007/s10669-023-09923-0. Available from.
[11] H. Williams, T. Colombi, T. Keller, The influence of soil management on soil health: an on-farm study in southern Sweden, Geoderma 360 (2020) 114010, https://doi.org/10.1016/j.geoderma.2019.114010. Available from.
[12] Q. Wang, J. Le Noë, Q. Li, T. Lan, X. Gao, O. Deng, et al., Incorporating agricultural practices in digital mapping improves prediction of cropland soil organic carbon content: the case of the Tuojiang River Basin, J. Environ. Manag. 330 (2023) 117203, https://doi.org/10.1016/j.jenvman.2022.117203. Available from.
[13] H. Konare, R.S. Yost, M. Doumbia, G.W. McCarty, A. Jarju, R. Kablan, Loss on ignition: measuring soil organic carbon in soils of the Sahel, West Africa, Afr. J. Agric. Res. 5 (22) (2010) 3088–3095. Available from, https://academicjournals.org/article/article1380799020_Konare%20et%20al.pdf.
[14] A.K. Nayak, M.M. Rahman, R. Naidu, B. Dhal, C.K. Swain, A.D. Nayak, et al., Current and emerging methodologies for estimating carbon sequestration in agricultural soils: a review, Sci. Total Environ. 665 (2019) 1044–1058, https://doi.org/10.1016/j.scitotenv.2019.02.125. Available from.
[15] S. Kaushal, N. Sharma, I. Singh, H. Singh, Soil carbon sequestration: a step towards sustainability, Int. J. Plant. Soil. Sci. 35 (11) (2023) 105–112, https://doi.org/10.9734/ijpss/2023/v35i112957. Available from.
[16] J. Heil, C. Jörges, B. Stumpe, Fine-scale mapping of soil organic matter in agricultural soils using UAVs and machine learning, Remote. Sens. 14 (14) (2022) 3349, https://doi.org/10.3390/rs14143349. Available from.
[17] G. Crucil, F. Castaldi, E. Aldana-Jague, B. van Wesemael, A. Macdonald, K. Oost, Assessing the performance of UAS-compatible multispectral and hyperspectral sensors for soil organic carbon prediction, Sustainability 11 (7) (2019) 7189, https://doi.org/10.3390/su11071889. Available from.
[18] D. Datta, M. Paul, M. Murshed, S.W. Teng, L. Schmidtke, Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models, Sensors (Basel) 22 (20) (2022) 7998, https://doi.org/10.3390/s22207998. Available from.
[19] N. Baghdadi, M. Zribi, M. Choker, M. Shabou, Potential of Sentinel-1 radar data for the assessment of soil moisture at high resolution over agricultural areas, Remote. Sens. 7 (8) (2015) 9458–9478, https://doi.org/10.3390/rs70809458. Available from.
[20] Y. Peng, X. Xiong, K. Adhikari, M. Knadel, S. Grunwald, M.H. Greve, Modeling soil organic carbon at regional scale by combining multi-spectral images with laboratory spectra, PLoS One 10 (11) (2015) e0142295, https://doi.org/10.1371/journal.pone.0142295. Available from.
[21] S. Liu, J. Chen, L. Guo, J. Wang, Z. Zhou, J. Luo, et al., Prediction of soil organic carbon in soil profiles based on visible–near-infrared hyperspectral imaging spectroscopy, Soil. Tillage Res. 232 (2023) 105736, https://doi.org/10.1016/j.still.2023.105736. Available from.
[22] Y. Yuan, H. Chen, L. Zhang, B. Ren, S. Xing, J. Tong, Prediction of spatial distribution of soil organic carbon in farmland based on multi-variables and random forest algorithm, Acta. Pedologica. Sin 58 (4) (2021) 623–636, https://doi.org/10.11766/trxb202001140623. Available from.
[23] M.C. de Castro Padilha, L.E. Vicente, J.A.M. Demattê, D. Gomes, A.K. Vicente, D. U. Salazar, et al., Soil organic carbon prediction and mapping in crop areas from a tropical region using Landsat satellite and clay content, Geoderma Reg 21 (2020) e00253, https://doi.org/10.1016/j.geodrs.2020.e00253. Available from.
[24] X. Zhu, E.H. Helmer, F. Gao, D. Liu, J. Chen, M.A Lefsky, Fusion of Landsat and MODIS data for high spatial- and temporal-resolution monitoring of forest disturbance, Remote Sens. Environ. 198 (2017) 298–309, https://doi.org/10.1016/j.rse.2017.06.022. Available from.
[25] K. Paustian, J. Lehmann, S. Ogle, D. Reay, G.P. Robertson, P. Smith, Climate-smart soils, Nature 532 (7597) (2016) 49–57, https://doi.org/10.1038/nature17174. Available from.
[26] R. Lal, Restoring soil quality to mitigate soil degradation, Sustainability 7 (5) (2015) 5875–5895, https://doi.org/10.3390/su7055875. Available from.
[27] R. Lal, Managing soils for resolving the conflict between agriculture and nature: the hard talk, Eur. J. Soil Sci. 71 (1) (2020) 1–9, https://doi.org/10.1111/ejss.12998. Available from.
