Environmental Technology & Innovation 43 (2026) 104971

Contents lists available at ScienceDirect

Environmental Technology & Innovation

journal homepage: www.elsevier.com/locate/eti

Optimizing Sentinel-2 temporal composites for soil organic carbon mapping and cropland management insights

Xiande Ji [a,*], R. Venkatesha Prasad [b], Binyuan Liu [c], Balamuralidhar Purushothaman [d], P.V. Aravind [a,e]

[a] Energy Conversion Group, Energy and Sustainability Research Institute Groningen, Faculty of Science and Engineering, University of Groningen, Nijenborgh 6, 9747AG Groningen, The Netherlands
[b] Delft University of Technology, Leeghwaterstraat 39, 2628CB Delft, The Netherlands
[c] Integrated Research on Energy, Environment and Society (IREES), Energy and Sustainability Research Institute Groningen, Faculty Science and Engineering, University of Groningen, Nijenborgh 6, 9747AG Groningen, The Netherlands
[d] TCS Research, Tata Consultancy Services, Bangalore, India
[e] Department of Process and Energy, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Leeghwaterstraat 39, 2628CB, Delft, The Netherlands

ARTICLE INFO

Keywords:
Soil organic carbon
Bare soil frequency
Sentinel-2
Digital soil mapping
Agricultural management

ABSTRACT

Accurate mapping of soil organic carbon (SOC) in intensive croplands is important for climate change mitigation and for guiding sustainable agricultural management. Despite the growing use of Sentinel-2 composites, evidence remains limited on how composite design affects SOC mapping accuracy in croplands and on whether satellite observations can capture management-relevant signals linked to SOC. This study compared four temporal Sentinel-2 spectral composites for SOC mapping using LUCAS 2015 and 2018 observations in Italy's Po Plain. Three machine learning models, random forest, XGBoost, and CatBoost, were trained, and SHAP was used to interpret variable contributions. Across models, composites targeting the bare soil period, based on multispectral reflectance and non-photosynthetic vegetation indices, achieved the best performance. CatBoost performed best and produced a high-resolution SOC map for the Po Plain. In contrast, traditional vegetation indices such as NDVI and EVI showed limited relevance across all composites. Importantly, we found a robust negative association between SOC and bare soil frequency derived from multi-temporal Sentinel-2 observations, with lower bare soil frequency corresponding to higher SOC. This highlights bare soil exposure duration as a practical indicator for monitoring and suggests that management practices that shorten bare soil windows may help maintain or enhance SOC. Overall, this study optimized Sentinel-2 temporal composites with machine learning to improve SOC mapping in the Po Plain and provides actionable insights for cropland management in intensively cultivated regions.

1. Introduction

Soil organic carbon (SOC) is a key component of the global carbon cycle and agricultural ecosystems. It plays an important role in soil fertility, plant growth, and water retention, and is therefore closely linked to sustainable crop production and food security (Crowther et al., 2016; Sanchez et al., 2009). However, cropland SOC stocks have declined over the long term under the combined pressures of climate change, soil degradation, and human activities (Schlesinger and Amundson, 2019; Huang et al.,

* Corresponding author.
E-mail address: x.ji.research@gmail.com (X. Ji).

https://doi.org/10.1016/j.eti.2026.104971
Received 25 March 2026; Received in revised form 22 April 2026; Accepted 4 May 2026
Available online 8 May 2026

2022). For example, one study estimated that 1.8 petagrams of SOC in the continental United States could be lost by 2100 due to climate change (Gautam et al., 2022). At the same time, cropland soils also have substantial potential to act as carbon sinks under improved management (Gattinger et al., 2012; Paustian et al., 2016; Minasny et al., 2017). Therefore, accurate SOC mapping and the identification of its key drivers are essential for monitoring SOC dynamics and supporting effective cropland management.

Digital soil mapping (DSM) has emerged as a powerful approach for predicting soil properties, including SOC, from local to global scales (Lamichhane et al., 2021; Rahmani et al., 2022; He et al., 2021; Zhang et al., 2023). This method identifies the relationships between measured SOC data and environmental variables to build predictive models that estimate the spatial distribution of SOC across entire study areas (Ji et al., 2024). Satellite remote sensing data, such as Sentinel-2, Landsat, and MODIS, together with spectral indices (SIs) derived from these data, have been widely used as important predictors in DSM for SOC prediction from local to large spatial scales (Zhou et al., 2020; Liu et al., 2023; Urbina-Salazar et al., 2023; Zhang et al., 2021). Previous studies have compared prediction accuracy by integrating different types of environmental covariates, including remote sensing and SIs (Jeong et al., 2017; Zhang et al., 2022a; Taghizadeh-Mehrjardi et al., 2021; Ai et al., 2025).

However, beyond the choice and combination of predictor types, the timing of environmental variable acquisition is also critical for SOC prediction (Wang et al., 2023a; Zepp et al., 2023; Sun et al., 2023; Yang et al., 2020). In croplands, the suitable data acquisition time for SOC prediction is unclear and controversial. On the one hand, some previous studies focused on the crop growth period, using vegetation information, such as NDVI and EVI, to characterize the spatial variations of SOC. For example, Liu et al. (2023) demonstrated that incorporating MODIS-derived phenological factors significantly enhanced SOC prediction in the Heihe River Basin, China. Wang et al. (2023a) confirmed that aboveground biomass had great potential for SOC prediction in cotton fields. In addition, vegetation height was introduced and provided additional explanatory power for SOC spatial variation (Takada et al., 2009). Moreover, Zhao et al. (2014) improved the model performance at the peak growing season based on the vegetation indices. On the other hand, some studies captured remote sensing data from the bare soil period (Zepp et al., 2023). Zhao et al. (2014) selected images at the spring and autumn seasons, when most crops were harvested, and the soil surface was exposed. Kalambukattu et al. (2018) confirmed that the Landsat-8 derived bare soil indices were highly related to SOC content.

These studies only focused on a single acquisition period, such as either the bare soil or vegetation cover stage, without systematically evaluating the effects of different combinations of multi-temporal spectral variables and indices. Therefore, there remains a lack of comparative analysis on how different temporal composites of remote sensing variables influence SOC prediction performance.

In addition, in large cropland areas, fields rarely remain in the same phenological phase because of spatial heterogeneity. As a result, bare soil, partial cover, and full vegetation can coexist across neighboring fields at a given time. Multi-temporal spectral observations can therefore be used to capture land surface conditions throughout the observation period, allowing the construction of state-specific reflectance composites and corresponding spectral indices, while also helping characterize crop phenology and management (Shafizadeh-Moghadam et al., 2022; Castaldi, 2021). Bare soil frequency (BSF), derived from multi-temporal observations, summarizes how often fields are observed under bare soil conditions and can serve as a proxy for the duration of non-vegetated intervals within cropping cycles (de Sousa et al., 2024). Because crop types and management practices strongly influence the balance between organic carbon inputs and losses, spatial variation in BSF provides an interpretable temporal indicator for examining how cropland management may affect SOC levels and their spatial distribution.

The Po Plain is the largest plain in Italy and one of the most productive agricultural regions in Europe (Monteleone et al., 2023), yet it lacks high-resolution SOC maps and bare soil period investigations. To address the above research gap and provide accurate SOC mapping in this area, this study aims to investigate how Sentinel-2 image acquisition dates affect DSM accuracy for SOC prediction and test whether areas with SOC distributions are linked to different agricultural practices by observing BSF. Specifically, we aim (1) to compare the performance and importance of Sentinel-2-based data including multispectral bands and SIs in different temporal composites; (2) to evaluate the effectiveness of three models, including random forest (RF), extreme gradient boosting (XGBoost), and categorical boosting (CatBoost) for SOC prediction in croplands; (3) to develop and evaluate the models for SOC prediction with the 30 m spatial resolution in croplands of the Po Plain; (4) to calculate the BSF and investigate its influence on SOC among crop types based on the predicted SOC map. This work will contribute to identifying the optimal approach for SOC mapping in agricultural landscapes and investigating the impacts of land surface exposure duration on SOC levels. Furthermore, the study will provide insights into the spatial distribution and agricultural management of cropland SOC.

2. Materials and methods

2.1. Study area

The study area covers cropland regions in the Po Plain, extending from 8.5°E to 13.1°E longitude and 43.7°N to 46.7°N latitude. It is situated between the Alps to the north and the Apennines to the south, in the Lombardy, Emilia-Romagna, and Veneto regions, with a total area of approximately 27300 km^2 (Fig. 1(a)). The region experiences a temperate climate, with a mean annual temperature of 11 °C and annual precipitation of 880 mm. The mean elevation of the region is about 50 m above sea level, while the eastern coastal region has relatively low elevations, with some areas even below sea level. The soils of the Po Plain are predominantly alluvial, formed by the deposition of sediments from the Po River and its tributaries (Costantini et al., 2004). In this study area, the dominant crop is maize, accounting for 46.6%, followed by common wheat (18.4%), grass (12.1%), and permanent crops located within woodland and shrubland areas (8.3%) (Fig. 1(b)). These crop statistical data were derived from the EU Crop Map (EUCROPMAP v1) dataset.

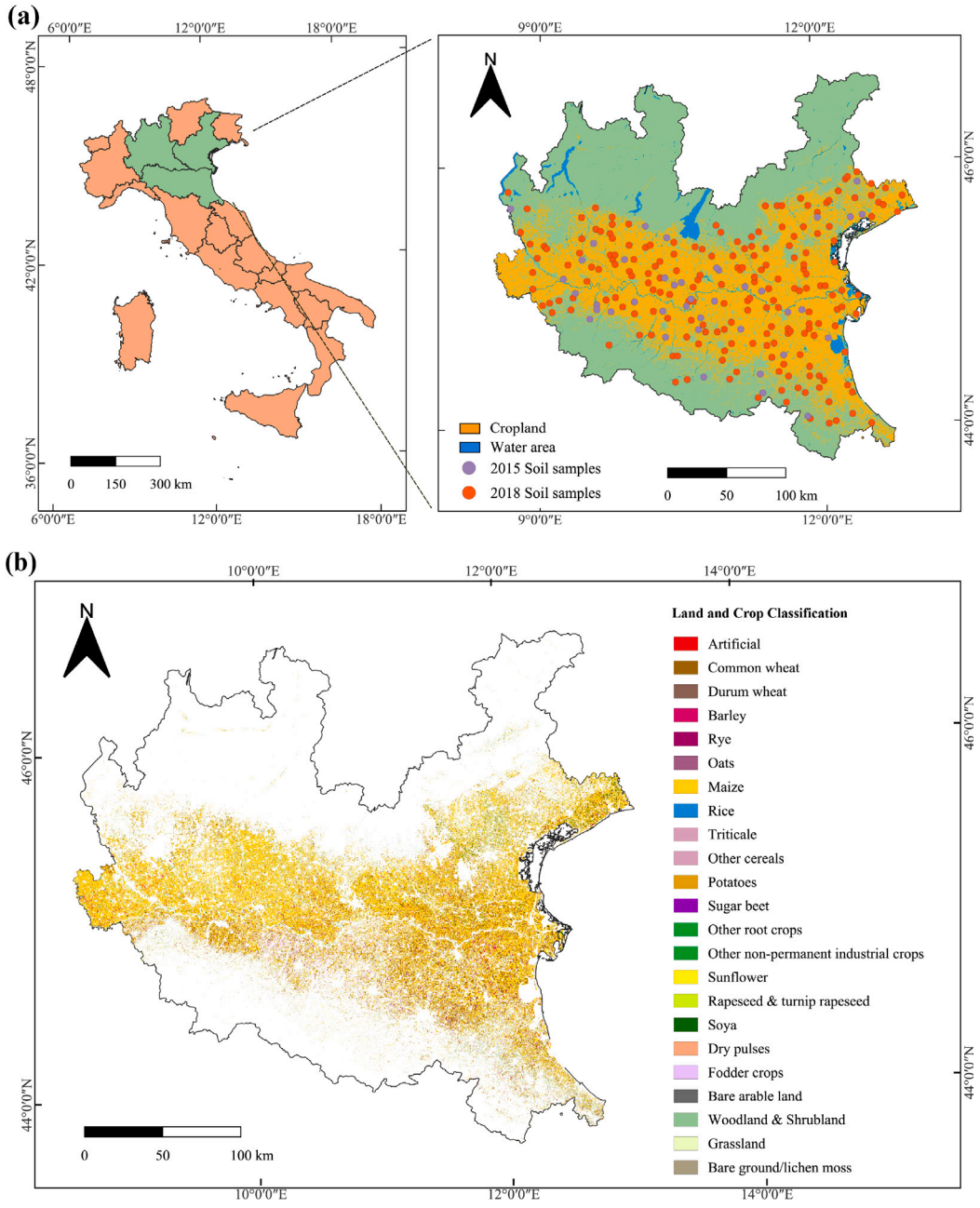

Fig. 1. (a) Location of the study area and distribution of LUCAS soil samples. (b) Land and crop classification in the study area based on EUCROPMAP v1.

2.2. Soil samples

Soil samples used in this study were obtained from the Land Use/Cover Area Frame Statistical Survey (LUCAS) topsoil (0–20 cm) database, which is the largest European soil spectral library provided by the European Soil Data Centre (ESDAC) (Orgiazzi et al., 2018). The LUCAS soil sampling campaign was initiated in 2009, with updates conducted every three years. Currently, the 2009/2012, 2015, and 2018 datasets are publicly available. Twelve physical and chemical properties, including SOC, were measured in a single laboratory to keep the data consistent. The measurement of SOC content used the ISO 10694:1995 protocol by dry combustion (Wang et al., 2020). SOC sequestration is usually a medium to long-term process (Chenu et al., 2019). Previous work has shown that SOC changes are often difficult to detect before 6 to 10 years, and studies of three years or less may be too short to robustly assess SOC stock changes (Chaplot and Smith, 2023; Smith, 2004). In this study, to improve model reliability while

keeping the temporal span of the soil observations relatively narrow, soil samples were selected from the LUCAS 2015 and LUCAS 2018 datasets for cropland within the study area. If a sampling location was revisited in LUCAS 2018, only the latest SOC values from the 2018 dataset were retained. In total, 210 soil samples were extracted, comprising 170 samples from LUCAS 2018 and 40 from LUCAS 2015 (Fig. 1(a)). These samples were randomly split into training data (70%) and testing data (30%) for further model training and evaluation (Huang et al., 2017; Ji et al., 2024; Hu et al., 2025).

2.3. Environmental variables

In this study, the environmental covariates, including climate, topography, and soil properties, were used with multispectral data and SIs to predict SOC content. All these variables were resampled to 30 m spatial resolution using the nearest neighbor method and were reprojected to the WGS 1984 UTM Zone 32N projection (Liu et al., 2022). All preprocessing and computation of variables were performed on the Google Earth Engine (GEE) platform.

2.3.1. Climate and topography

The annual mean temperature and annual precipitation were selected as climatic variables for this study. These climate data were obtained from WorldClim version 2.1, which provides long-term climate data (1970–2000) at a 1 km resolution (Fick and Hijmans, 2017). Elevation in our study area was extracted from the Shuttle Radar Topography Mission (SRTM) V3 product with a 30 m resolution accessed via the GEE platform. Additional topographic variables, including slope, aspect, topographic wetness index (TWI), length-slope factor (LSF), valley depth (VD), relative slope position (SP), plan curvature (PC), and convergence index (Conver), were derived from the elevation map using SAGA GIS.

2.3.2. Soil properties

Soil properties are strongly associated with SOC content. Total nitrogen (TN) and pH were identified as key predictors in DSM for SOC prediction (Deng et al., 2018; Castro-Franco et al., 2015). This study selected TN and pH derived from the SoilGrids 250 m product. Soil properties at a depth of 0–20 cm were derived by aggregating the original depth layers (0–5 cm, 5–15 cm, and 15–20 cm) using weighted averaging techniques (Huang et al., 2024).

2.3.3. Sentinel-2 data

This study extracted multispectral bands from Sentinel-2 imagery via the GEE platform during the period 2018–2020. Sentinel-2 sensors capture 13 spectral bands, spanning the visible to near-infrared (VNIR) and shortwave infrared (SWIR) ranges, enabling applications in land cover and land use monitoring, vegetation analysis, and agricultural management (Segarra et al., 2020; He et al., 2021). In this study, ten multispectral bands (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12) were selected. Additionally, spectral-based vegetation indices were calculated using Sentinel-2 bands as indicators of vegetation and land cover. NDVI, EVI, and soil-adjusted vegetation index (SAVI) are widely used to represent vegetation growth for SOC prediction, and their importance for modeling has been indicated in previous studies (Xiao et al., 2019; Yang et al., 2020; Nguyen et al., 2022). Kernel NDVI (kNDVI) is an advanced vegetation index that improves the characterization of vegetation dynamics and enhances its sensitivity and robustness compared to NDVI (Wang et al., 2023b). In addition to vegetation indices, the normalized burn ratio 2 (NBR2) is used to assess soil moisture levels and the presence of straw and crop residues (Castaldi et al., 2019). The land surface water index (LSWI) is a moisture index calculated by the SWIR band, which is sensitive to changes in vegetation water content (John et al., 2018). Clay index (CI) is used to capture spectral variation related to clay content in soil (Li et al., 2025). The non-photosynthetic vegetation-soil separation index (NSSI) is a novel index proposed by Tian et al. (2021) to improve the distinction between bare soil and non-photosynthetic vegetation, which has shown promising results in SOC mapping (Ji et al., 2024). The formulas for all eight indices are given in Table A.1. This study used only images with less than 10% cloud cover to ensure data quality and reliability. Additionally, images were restricted to the period from March to October because the high solar zenith angle during winter months results in receiving low energy levels by the Sentinel-2 sensors (Vermote et al., 2016). The Sentinel-2 images were categorized into three temporal periods to explore the impacts of temporal composites. The first period focuses on the bare soil situation, including post-harvest and fallow periods. NDVI values were used to identify bare soil pixels, with values below 0.3 classified as bare soil (Davis et al., 2019; Xu and Zhai, 2023). Pixels meeting this criterion were extracted and averaged to represent bare soil conditions. The second period aimed to account for the effects of the entire vegetation growth cycle. For this purpose, all images, with NDVI values greater than 0.3, from March to October were collected and averaged to derive variables representing the full growth period of vegetation cover. The reference group covered the entire observation period from March to October. Ten multispectral bands and eight SIs were processed according to this standard for each period.

2.4. Variable selection

Four variable groups were constructed to compare the effects of different Sentinel-2 temporal composites on SOC prediction (Table 1). Group 1 included multispectral bands and spectral indices (SIs) derived from the bare soil period. Group 2 combined multispectral bands from the bare soil period with SIs from the vegetation cover period. Group 3 used both multispectral bands and SIs from the vegetation cover period, whereas Group 4 used both multispectral bands and SIs from the entire observation period from March to October. In all groups, these Sentinel-2 variables were combined with climate, topography, and soil properties for SOC prediction.

Table 1
The environmental variables for each group and the selected numbers for predictors based on the Boruta algorithm.

Group name	The included variable categories	Variable numbers	Number of variables after Boruta
Group 1	Climate, Topography, Soil property, S2_b, SI_b	31	15
Group 2	Climate, Topography, Soil property, S2_b, SI_v	31	14
Group 3	Climate, Topography, Soil property, S2_v, SI_v	31	11
Group 4	Climate, Topography, Soil property, S2_e, SI_e	31	12

* Note: S2_e: Sentinel-2 bands in entire observation period, S2_b: Sentinel-2 bands selected with NDVI less than 0.3 period, S2_v: Sentinel-2 bands selected with NDVI greater than 0.3 period, SI_e: Spectral index in entire observation period, SI_b: spectral index selected with NDVI less than 0.3 period, SI_v: Spectral index selected with NDVI greater than 0.3 period.

For the four variable groups, to avoid multicollinearity and redundancy while extracting significant variables to improve prediction accuracy, the Boruta approach was used to select significant predictors for modeling SOC prediction (Rahmani et al., 2022). The Boruta approach is a feature selection algorithm that identifies relevant features in a large dataset by comparing the importance of original features against randomized shadow features based on the random forest algorithm. It is especially effective for high-dimensional datasets to distinguish useful predictors (Keskin et al., 2019; Xu et al., 2017). In this study, the Boruta method was applied to each group. To improve the stability of feature selection and to identify a more consistent subset of predictors, the Boruta algorithm was iterated 20 times for each group to calculate the average importance of all variables (Meinshausen and Bühlmann, 2010; Khaire and Dhanalakshmi, 2022). Variables that ranked within the top ten in at least 10 iterations were selected as predictors for SOC modeling (Szul et al., 2021). The detailed rankings are shown in the Supplementary Materials (Fig. A.1). As a result, we selected 15, 14, 11, and 12 variables for Group 1, Group 2, Group 3, and Group 4, respectively (see Fig. 2).

2.5. Predictive models

Machine learning models have been widely used for modeling the nonlinear relationships between SOC and the corresponding environmental variables (Kalambukattu et al., 2018; Zeraatpisheh et al., 2019; Wang et al., 2018). In this study, to investigate the performance of temporal composites from Sentinel-2 data, three widely used models, including random forest (RF), extreme gradient boosting (XGBoost), and categorical boosting (CatBoost), were employed. Each model predicted SOC for the four variable groups using predictors selected by the Boruta algorithm.