[28] A.J. Franzluebbers, Achieving soil organic carbon sequestration with conservation agricultural systems in the southeastern United States, Soil Sci. Soc. Am. J. 74 (2) (2010) 347–357, https://doi.org/10.2136/sssaj2009.0079. Available from.
[29] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785. Available from.
[30] D.F. Ball, Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils, J. Soil Sci. [Internet] 15 (1) (1964) 84–92, https://doi.org/10.1111/j.1365-2389.1964.tb00247.x. Mar 1Available from.
[31] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis, D. Bochtis, Remote sensing techniques for soil organic carbon estimation: a review, Remote Sensing [Internet] 11 (6) (2019) 676, https://doi.org/10.3390/rs11060676. Mar 21Available from.
[32] T. Li, L. Cui, Y. Wu, T.I. McLaren, A. Xia, R. Pandey, H. Liu, Soil organic carbon estimation via remote sensing and machine learning techniques: global topic modeling and research trend exploration, Remote Sens (Basel) 16 (17) (2024) 3168, https://doi.org/10.3390/rs16173168. Available from.
[33] F.A. Diaz-Gonzalez, J. Vuelvas, C.A. Correa, V.E. Vallejo, Machine learning and remote sensing techniques applied to estimate soil indicators–review, Ecol. Indic. 136 (2022) 108655, https://doi.org/10.1016/j.ecolind.2021.108655. Available from.
[34] B. Van Wesemael, S. Chabrillat, A. Sanz Dias, M. Berger, Remote sensing for soil organic carbon mapping and monitoring, Remote Sens (Basel) 15 (14) (2023) 3464, https://doi.org/10.3390/rs15143464. Available from.
[35] T. Li, L. Cui, M. Kuhnert, T.I. McLaren, R. Pandey, A comprehensive review of soil organic carbon estimates: integrating remote sensing and machine learning technologies, J. Soils Sediments (2024), https://doi.org/10.1007/s11368-024-03913-8. Available from.
[36] M. Pavlovic, S. Ilic, N. Ralevic, N. Antonic, D.W. Raffa, A deep learning approach to estimate soil organic carbon from remote sensing, Remote Sens (Basel) 16 (4) (2024) 655, https://doi.org/10.3390/rs16040655. Available from.
[37] D. Radočaj, M. Gašparović, M. Jurišić, Open remote sensing data in digital soil organic carbon mapping: a review, Agriculture 14 (7) (2024) 1005, https://doi.org/10.3390/agriculture14071005. Available from.
[38] P. Abdoli, A. Khanmirzaei, S. Hamzeh, S. Rezaei, Use of remote sensing data to predict soil organic carbon in some agricultural soils of Iran, Remote Sens. Appl.: Society Environ. 28 (2023) 100579, https://doi.org/10.1016/j.rsase.2023.100579. Available from.
[39] H. Zayani, Y. Fouad, D. Michot, Z. Kassouk, N. Baghdadi, Using machine-learning algorithms to predict soil organic carbon content from combined remote sensing imagery and laboratory Vis-NIR spectral datasets, Remote Sens (Basel) 15 (17) (2023) 4264, https://doi.org/10.3390/rs15174264. Available from.
[40] Q. Chen, Y. Wang, X. Zhu, Soil organic carbon estimation using remote sensing data-driven machine learning, PeerJ 12 (2024) e17836, https://doi.org/10.7717/peerj.17836. Available from.
[41] S. Wang, K. Guan, C. Zhang, D.K. Lee, Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: assessing potential of airborne and spaceborne optical soil sensing, Remote Sens. Environ. 280 (2022) 113195, https://doi.org/10.1016/j.rse.2022.113195. Available from.
[42] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis, Remote sensing techniques for soil organic carbon estimation: a review, Remote Sens (Basel) 11 (6) (2019) 676, https://doi.org/10.3390/rs11060676. Available from.
[43] O. Odebiri, O. Mutanga, J. Odindi, R. Naicker, Deep learning approaches in remote sensing of soil organic carbon: a review of utility, challenges, and prospects, Environ. Monit. Assess. 193 (2021) 561, https://doi.org/10.1007/s10661-021-09561-6. Available from.
[44] J. Heil, C. Jörges, B. Stumpe, Fine-scale mapping of soil organic matter in agricultural soils using UAVs and machine learning, Remote Sens (Basel) 14 (14) (2022) 3349, https://doi.org/10.3390/rs14143349. Available from.
[45] Emadi M., Taghizadeh-Mehrjardi R., Cherati A. Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. Remote Sensing. 2020;12(14):2234. Available from: doi:10.3390/rs12142234.
[46] C. Zhu, Y. Wei, F. Zhu, W. Lu, Z. Fang, Z. Li, J. Pan, Digital mapping of soil organic carbon based on machine learning and regression kriging, Sensors 22 (22) (2022) 8997, https://doi.org/10.3390/s22228997. Available from.