2.5.1. Random forest

RF is an ensemble model that employs a bagging strategy to generate multiple decision trees, aggregating their predictions to improve model robustness (Breiman, 2001). Unlike single-tree models, RF averages outputs across all trees in regression tasks, thereby reducing noise and variability (Rasaei and Bogaert, 2019). Each tree is built using a bootstrap sample of the data and a random subset of features, which helps mitigate overfitting and improve generalization performance. In the previous SOC mapping studies (Hengl et al., 2015; Castro-Franco et al., 2015), RF has been adapted and reported as an effective model for SOC prediction at various scales.

2.5.2. XGBoost

XGBoost adopts a gradient boosting framework, which sequentially builds decision trees, where each new tree attempts to correct the errors of the previous ones (Chen and Guestrin, 2016). XGBoost integrates L1/L2 regularization and depth-based post-pruning to reduce overfitting while enhancing model generalization. A key advantage of XGBoost is its support for parallel computing, enabling efficient use of multi-core CPUs and GPUs to accelerate training (He et al., 2024). XGBoost has been gradually used for SOC mapping due to its superior predictive performance and speed (Chen et al., 2024; Batunacun et al., 2021).

2.5.3. CatBoost

CatBoost is an advanced gradient boosting algorithm specifically designed to deal with categorical features effectively (Hancock and Khoshgoftaar, 2020). It natively handles categorical variables through an efficient ordered target encoding scheme, which dynamically optimizes feature representations during training while eliminating manual feature engineering. Furthermore, its ordered boosting mechanism sequentially trains on permuted data partitions, preventing target leakage, significantly reducing prediction bias, and enhancing generalization performance (Mantena et al., 2023). Despite its proven efficacy in a large variety of applications, such as medical and financial fields, its potential for SOC prediction in croplands remains unexplored.

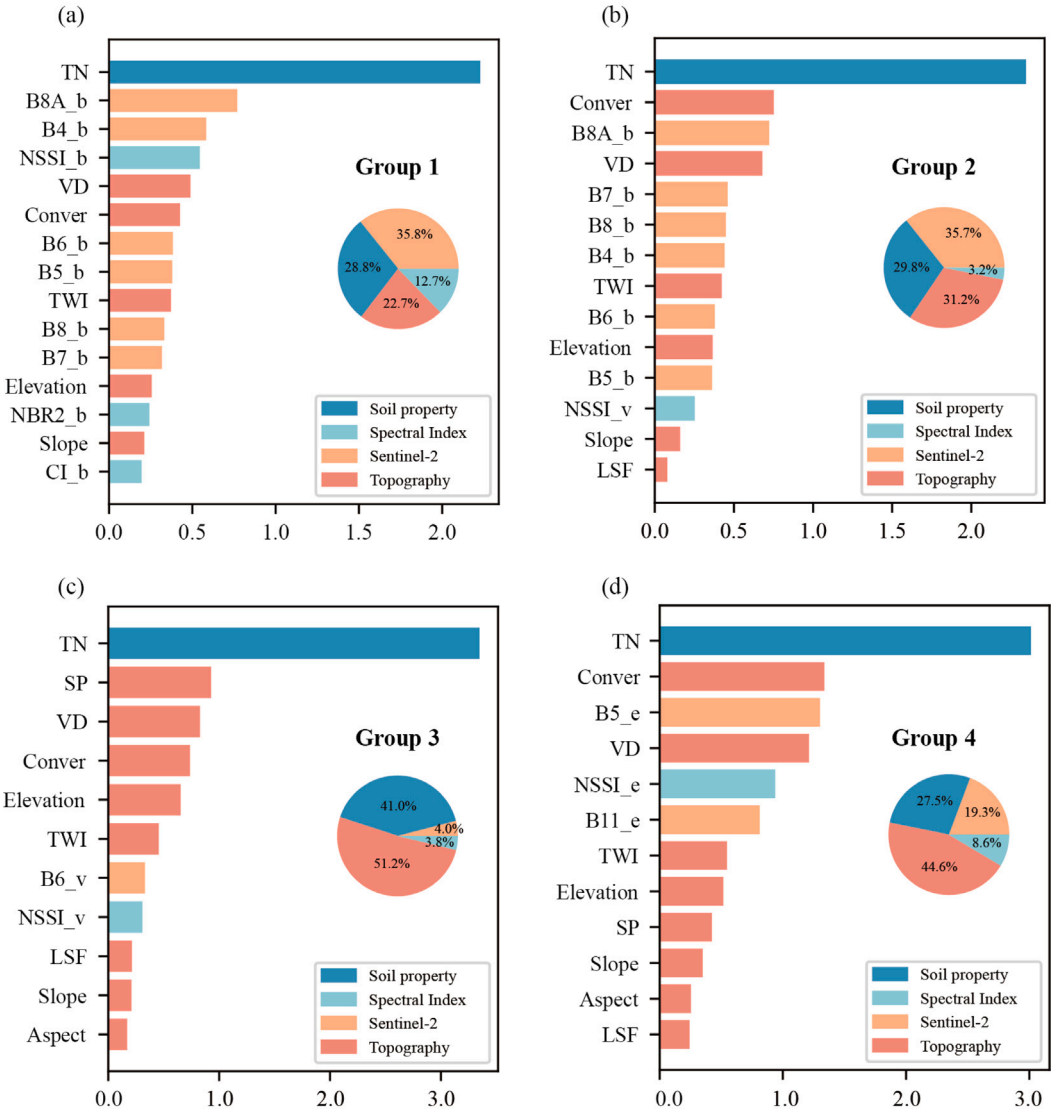

Fig. 2. The variable importance for Group 1, Group 2, Group 3, and Group 4 in the CatBoost models is quantified by $|SHAP|$ values. "_b": bands or SIs in the bare soil period, "_v": bands or SIs in the vegetation cover period, "_e": bands or SIs in the entire observation period.

2.6. Model evaluation

To achieve high prediction accuracy, a comprehensive grid search integrated with 5-fold cross-validation was conducted to identify optimal values for key hyperparameters for each model in the training data. The optimal parameter set was determined by minimizing the averaged root mean squared error (RMSE) on the validation dataset, ensuring robust generalization to testing data.

The testing data were used to evaluate the performance of each model for predicting SOC. The accuracy evaluation indicators were the coefficient of determination (R^2), RMSE, and Lin's concordance correlation coefficient (CCC). Higher values of R^2 and CCC, combined with a lower RMSE, demonstrate better predictive accuracy and lower errors for the model (Ugbemuna Ugbaje et al., 2024; Wu et al., 2024). The formula for these indicators is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

Table 2
Performance evaluation of predictive models.

Model	Variable	R^2	RMSE (g/kg)	CCC
RF	Group 1	0.415	12.726	0.559
	Group 2	0.365	13.252	0.516
	Group 3	0.203	14.849	0.281
	Group 4	0.239	14.509	0.317
XGBoost	Group 1	0.416	12.712	0.555
	Group 2	0.407	12.807	0.519
	Group 3	0.106	15.728	0.160
	Group 4	0.187	15.001	0.238
CatBoost	**Group 1**	**0.440**	**12.450**	**0.557**
	Group 2	0.417	12.703	0.535
	Group 3	0.205	14.832	0.286
	Group 4	0.282	14.095	0.381

$$CCC = \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\bar{\hat{y}} - \bar{y})^2} \tag{3}$$

where y is the measured SOC content, $\hat{y}$ is the predicted SOC content from a predictive model, n is the number of testing samples, $\bar{y}$ and $\bar{\hat{y}}$ are the average values of measured and predicted SOC content, ρ is the correlation coefficient between the measurements and predictions, σ_y^2 and $\sigma_{\hat{y}}^2$ are the corresponding variances.

The spatial distribution of SOC content and its uncertainty was predicted using predictive models by the bootstrap approach (Wu et al., 2024). Specifically, the predictive model was trained over 10 bootstrap iterations using the training data to map the distribution of SOC across the whole spatial scale. The final SOC prediction was obtained by averaging the results, with uncertainty quantified as the standard deviation of these iterations.

The Shapley additive explanations (SHAP) framework was used to evaluate the variables' significance. SHAP is a game-theoretic approach that quantifies the marginal contribution of each variable within a predictive model (Li, 2022). It also provides global feature importance rankings and instance-level explanations, ensuring consistency and fairness in attribution across all model outputs. This rigorous quantification of variable contributions bridges the gap between complex model behavior and actionable interpretability (Lundberg and Lee, 2017; He et al., 2024). In this study, we utilized the SHAP framework to determine the relative importance of predictors for each temporal composite in the predictive models.

2.7. Statistics for bare soil frequency effects

BSF has been used as a management-relevant temporal indicator of bare soil exposure in croplands (de Sousa et al., 2024). Building on this, it is important to examine how SOC varies along the BSF spatial distribution. This analysis further evaluates whether BSF captures management-related differences that may contribute to spatial variation in SOC. BSF was calculated as the proportion of remote sensing observations (from March to October in this study) that were identified as bare soil (Nascimento et al., 2021). To further assess its influence on SOC variation, an analysis of variance (ANOVA) approach was applied to test the effects of BSF classes, crop types, and their interactions. The calculated BSF values were categorized into three classes: Class 1 with low BSF (0–0.25), Class 2 with medium BSF (0.25–0.5), and Class 3 with high BSF (0.5–1). Moreover, to minimize the effects of spatial heterogeneity in environmental covariates, we aimed to select three small regions with low uncertainty from the predicted SOC maps. For each region, 200 pixels were randomly sampled from the main crops within each BSF class for ANOVA analysis. Aligned rank transform ANOVA (ART-ANOVA) was employed as it relaxes the assumptions of normality and homoscedasticity required in conventional ANOVA, making it well suited for ecological and soil datasets that often deviate from these assumptions (Wobbrock et al., 2011; Zhou et al., 2021).

3. Results

3.1. The performance of predictive models

Table 2 shows the predictive performance of the three models for SOC estimation in four variable groups. Group 1 consistently showed better performance metrics across all models than the others. The result indicated that Sentinel-2 temporal composite in the bare soil period yielded the highest accuracy, with R^2 ranging from 0.415 to 0.440, RMSE from 12.450 g/kg to 12.726 g/kg, and CCC from 0.555 to 0.557. In contrast, variables selected in the entire vegetation period (Group 3) exhibited the lowest performance throughout these three models. Among the predictive models, CatBoost outperformed RF and XGBoost across four groups, particularly in Group 1, indicating the CatBoost model achieved the highest predictive accuracy with the lowest errors for SOC prediction in this study. Compared to RF and XGBoost in Group 1, the R^2 of CatBoost increased by 6.02% and 5.77%, respectively. The CatBoost model in Group 1 demonstrated the best performance for predicting SOC with high R^2 and CCC and low RMSE.

3.2. The importance of temporal composites for SOC prediction

Fig. 2 presents the selected variables from the Boruta method and their importance rankings for SOC prediction in the CatBoost model across the four variable groups. Across all groups, TN had the highest predictive importance for mapping SOC in cropland. Topography variables also had significant importance for SOC prediction. However, the performance of Sentinel-2 spectral data varied in different observation periods. In Group 1, multispectral bands accounted for the main contribution to prediction, a total of 35.8%. In addition, SIs, such as NSSI_b in bare soil, played a significant role. For Group 2, bare soil-based spectral bands had similar importance (35.7%) as in Group 1, whereas the contribution of SIs decreased significantly when they were selected from the vegetation period. In this period, only NSSI was regarded by Boruta as an important predictor. For Group 3, compared with the impacts of remote sensing images in bare soil on prediction, in the vegetation period, only B6 and NSSI remained influential. For Group 4, spectral data from the entire observation period were not more significant than those from the bare soil period, but outperformed the data from the vegetation cover period. Overall, the temporal composite of spectral bands and non-vegetation indices in the bare soil period performed well for SOC prediction, whereas the widely used vegetation indices, such as NDVI and EVI, have not shown significant results via the Boruta method in all composites for SOC mapping, even in the vegetation period.

3.3. The spatial distribution of soil organic carbon

Fig. 3(a–c) displays the spatial distribution of SOC content (0–20 cm) at a 30 m resolution, as predicted by the CatBoost model with the best variable combination group (Group 1). The SOC content of cropland in this study area was between 11.92 g/kg and 70.55 g/kg, with an average of 19.84 g/kg in Group 1. High SOC content was concentrated in the northeastern and eastern coastal regions, whereas the areas characterized by low SOC distributions were in the northwestern region. The prediction uncertainty of SOC is presented in Fig. 3(d–f). High uncertainty values were predominantly associated with high SOC regions, particularly in the eastern coastal areas. While some eastern regions exhibited elevated uncertainty, the majority of the study area demonstrated low uncertainty levels, with an average of 2.82 g/kg. Three small regions were selected in the low uncertainty areas for investigating the response of BSF on SOC levels (Fig. A.2).

Fig. 4 shows the distribution of SOC content across major crop types and BSF classes in the study area. In general, SOC distributions were right-skewed, reflecting strong spatial heterogeneity. As a result, the mean SOC was strongly influenced by these extremes, making the median a more reliable measure of central tendency. Median SOC levels were highest in soya, followed by grass, feed crops, and permanent crops in woodland and shrubland areas. All of these crops exceeded the overall regional median, indicating that they might store more SOC content in the soils. In contrast, potatoes, sunflowers, rapeseed, and barley showed comparatively lower median SOC levels. The median SOC across BSF classes showed Class 1 (low BSF) exhibited higher SOC values than the other two classes. This finding indicated that shorter bare soil durations might have greater SOC accumulation and stability.

3.4. The effects of bare soil frequency on SOC levels

The distributions of SOC content varied in different crop types and BSF classes in each selected region (Fig. A.3). For common wheat, maize, and grass across all three regions, SOC values tended to be higher under low BSF class (Class 1), while medium and high BSF classes (Classes 2 and 3) were associated with lower SOC distributions. Similar trends were also observed in sunflower, fodder crop, and potato in Region 1. These findings are consistent with the results obtained for the whole study area based on the predicted SOC distribution. The ART-ANOVA results confirmed that BSF had statistically significant negative effects on SOC values ($p < 0.001$) across three regions, and crop types also influenced SOC (Table 3). In particular, post-hoc tests further showed BSF Class 1 had significantly higher SOC than those with Class 2 and Class 3 ($p < 0.001$; Table A.2 in the Supplementary Materials). However, the interaction effects were not consistently significant across the three regions and were only significant in Region 2 and Region 3 ($p = 0.015$, $p = 0.014$). These results highlighted the highly significant relationship between BSF and SOC distribution on local scales. Across the analyzed regions, lower BSF was associated with higher SOC. This pattern suggests that cropping and management practices that maintain more surface cover and less bare soil exposure are linked to higher SOC levels.

4. Discussion

4.1. Comparison of predictive models for cropland SOC

This study used three machine learning models to predict agricultural SOC across four variable groups. The results showed that CatBoost performed better than RF and XGBoost in all groups, indicating its potential for DSM applications. For comparison, van Wesemael et al. (2024) used convolutional neural networks to predict SOC in European croplands and reported an R^2 of 0.41. In addition, a recent review based on 67 studies across a variety of spatial scales reported a median R^2 of 0.47 for SOC prediction in agricultural lands. Therefore, although the best CatBoost model in this study did not achieve high predictive accuracy, its performance remained within the commonly reported range and can be regarded as moderate for cropland SOC mapping.

However, several factors may have contributed to the observed prediction error. One source of uncertainty may arise from the harmonization of predictors with different spatial resolutions (Garosi et al., 2022; Zeraatpisheh et al., 2023). Although all variables were aligned to a common 30 m grid for pixel-level modeling, some important covariates, such as SoilGrids and climate data, were originally available at much coarser resolutions. This preprocessing step was necessary for spatial alignment, but coarse predictors

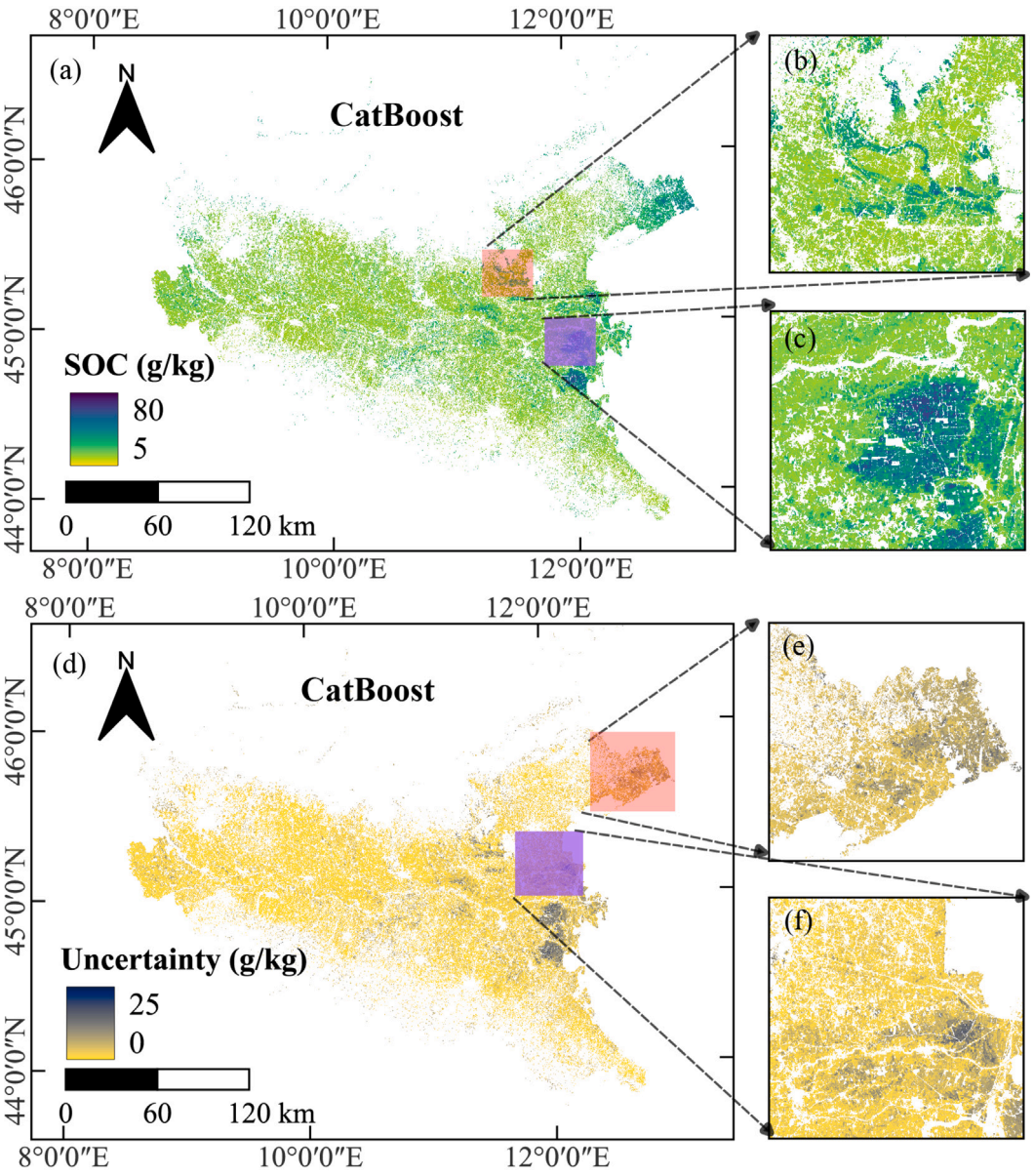

Fig. 3. Spatial distribution pattern of SOC content (g/kg) (a–c) and uncertainty (d–f) prediction estimated by the CatBoost model in Group 1.

cannot provide fine scale spatial detail and may not fully match finer resolution Sentinel-2 variables, which may have reduced the model's ability to capture local SOC variability and limited predictive accuracy (Guo et al., 2019). In addition, the limited sample size may have restricted the model's ability to fully represent the spatial heterogeneity of cropland SOC in the study area. This issue may be particularly relevant for areas with relatively high SOC, which appeared to be less well represented in the current dataset and were associated with higher prediction uncertainty (Ji et al., 2024; van Wesemael et al., 2024). Future work should therefore increase the number of field observations, especially in high SOC areas, to improve the representativeness of the training data. Moreover, transfer modeling approaches could be explored to improve local SOC prediction under limited sample conditions by transferring useful information from larger scale datasets or environmentally similar regions (Shen et al., 2022; Zhang et al., 2025).