[47] O. Odebiri, O. Mutanga, J. Odindi, R. Naicker, Modelling soil organic carbon stock distribution across different land-uses in South Africa: a remote sensing and deep learning approach, ISPRS J. Photogramm. Remote Sens. 184 (2022) 284–296, https://doi.org/10.1016/j.isprsjprs.2022.05.017. Available from.
[48] Y. Zhou, X. Zhao, X. Guo, Y. Li, Mapping of soil organic carbon using machine learning models: combination of optical and radar remote sensing data, Soil Sci. Soc. Am. J. 86 (3) (2022) 837–853, https://doi.org/10.1002/saj2.20371. Available from.
[49] B.D.A. Bartsch, N.A. Rosin, J.T.F. Rosas, Space-time mapping of soil organic carbon through remote sensing and machine learning, Soil Tillage Res. (2025), https://doi.org/10.1016/j.still.2024.107887. Available from.

[50] S. Nawar, M. Abdul Munnaf, A.M Mouazen, Machine learning based on-line prediction of soil organic carbon after removal of soil moisture effect, Remote Sens (Basel) 12 (8) (2020) 1308, https://doi.org/10.3390/rs12081308. Available from.
[51] T. Zhang, Y. Li, M. Wang, Remote sensing-based prediction of organic carbon in agricultural and natural soils influenced by salt and sand mining using machine learning, J. Environ. Manage. 340 (2024) 117264, https://doi.org/10.1016/j.jenvman.2024.117264. Available from.
[52] Statistical Procedures for Agricultural Research, 2nd Edition (1984) Available online: https://www.wiley.com/en-us/Statistical+Procedures+for+Agricultural+Research%2C+2nd+Edition-p-9780471870920.
[53] H.O. Tuffour, I.B. Yeboah, M. Bonsu, T. Adjei-Gyapong, A.A. Khalid, A. Abubakar, C. Melenya, P. Kpotor, Soil organic carbon: relating the walkley-black wet oxidation method to loss on ignition and clay content, Int. J. Scientific Res. Knowl. 2 (2014) 249–256, https://doi.org/10.12983/ijsrk-2014-p0249-0256.
[54] M. Oukhattar, S. Gadal, Y. Robert, N. Saby, et al., Variability analysis of soil organic carbon content across land use types and its digital mapping using machine learning and deep learning algorithms, Environ. Monit. Assess. 197 (5) (2025). Available from, https://link.springer.com/article/10.1007/s10661-025-13972-0.
[55] J. He, Y. Zhang, M. Liu, L. Chen, W. Man, et al., Prediction of Soil Organic Carbon Content in Spartina Alterniflora by Using UAV Multispectral and LiDAR Data, IEEE J. Select. Topics Appl. Earth Observ. Remote Sens. (2025). Available from, https://ieeexplore.ieee.org/abstract/document/10854657/.
[56] S. Li, X. Li, X. Ge, Prediction and mapping of soil organic carbon in the Bosten Lake oasis based on Sentinel-2 data and environmental variables, Int. Soil Water Conservat. Res. (2025). Available from, https://www.sciencedirect.com/science/article/pii/S2095633924000911.
[57] J. Xiao, W. Zhou, T. Wang, Y. Peng, Z. Shi, et al., Surface soil organic carbon estimation based on habitat patches in southwest China, IEEE J. Selected Topics Appl. Earth Observ. Remote Sens. (2024). Available from, https://ieeexplore.ieee.org/abstract/document/10812010/.
[58] T.P. Nguyen, P.K. Nguyen, H.N. Nguyen, et al., A new approach in soil organic carbon estimation using machine learning algorithms: a study in a tropical forest in Vietnam, J. Forest Res. (2025). Available from, https://www.tandfonline.com/doi/abs/10.1080/13416979.2024.2436748.
[59] M. Dadgar, S.E. Faramarzi, Assessing the performance of machine learning models for predicting soil organic carbon variability across diverse landforms, Environ Earth Sci (2024). Available from, https://link.springer.com/article/10.1007/s12665-024-11960-0.
[60] Q. Chen, Y. Wang, X. Zhu, Soil organic carbon estimation using remote sensing data-driven machine learning, PeerJ 12 (2024) e17836. Available from, https://peerj.com/articles/17836/.
[61] X. Subi, M. Eziz, N. Wang, Improving the estimation accuracy of soil organic matter content based on the spectral reflectance from soils with different grain sizes, Land (Basel) 13 (7) (2024) 1111. Available from, https://www.mdpi.com/2073-445X/13/7/1111.
[62] S. Falahatkar, S.M. Hosseini, S. Ayoubi, A. Salmanmahiny, Predicting soil organic carbon density using auxiliary environmental variables in northern Iran, Arch. Agron. Soil Sci. 62 (2015) 375–393, https://doi.org/10.1080/03650340.2015.1051472.
[63] M. Ajami, A. Heidari, F. Khormali, M. Gorji, S. Ayoubi, Environmental factors controlling soil organic carbon storage in loess soils of a subhumid region, northern Iran, Geoderma 281 (2016) 1–10, https://doi.org/10.1016/j.geoderma.2016.06.017.