4.2. Relative importance of the variables in different composites

Selecting image acquisition dates should consider the cropland surface conditions. In this study, the composite of the bare soil period provides a more appropriate timing for predicting SOC content compared to the other period composites for croplands. The

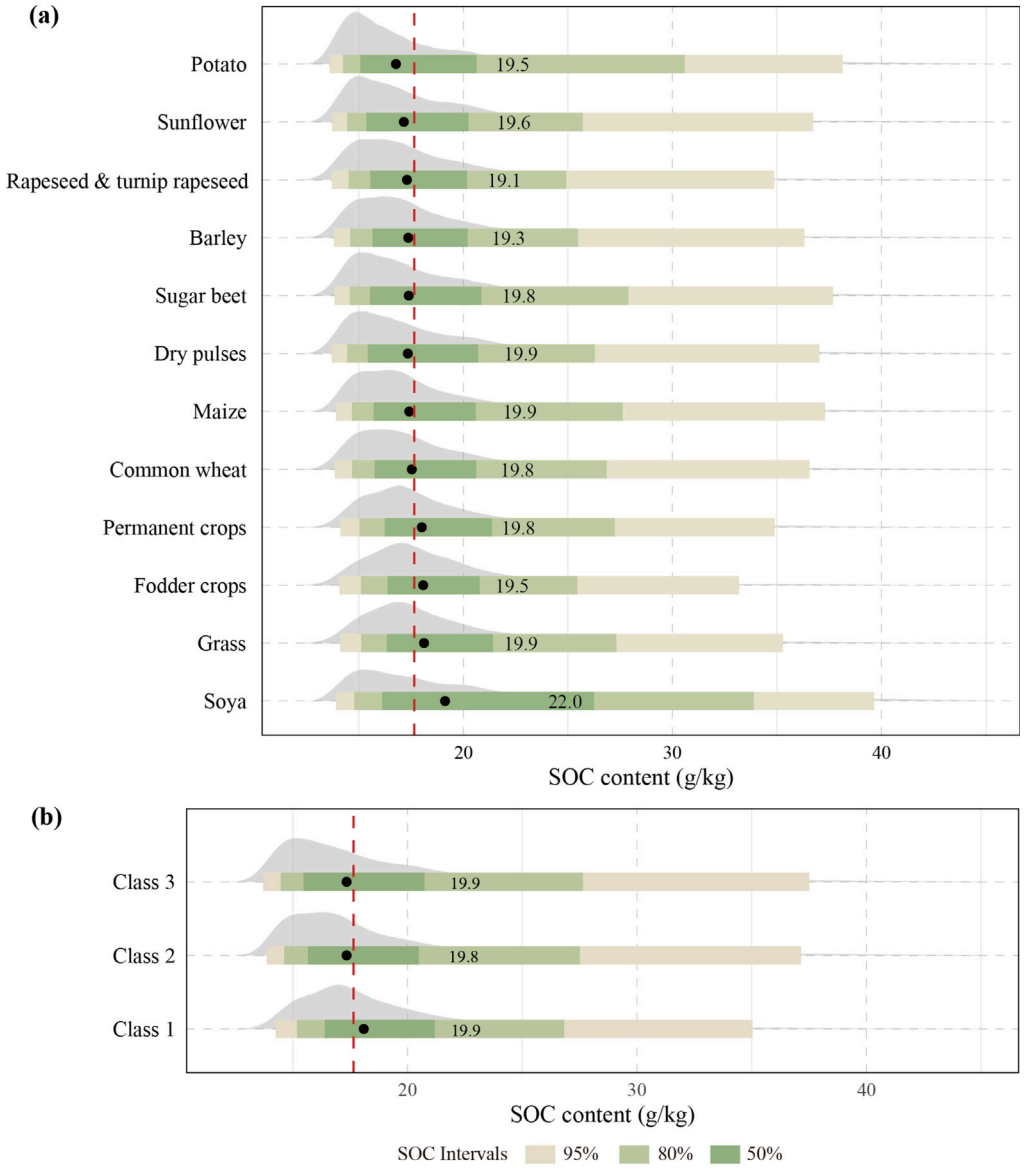

Fig. 4. Distribution of SOC (g/kg) by main crop type (a) and each bare soil frequency class (b). Ridge represents the density distribution of SOC (g/kg) for each crop or each class; Bands indicate 50%, 80%, and 95% credible intervals; black dots mark the median SOC, and numbers show the mean SOC per crop or per class; the red dashed line marks the overall regional median SOC.

variables of Group 1 presented the best SOC prediction across three machine-learning models. Bare soil reflectance had significant importance for explaining the variations of SOC in machine learning models, especially in the RF, where the relative importance accounted for 43.4% (Fig. A.4). During the vegetation cover period (Group 3), the importance of multispectral bands declined sharply and ranked lowest in both the RF and XGBoost models. This variation indicated that bare soil-based spectral bands captured effective reflectance that directly contributed to SOC prediction. In contrast, due to interference from vegetation cover on the soil surface, spectral bands alone cannot be used to predict SOC with reasonable accuracy. Additionally, the SIs derived from the bare soil period (Group 1) had better performance for SOC prediction than those in the vegetation cover period (Groups 2 and 3). This finding demonstrated that the importance of remote sensing image-based SIs was affected by the land surface dynamics.

In the bare soil composite, non-vegetation indices, including NSSI, NBR2, and CI, played a critical role in SOC prediction (Fig. 2). They were specifically designed to capture bare soil characteristics and were strongly correlated with SOC in Group 1, while in the vegetation cover period or the entire observation period, only NSSI was selected as a predictor by the Boruta algorithm.

Table 3
Results of ART-ANOVA for the effects of bare soil class, crop type, and their interaction on SOC across the three study regions.

Region	Variable	Df	F value	p value
Region 1	BSF Class	2	197.186	<0.001
	Crop type	6	12.344	<0.001
	Interaction	12	1.287	0.219
Region 2	BSF Class	2	98.063	<0.001
	Crop type	5	13.963	<0.001
	Interaction	10	2.202	0.015
Region 3	BSF Class	2	32.055	<0.001
	Crop type	3	11.613	<0.001
	Interaction	6	2.666	0.014

* Note: BSF: Bare soil frequency; Df: Degree of freedom.

It confirmed that NBR2 and CI were tailored for bare soil conditions, highlighting their limited applicability for SOC prediction when vegetation cover interferes with soil reflectance signals. NSSI was the most important VI for SOC prediction regardless of the acquisition time of remote sensing imagery. NSSI effectively separates non-photosynthetic vegetation and bare soil (Tian et al., 2021). The non-photosynthetic vegetation (NPV), including litter, dead vegetation, and roots, significantly influences SOC via altering soil structure, nutrient cycling, and erosion protection (Li and Guo, 2018; Jackson and Prince, 2016). Particularly after harvesting, a large amount of crop residues is one of the main SOC resources for agricultural soils. Therefore, in the bare soil period, NSSI presented a significant impact on SOC prediction. However, there are few previous studies in which the quantification of NPV has been shown to improve SOC prediction in croplands using DSM.

During the vegetation cover period, SOC prediction performance declined not only because vegetation obstructed the soil signal captured by spectral bands but also because conventional vegetation indices, such as NDVI and EVI, failed to provide reliable indirect information about SOC. Their reduced predictive power likely stems from the complex interactions between vegetation cover and soil conditions, such as moisture and roughness, which obscure the relationship between vegetation growth and SOC (Castaldi, 2021). Zeraatpisheh et al. (2022) reported that their results were similar to our study. They illustrated that remote sensing time-series of indices, including NDVI, EVI, and soil-adjusted total vegetation index (SATVI), were not correlated significantly with SOC, and all SIs derived from each month did not have enough potential to increase SOC prediction accuracy. It is suggested that the effectiveness of widely used SIs, such as NDVI and EVI, for SOC prediction should be explored in further studies.

Overall, our results highlighted that the timing of image acquisition has a strong impact on SOC prediction. Bare soil periods provided clear and direct spectral reflectance, while vegetation cover introduced interference that weakens the usefulness of both spectral bands and conventional SIs. Among SIs, indices specifically designed for bare soil, such as NSSI, NBR2, and CI, were most effective, whereas widely used vegetation indices like NDVI and EVI showed limited predictive value, even in the vegetation cover period.

While remote sensing data are important for SOC prediction, other environmental covariates, such as soil properties and topography, also play a crucial role, especially when the contribution of spectral data is limited. In this study, TN was the most important variable in all groups and machine learning models. The significant role has been reported in many previous studies (Deng et al., 2018; Zhang et al., 2022b). TN is closely related to the variation of SOC by changing the nutrient cycling and microbial activities (Wu, 2020). Topographical variables played a crucial role in influencing the distribution of SOC. Especially in Group 3, due to the weak impact of remote sensing data on SOC prediction, terrain factors dominated, accounting for 51.2% of all variables. Topography impacts temperature, water movement, and water distribution, resulting in variations in plant growth, accumulation, and decomposition of organic matter (Maleki et al., 2014; Zhou et al., 2020). The importance of topography for SOC prediction has been reported by Zhou et al. (2020), Ji et al. (2024), Wang et al. (2020).

4.3. Implications of agricultural management for soil organic carbon

Sustainable agricultural management is crucial for maintaining SOC in the Po Plain. In this study, SOC was found to be significantly associated with BSF, suggesting that cropland areas with less frequent bare soil exposure tended to exhibit higher SOC. Lower BSF generally indicates more continuous vegetation or residue cover, which can increase carbon inputs from roots and crop residues, thereby promoting SOC accumulation (Jian et al., 2020; van Wesemael et al., 2024). In addition, lower BSF also means shorter periods of bare soil exposure, which can reduce runoff and erosion and help retain carbon-rich topsoil (Márquez-García et al., 2024; Mzid et al., 2021). Therefore, this negative relationship may provide an interpretable management-related perspective on SOC accumulation in croplands.

Specifically, grass and perennial crops exhibited significantly higher SOC content in the Po Plain. This was likely attributed to their long-term vegetation cover, resulting in low soil disturbance and continuous organic matter inputs. In contrast, potato and sunflower croplands showed consistently lower SOC than other crops in the post-hoc comparisons (Table A.3). This might be related to their shorter growing cycles and extended bare soil periods (Carter et al., 2003). These results highlight that, for short-season crops with long post-harvest bare soil windows, management should explicitly target the fallow interval by maintaining surface cover to narrow this window and potentially lower BSF. Residue retention and cover cropping are two widely adopted practical strategies to

maintain surface cover during otherwise non-vegetated intervals, which can reduce bare soil exposure and provide organic carbon inputs that may support SOC formation through microbial processing (Aertsens et al., 2013; Lal, 2004). The post-hoc analysis within interaction effects supported that common wheat, maize, and sunflower crops had higher SOC content in the low BSF conditions (Class 1) than under the high BSF conditions (Table A.4). Reducing the soil exposure might be beneficial for increasing SOC in cropland systems.

Moreover, well-designed crop rotations provide an additional pathway to maintain surface cover and can contribute to long-term agricultural sustainability and carbon sequestration by enhancing soil structure, nutrient cycling, and pest management (Cortignani and Dono, 2020). In the Po Plain, rotation design may be particularly relevant for SOC management because it offers a practical way to combine cover-maintaining phases with periods of higher disturbance. For example, soya can serve as a valuable rotational crop and show relatively high SOC in our study areas. By introducing a legume phase, rotations can support soil fertility through biological nitrogen fixation and may reduce the need for intensive soil operations in subsequent crops, which is relevant for maintaining SOC (Drinkwater et al., 1998). This is especially important for systems dominated by disturbance-intensive crops such as potato, sugar beet, rapeseed, and maize, which might accelerate SOC mineralization and constrain carbon accumulation (West and Post, 2002; Merante et al., 2017; Piccoli et al., 2016). Therefore, integrating soya (or other legumes) into such rotations could help offset SOC losses by improving nutrient supply and supporting more continuous cover and organic inputs over the rotation cycle. A previous study demonstrated that rotating potatoes with legumes or cereals can break pest cycles and benefit from nitrogen-fixing crops, thereby improving soil fertility and contributing to SOC stability (Nelson et al., 2009).

In recent years, climate change has intensified the frequency and severity of droughts in the Po Plain, increasing pressure on local agricultural systems. Under drier and more variable conditions, prolonged bare soil exposure can exacerbate soil moisture losses and degradation risks, making SOC maintenance more challenging (Caddeo et al., 2019; Xie et al., 2024). In this context, the significant association observed in this study between SOC and BSF indicates that reducing bare soil exposure is a key management leverage point for limiting SOC losses in croplands. Therefore, cropland management strategies that prioritize reducing bare soil exposure should be promoted, particularly in intensively cultivated regions and short-season cropping systems. Management strategies, including residue retention, cover cropping, and diversified rotations, are likely to become increasingly important for sustaining SOC and improving cropland resilience in the Po Plain Perego et al. (2019), Li et al. (2024), Corbari et al. (2019). Overall, using BSF as a management-relevant temporal indicator can support targeted crop planning and land management aimed at enhancing SOC and climate resilience.

4.4. The limitations and prospects

In this study, we further investigated whether adding BSF classes to the trained CatBoost model could enhance SOC prediction, leveraging its ability to handle categorical predictors (Hancock and Khoshgoftaar, 2020). However, the inclusion of BSF into Group 1 did not lead to a significant improvement over the baseline model (R^2 = 0.437, RMSE = 12.485, CCC = 0.560). This result likely reflected that the information captured by BSF overlaps with that contained in the existing spectral variables and indices. Additionally, due to the limited number of samples, although crop type showed significant associations with SOC in the ART-ANOVA analysis, its predictive contribution was constrained by the highly imbalanced distribution of crop categories, with some types being underrepresented or absent from the sample set. The importance of these categorical predictors might be tested for prediction accuracy when they are better represented and processed in the dataset.

Agricultural activities are playing an increasingly important role in shaping SOC dynamics (Herzfeld et al., 2021). Future research could further explore the interactions between BSF and other management variables within machine learning models, ensuring more adequate samples to balance crop type representation using harmonized agricultural statistics or precise crop maps. In addition to acquiring indirect indicators, future studies should combine remote sensing data with field observations, field-based investigations, and management related records to obtain more direct and dynamic information on agricultural activities and crop conditions. For example, variables such as crop yield, crop nitrogen content, and other crop growth characteristics could be incorporated into SOC mapping to better capture management related SOC variation (Huang et al., 2022).

5. Conclusion

This study showed that the acquisition period is critical for remote sensing based SOC mapping in intensively managed croplands. Sentinel-2 composites from the bare soil period provided the most informative signal for SOC prediction, with NSSI consistently emerging as a key predictor, while NDVI and EVI contributed little in this setting. The CatBoost model generated a high-resolution SOC map, revealing that soya fields exhibited the highest SOC levels. In addition, this study introduced BSF as a management-relevant temporal indicator, which was negatively associated with SOC across multiple crop types, with the lowest bare soil frequency class consistently showing higher SOC than higher bare soil frequency classes. Overall, this work highlights the importance of acquisition timing in remote sensing-based SOC mapping. The high-resolution SOC maps and BSF investigation provide new insights into SOC distributions and offer valuable guidance for sustainable agricultural management planning in the Po Plain.

CRediT authorship contribution statement

Xiande Ji: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **R. Venkatesha Prasad:** Writing – review & editing, Supervision. **Binyuan Liu:** Writing – review & editing, Funding acquisition. **Balamuralidhar Purushothaman:** Writing – review & editing, Supervision, Conceptualization. **P.V. Aravind:** Supervision.

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.

Acknowledgments

The authors are particularly grateful to the European Soil Data Center (https://esdac.jrc.ec.europa.eu/projects/lucas) for its support of soil data. This research was supported by the China Scholarship Council PhD program (Xiande Ji and Binyuan Liu).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eti.2026.104971.

Data availability

Data will be made available on request.

References

Aertsens, J., De Nocker, L., Gobin, A., 2013. Valuing the carbon sequestration potential for European agriculture. Land Use Policy 31, 584–594. http://dx.doi.org/10.1016/j.landusepol.2012.09.003.

Ai, J., Zhang, Z., Yang, C., Cao, J., Zhou, Z., Ge, X., Chen, X., Wang, J., 2025. Unveiling the dynamic patterns and driving forces of soil organic carbon in Chinese croplands from 1980 to 2020. Land Degrad. Dev. 36 (10), 3587–3603. http://dx.doi.org/10.1002/ldr.5587.

Batunacun, Wieland, R., Lakes, T., Nendel, C., 2021. Using Shapley additive explanations to interpret extreme gradient boosting predictions of grassland degradation in Xilingol, China. Geosci. Model. Dev. 14 (3), 1493–1510. http://dx.doi.org/10.5194/gmd-14-1493-2021.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Caddeo, A., Marras, S., Sallustio, L., Spano, D., Sirca, C., 2019. Soil organic carbon in Italian forests and agroecosystems: estimating current stock and future changes with a spatial modelling approach. Agricult. Forest. Meterol. 278, 107654. http://dx.doi.org/10.1016/j.agrformet.2019.107654.

Carter, M.R., Kunelius, H.T., Sanderson, J.B., Kimpinski, J., Platt, H.W., Bolinder, M.A., 2003. Productivity parameters and soil health dynamics under long-term 2-year potato rotations in Atlantic Canada. Soil Tillage Res. 72 (2), 153–168. http://dx.doi.org/10.1016/S0167-1987(03)00085-0.

Castaldi, F., 2021. Sentinel-2 and landsat-8 multi-temporal series to estimate topsoil properties on croplands. Remote. Sens. 13 (17), http://dx.doi.org/10.3390/rs13173345.

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019. Soil organic carbon mapping using LUCAS topsoil database and sentinel-2 data: an approach to reduce soil moisture and crop residue effects. Remote. Sens. 11 (18), 2121. http://dx.doi.org/10.3390/rs11182121.

Castro-Franco, M., Costa, J.L., Peralta, N., Aparicio, V., 2015. Prediction of soil properties at farm scale using a model-based soil sampling scheme and random forest. Soil Sci. 180 (2), 74. http://dx.doi.org/10.1097/SS.0000000000000115.

Chaplot, V., Smith, P., 2023. Cover crops do not increase soil organic carbon stocks as much as has been claimed: what is the way forward? Global Change Biol. 29 (22), 6163–6169. http://dx.doi.org/10.1111/gcb.16917.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. In: KDD '16, Association for Computing Machinery, New York, NY, USA, pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Chen, Q., Wang, Y., Zhu, X., 2024. Soil organic carbon estimation using remote sensing data-driven machine learning. PeerJ 12, e17836. http://dx.doi.org/10.7717/peerj.17836.

Chenu, C., Angers, D.A., Barré, P., Derrien, D., Arrouays, D., Balesdent, J., 2019. Increasing organic stocks in agricultural soils: knowledge gaps and potential innovations. Soil Tillage Res. 188, 41–52. http://dx.doi.org/10.1016/j.still.2018.04.011.

Corbari, C., Salerno, R., Ceppi, A., Telesca, V., Mancini, M., 2019. Smart irrigation forecast using satellite LANDSAT data and meteo-hydrological modeling. Agricult. Water. Manag. 212, 283–294. http://dx.doi.org/10.1016/j.agwat.2018.09.005.

Cortignani, R., Dono, G., 2020. Greening and legume-supported crop rotations: An impacts assessment on Italian arable farms. Sci. Total Environ. 734, 139464. http://dx.doi.org/10.1016/j.scitotenv.2020.139464.

Costantini, E., Urbano, F., L'Abate, G., 2004. Soil regions of Italy. CRA-ISSDS, Firenze.

Crowther, T.W., Todd-Brown, K.E.O., Rowe, C.W., Wieder, W.R., Carey, J.C., Machmuller, M.B., Snoek, B.L., Fang, S., Zhou, G., Allison, S.D., Blair, J.M., Bridgham, S.D., Burton, A.J., Carrillo, Y., Reich, P.B., Clark, J.S., Classen, A.T., Dijkstra, F.A., Elberling, B., Emmett, B.A., Estiarte, M., Frey, S.D., Guo, J., Harte, J., Jiang, L., Johnson, B.R., Kröel-Dulay, G., Larsen, K.S., Laudon, H., Lavallee, J.M., Luo, Y., Lupascu, M., Ma, L.N., Marhan, S., Michelsen, A., Mohan, J., Niu, S., Pendall, E., Peñuelas, J., Pfeifer-Meister, L., Poll, C., Reinsch, S., Reynolds, L.L., Schmidt, I.K., Sistla, S., Sokol, N.W., Templer, P.H., Treseder, K.K., Welker, J.M., Bradford, M.A., 2016. Quantifying global soil carbon losses in response to warming. Nature 540 (7631), 104–108. http://dx.doi.org/10.1038/nature20150.

Davis, E., Wang, C., Dow, K., 2019. Comparing sentinel-2 MSI and landsat 8 OLI in soil salinity detection: A case study of agricultural lands in coastal North Carolina. Int. J. Remote Sens. 40 (16), 6134–6153. http://dx.doi.org/10.1080/01431161.2019.1587205.

de Sousa, G.P.B., Bellinaso, H., Rosas, J.T.F., de Mello, D.C., Rosin, N.A., Amorim, M.T.A., dos Anjos Bartsch, B., Cardoso, M.C., Mallah, S., Francelino, M.R., Falcioni, R., Alves, M.R., Demattê, J.A.M., 2024. Assessing soil degradation in Brazilian agriculture by a remote sensing approach to monitor bare soil frequency: Impact on soil carbon. Soil Adv. 2, 100011. http://dx.doi.org/10.1016/j.soilad.2024.100011.

Deng, X., Chen, X., Ma, W., Ren, Z., Zhang, M., Grieneisen, M.L., Long, W., Ni, Z., Zhan, Y., Lv, X., 2018. Baseline map of organic carbon stock in farmland topsoil in East China. Agric. Ecosyst. Environ. 254, 213–223. http://dx.doi.org/10.1016/j.agee.2017.11.022.

Drinkwater, L.E., Wagoner, P., Sarrantonio, M., 1998. Legume-based cropping systems have reduced carbon and nitrogen losses. Nature 396 (6708), 262–265. http://dx.doi.org/10.1038/24376.

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: New 1-Km spatial resolution climate surfaces for Global Land Areas. Int. J. Climatol. 37 (12), 4302–4315. http://dx.doi.org/10.1002/joc.5086.

Garosi, Y., Ayoubi, S., Nussbaum, M., Sheklabadi, M., 2022. Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of Iran. Geoderma Reg. 29, e00513. http://dx.doi.org/10.1016/j.geodrs.2022.e00513.

Gattinger, A., Muller, A., Haeni, M., Skinner, C., Fliessbach, A., Buchmann, N., Mäder, P., Stolze, M., Smith, P., Scialabba, N.E.-H., Niggli, U., 2012. Enhanced top soil carbon stocks under organic farming. Proc. Natl. Acad. Sci. 109 (44), 18226–18231. http://dx.doi.org/10.1073/pnas.1209429109.

Gautam, S., Mishra, U., Scown, C.D., Wills, S.A., Adhikari, K., Drewniak, B.A., 2022. Continental United States may lose 1.8 petagrams of soil organic carbon under climate change by 2100. Glob. Ecol. Biogeogr. 31 (6), 1147–1160. http://dx.doi.org/10.1111/geb.13489.

Guo, Z., Adhikari, K., Chellasamy, M., Greve, M.B., Owens, P.R., Greve, M.H., 2019. Selection of terrain attributes and its scale dependency on soil organic carbon prediction. Geoderma 340, 303–312. http://dx.doi.org/10.1016/j.geoderma.2019.01.023.

Hancock, J.T., Khoshgoftaar, T.M., 2020. CatBoost for big data: An interdisciplinary review. J. Big Data 7 (1), 94. http://dx.doi.org/10.1186/s40537-020-00369-8.

He, X., Yang, L., Li, A., Zhang, L., Shen, F., Cai, Y., Zhou, C., 2021. Soil organic carbon prediction using phenological parameters and remote sensing variables generated from sentinel-2 images. CATENA 205, 105442. http://dx.doi.org/10.1016/j.catena.2021.105442.

He, S., Zhou, L., Xie, H., Tan, S., 2024. Enhancing XGBoost's accuracy in soil organic matter prediction through feature fusion. Paddy Water Environ. 22 (3), 475–489. http://dx.doi.org/10.1007/s10333-024-00980-y.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLOS ONE 10 (6), e0125814. http://dx.doi.org/10.1371/journal.pone.0125814.

Herzfeld, T., Heinke, J., Rolinski, S., Müller, C., 2021. Soil organic carbon dynamics from agricultural management practices under climate change. Earth Syst. Dyn. 12 (4), 1037–1055. http://dx.doi.org/10.5194/esd-12-1037-2021.

Hu, B., Geng, Y., Lin, Y., Ni, H., Xie, M., Wang, N., Hu, J., Zou, Q., Chen, S., Zhou, Y., Li, H., Shi, Z., 2025. Comparison of machine learning and geostatistical methods on mapping soil organic carbon density in regional croplands and visualizing its location-specific dominators via interpretable model. Land Degrad. Dev. 36 (10), 3405–3427. http://dx.doi.org/10.1002/ldr.5573.

Huang, Y., Song, X., Wang, Y.-P., Canadell, J.G., Luo, Y., Ciais, P., Chen, A., Hong, S., Wang, Y., Tao, F., Li, W., Xu, Y., Mirzaeitalarposhti, R., Elbasiouny, H., Savin, I., Shchepashchenko, D., Rossel, R.A.V., Goll, D.S., Chang, J., Houlton, B.Z., Wu, H., Yang, F., Feng, X., Chen, Y., Liu, Y., Niu, S., Zhang, G.-L., 2024. Size, distribution, and vulnerability of the global soil inorganic carbon. Science 384 (6692), 233–239. http://dx.doi.org/10.1126/science.adi7918.

Huang, H., Yang, L., Zhang, L., Pu, Y., Yang, C., Wu, Q., Cai, Y., Shen, F., Zhou, C., 2022. A review on digital mapping of soil carbon in cropland: Progress, challenge, and prospect. Environ. Res. Lett. 17 (12), 123004. http://dx.doi.org/10.1088/1748-9326/aca41e.

Huang, F., Yin, K., Huang, J., Gui, L., Wang, P., 2017. Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. Eng. Geol. 223, 11–22. http://dx.doi.org/10.1016/j.enggeo.2017.04.013.

Jackson, H., Prince, S.D., 2016. Degradation of non-photosynthetic vegetation in a semi-arid rangeland. Remote. Sens. 8 (8), 692. http://dx.doi.org/10.3390/rs8080692.

Jeong, G., Oeverdieck, H., Park, S.J., Huwe, B., Ließ, M., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. CATENA 154, 73–84. http://dx.doi.org/10.1016/j.catena.2017.02.006.

Ji, X., Purushothaman, B., Prasad, R.V., Aravind, P.V., 2024. Developing a digital mapping of soil organic carbon on a national scale using sentinel-2 and hybrid models at varying spatial resolutions. Ecol. Indic. 167, 112654. http://dx.doi.org/10.1016/j.ecolind.2024.112654.

Jian, J., Du, X., Reiter, M.S., Stewart, R.D., 2020. A meta-analysis of global cropland soil carbon changes due to cover cropping. Soil Biol. Biochem. 143, 107735. http://dx.doi.org/10.1016/j.soilbio.2020.107735.

John, R., Chen, J., Giannico, V., Park, H., Xiao, J., Shirkey, G., Ouyang, Z., Shao, C., Lafortezza, R., Qi, J., 2018. Grassland canopy cover and aboveground biomass in Mongolia and Inner Mongolia: spatiotemporal estimates and controlling factors. Remote Sens. Environ. 213 (May), 34–48. http://dx.doi.org/10.1016/j.rse.2018.05.002.

Kalambukattu, J.G., Kumar, S., Arya Raj, R., 2018. Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model. Environ. Earth Sci. 77 (5), 1–14. http://dx.doi.org/10.1007/s12665-018-7367-9.

Keskin, H., Grunwald, S., Harris, W.G., 2019. Digital mapping of soil carbon fractions with machine learning. Geoderma 339, 40–58. http://dx.doi.org/10.1016/j.geoderma.2018.12.037.

Khaire, U.M., Dhanalakshmi, R., 2022. Stability of feature selection algorithm: A review. J. King Saud Univ. - Comput. Inf. Sci. 34 (4), 1060–1073. http://dx.doi.org/10.1016/j.jksuci.2019.06.012.

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

Lamichhane, S., Kumar, L., Adhikari, K., 2021. Digital mapping of topsoil organic carbon content in an Alluvial Plain Area of the Terai region of Nepal. Catena 202 (September 2019), 105299. http://dx.doi.org/10.1016/j.catena.2021.105299.

Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Comput. Environ. Urban Syst. 96, 101845. http://dx.doi.org/10.1016/j.compenvurbsys.2022.101845.

Li, Z., Guo, X., 2018. Non-photosynthetic vegetation biomass estimation in semiarid Canadian mixed grasslands using ground hyperspectral data, landsat 8 OLI, and sentinel-2 images. Int. J. Remote Sens. 39 (20), 6893–6913. http://dx.doi.org/10.1080/01431161.2018.1468105.

Li, S., Li, X., Ge, X., 2025. Prediction and mapping of soil organic carbon in the Bosten Lake oasis based on sentinel-2 data and environmental variables. Int. Soil Water Conserv. Res. 13 (2), 436–446. http://dx.doi.org/10.1016/j.iswcr.2024.12.002.

Li, M., Zhou, S., Shen, S., Wang, J., Yang, Y., Wu, Y., Chen, F., Lei, Y., 2024. Climate-smart irrigation strategy can mitigate agricultural water consumption while ensuring food security under a changing climate. Agricult. Water. Manag. 292, 108663. http://dx.doi.org/10.1016/j.agwat.2023.108663.

Liu, X., Li, S., Wang, S., Bian, Z., Zhou, W., Wang, C., 2022. Effects of farmland landscape pattern on spatial distribution of soil organic carbon in Lower Liaohe Plain of northeastern China. Ecol. Indic. 145, 109652. http://dx.doi.org/10.1016/j.ecolind.2022.109652.

Liu, X., Wang, J., Song, X., 2023. Improving the spatial prediction of soil organic carbon content using phenological factors: a case study in the middle and upper reaches of heihe river basin, China. Remote. Sens. 15 (7), 1847. http://dx.doi.org/10.3390/rs15071847.

Lundberg, S., Lee, S.-I., 2017. A unified approach to interpreting model predictions. http://dx.doi.org/10.48550/arXiv.1705.07874, arXiv:1705.07874.

Maleki, S., Khormali, F., Karimi, A.R., 2014. Mapping soil organic matter using topographic attributes and geostatistic approaches in toshan area, golestan province, Iran. Iran. J. Soil Res. 28 (2), 459–468. http://dx.doi.org/10.22092/ijsr.2014.126214.

Mantena, S., Mahammood, V., Rao, K.N., 2023. Prediction of soil salinity in the upputeru River Estuary catchment, India, using machine learning techniques. Environ. Monit. Assess. 195 (8), 1006. http://dx.doi.org/10.1007/s10661-023-11613-y.

Márquez-García, F., Hayas, A., Peña, A., Ordóñez-Fernández, R., González-Sánchez, E.J., 2024. Influence of cover crops and tillage on organic carbon loss in mediterranean olive orchards. Soil Tillage Res. 235, 105905. http://dx.doi.org/10.1016/j.still.2023.105905.

Meinshausen, N., Bühlmann, P., 2010. Stability selection. J. R. Stat. Soc. Ser. B Stat. Methodol. 72 (4), 417–473. http://dx.doi.org/10.1111/j.1467-9868.2010.00740.x.

Merante, P., Dibari, C., Ferrise, R., Sánchez, B., Iglesias, A., Lesschen, J.P., Kuikman, P., Yeluripati, J., Smith, P., Bindi, M., 2017. Adopting soil organic carbon management practices in soils of varying quality: implications and perspectives in Europe. Soil Tillage Res. 165, 95–106. http://dx.doi.org/10.1016/j.still.2016.08.001.

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. http://dx.doi.org/10.1016/j.geoderma.2017.01.002.

Monteleone, B., Borzí, I., Arosio, M., Cesarini, L., Bonaccorso, B., Martina, M., 2023. Modelling the response of wheat yield to stage-specific water stress in the po plain. Agricult. Water. Manag. 287, 108444. http://dx.doi.org/10.1016/j.agwat.2023.108444.

Mzid, N., Pignatti, S., Huang, W., Casa, R., 2021. An analysis of bare soil occurrence in arable croplands for remote sensing topsoil applications. Remote. Sens. 13 (3), 474. http://dx.doi.org/10.3390/rs13030474.

Nascimento, C.M., de Sousa Mendes, W., Quiñonez Silvero, N.E., Poppiel, R.R., Sayão, V.M., Dotto, A.C., Valadares dos Santos, N., Accorsi Amorim, M.T., Demattê, J.A.M., 2021. Soil degradation index developed by multitemporal remote sensing images, climate variables, terrain and soil atributes. J. Environ. Manag. 277, 111316. http://dx.doi.org/10.1016/j.jenvman.2020.111316.

Nelson, K.L., Lynch, D.H., Boiteau, G., 2009. Assessment of changes in soil health throughout organic potato rotation sequences. Agric. Ecosyst. Environ. 131 (3), 220–228. http://dx.doi.org/10.1016/j.agee.2009.01.014.

Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H., 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. Sci. Total Environ. 804, 150187. http://dx.doi.org/10.1016/j.scitotenv.2021.150187.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS soil, the largest expandable soil dataset for Europe: A review. Eur. J. Soil Sci. 69 (1), 140–153. http://dx.doi.org/10.1111/ejss.12499.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P., 2016. Climate-smart soils. Nature 532 (7597), 49–57. http://dx.doi.org/10.1038/nature17174.

Perego, A., Rocca, A., Cattivelli, V., Tabaglio, V., Fiorini, A., Barbieri, S., Schillaci, C., Chiodini, M.E., Brenna, S., Acutis, M., 2019. Agro-environmental aspects of conservation agriculture compared to conventional systems: A 3-year experience on 20 farms in the Po valley (Northern Italy). Agricult. Sys. 168, 73–87. http://dx.doi.org/10.1016/j.agsy.2018.10.008.

Piccoli, I., Chiarini, F., Carletti, P., Furlan, L., Lazzaro, B., Nardi, S., Berti, A., Sartori, L., Dalconi, M.C., Morari, F., 2016. Disentangling the effects of conservation agriculture practices on the vertical distribution of soil organic carbon. evidence of poor carbon sequestration in North- Eastern Italy. Agric. Ecosyst. Environ. 230, 68–78. http://dx.doi.org/10.1016/j.agee.2016.05.035.

Rahmani, S.R., Ackerson, J.P., Schulze, D., Adhikari, K., Libohova, Z., 2022. Digital mapping of soil organic matter and cation exchange capacity in a low relief landscape using LiDAR data. Agronomy 12 (6), 1338. http://dx.doi.org/10.3390/agronomy12061338.

Rasaei, Z., Bogaert, P., 2019. Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. Geoderma 344, 50–62. http://dx.doi.org/10.1016/j.geoderma.2019.02.031.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.d., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.-L., 2009. Digital soil map of the world. Science 325 (5941), 680–681. http://dx.doi.org/10.1126/science.1175084.

Schlesinger, W.H., Amundson, R., 2019. Managing for soil carbon sequestration: let's get realistic. Global Change Biol. 25 (2), 386–389. http://dx.doi.org/10.1111/gcb.14478.

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C., 2020. Remote sensing for precision agriculture: sentinel-2 improved features and applications. Agronomy 10 (5), http://dx.doi.org/10.3390/agronomy10050641.

Shafizadeh-Moghadam, H., Minaei, F., Talebi-khiyavi, H., Xu, T., Homaee, M., 2022. Synergetic use of multi-temporal sentinel-1, sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. CATENA 212, 106077. http://dx.doi.org/10.1016/j.catena.2022.106077.

Shen, Z., Ramirez-Lopez, L., Behrens, T., Cui, L., Zhang, M., Walden, L., Wetterlind, J., Shi, Z., Sudduth, K.A., Baumann, P., Song, Y., Catambay, K., Viscarra Rossel, R.A., 2022. Deep transfer learning of global spectra for local soil carbon monitoring. ISPRS J. Photogramm. Remote Sens. 188, 190–200. http://dx.doi.org/10.1016/j.isprsjprs.2022.04.009.

Smith, P., 2004. How long before a change in soil organic carbon can be detected? Global Change Biol. 10 (11), 1878–1883. http://dx.doi.org/10.1111/j.1365-2486.2004.00854.x.

Sun, C., Li, J., Liu, Y., Zhao, S., Zheng, J., Zhang, S., 2023. Tracking annual changes in the distribution and composition of saltmarsh vegetation on the Jiangsu coast of China using landsat time series–based phenological parameters. Remote Sens. Environ. 284, 113370. http://dx.doi.org/10.1016/j.rse.2022.113370.

Szul, T., Tabor, S., Pancerz, K., 2021. Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to prediction energy consumption for building heating. Energies 14 (10), http://dx.doi.org/10.3390/en14102779.

Taghizadeh-Mehrjardi, R., Hamzehpour, N., Hassanzadeh, M., Heung, B., Ghebleh Goydaragh, M., Schmidt, K., Scholten, T., 2021. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. Geoderma 399 (March), 115108. http://dx.doi.org/10.1016/j.geoderma.2021.115108.

Takada, M., Mishima, Y., Natsume, S., 2009. Estimation of surface soil properties in peatland using ALOS/PALSAR. Landsc. Ecol. Eng. 5 (1), 45–58. http://dx.doi.org/10.1007/s11355-008-0061-4.

Tian, J., Su, S., Tian, Q., Zhan, W., Xi, Y., Wang, N., 2021. A novel spectral index for estimating fractional cover of non-photosynthetic vegetation using near-infrared bands of sentinel satellite. Int. J. Appl. Earth Obs. Geoinf. 101, 102361. http://dx.doi.org/10.1016/j.jag.2021.102361.

Ugbemuna Ugbaje, S., Karunaratne, S., Bishop, T., Gregory, L., Searle, R., Coelli, K., Farrell, M., 2024. Space-time mapping of soil organic carbon stock and its local drivers: potential for use in carbon accounting. Geoderma 441, 116771. http://dx.doi.org/10.1016/j.geoderma.2023.116771.

Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Martelet, G., Baghdadi, N., Arrouays, D., 2023. Sentinel-2 and sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in central France. Remote. Sens. 15 (9), 2410. http://dx.doi.org/10.3390/rs15092410.

van Wesemael, B., Abdelbaki, A., Ben-Dor, E., Chabrillat, S., d'Angelo, P., Demattê, J.A.M., Genova, G., Gholizadeh, A., Heiden, U., Karlshoefer, P., Milewski, R., Poggio, L., Sabetizade, M., Sanz, A., Schwind, P., Tsakiridis, N., Tziolas, N., Yagüe, J., Žížala, D., 2024. A European soil organic carbon monitoring system leveraging sentinel 2 imagery and the LUCAS soil data base. Geoderma 452, 117113. http://dx.doi.org/10.1016/j.geoderma.2024.117113.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. http://dx.doi.org/10.1016/j.rse.2016.04.008.

Wang, J., Feng, C., Hu, B., Chen, S., Hong, Y., Arrouays, D., Peng, J., Shi, Z., 2023a. A novel framework for improving soil organic matter prediction accuracy in cropland by integrating soil, vegetation and human activity information. Sci. Total Environ. 903, 166112. http://dx.doi.org/10.1016/j.scitotenv.2023.166112.

Wang, Q., Moreno-Martínez, Á., Muñoz-Marí, J., Campos-Taberner, M., Camps-Valls, G., 2023b. Estimation of vegetation traits with kernel NDVI. ISPRS J. Photogramm. Remote Sens. 195, 408–417. http://dx.doi.org/10.1016/j.isprsjprs.2022.12.019.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Sci. Total Environ. 630, 367–378. http://dx.doi.org/10.1016/j.scitotenv.2018.02.204.

Wang, S., Zhuang, Q., Jin, X., Yang, Z., Liu, H., 2020. Predicting soil organic carbon and soil nitrogen stocks in topsoil of forest ecosystems in northeastern China using remote sensing data. Remote. Sens. 12 (7), 1115. http://dx.doi.org/10.3390/rs12071115.

West, T.O., Post, W.M., 2002. Soil organic carbon sequestration rates by tillage and crop rotation. Soil Sci. Am. J. 66 (6), 1930–1946. http://dx.doi.org/10.2136/sssaj2002.1930.

Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J., 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Vancouver BC Canada, pp. 143–146. http://dx.doi.org/10.1145/1978942.1978963.

Wu, J., 2020. Change in soil microbial biomass and regulating factors in an alpine meadow site on the Qinghai-Tibetan Plateau. Soil Sci. Plant Nutr. 66 (1), 177–194. http://dx.doi.org/10.1080/00380768.2019.1705181.

Wu, J., Liu, S., Peng, C., Luo, Y., Terrer, C., Yue, C., Peng, S., Li, J., Wang, B., Shangguan, Z., Deng, L., 2024. Future soil organic carbon stocks in China under climate change. Cell Rep. Sustain. http://dx.doi.org/10.1016/j.crsus.2024.100179.

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J.A., Huete, A.R., Ichii, K., Ni, W., Pang, Y., Rahman, A.F., Sun, G., Yuan, W., Zhang, L., Zhang, X., 2019. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. Remote Sens. Environ. 233 (August), 111383. http://dx.doi.org/10.1016/j.rse.2019.111383.

Xie, E., Chen, J., Peng, Y., Yan, G., Zhao, Y., 2024. Historical and future dynamics of cropland soil organic carbon stocks in an intensive human-impacted area of southeastern China. Agric. Ecosyst. Environ. 372, 109098. http://dx.doi.org/10.1016/j.agee.2024.109098.

Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., 2017. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. ISPRS J. Photogramm. Remote Sens. 123, 1–19. http://dx.doi.org/10.1016/j.isprsjprs.2016.11.001.

Xu, X., Zhai, X., 2023. Mapping soil organic matter content during the bare soil period by using satellite data and an improved deep learning network. Sustainability 15 (1), 323. http://dx.doi.org/10.3390/su15010323.

Yang, L., He, X., Shen, F., Zhou, C., Zhu, A.-X., Gao, B., Chen, Z., Li, M., 2020. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. Soil Tillage Res. 196, 104465. http://dx.doi.org/10.1016/j.still.2019.104465.

Zepp, S., Heiden, U., Bachmann, M., Möller, M., Wiesmeier, M., van Wesemael, B., 2023. Optimized bare soil compositing for soil organic carbon prediction of topsoil croplands in Bavaria using landsat. ISPRS J. Photogramm. Remote Sens. 202, 287–302. http://dx.doi.org/10.1016/j.isprsjprs.2023.06.003.

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma 338, 445–452. http://dx.doi.org/10.1016/j.geoderma.2018.09.006.

Zeraatpisheh, M., Galford, G.L., White, A., Noel, A., Darby, H., Adair, E.C., 2023. Soil organic carbon stock prediction using multi-spatial resolutions of environmental variables: how well does the prediction match local references? CATENA 229, 107197. http://dx.doi.org/10.1016/j.catena.2023.107197.

Zeraatpisheh, M., Garosi, Y., Reza Owliaie, H., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., Xu, M., 2022. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. CATENA 208, 105723. http://dx.doi.org/10.1016/j.catena.2021.105723.

Zhang, L., Cai, Y., Huang, H., Li, A., Yang, L., Zhou, C., 2022a. A CNN-LSTM model for soil organic carbon content prediction with long time series of MODIS-based phenological variables. Remote. Sens. 14 (18), 4441. http://dx.doi.org/10.3390/rs14184441.

Zhang, Z., Ding, J., Zhu, C., Chen, X., Wang, J., Han, L., Ma, X., Xu, D., 2021. Bivariate empirical mode decomposition of the spatial variation in the soil organic matter content: A case study from NW China. CATENA 206, 105572. http://dx.doi.org/10.1016/j.catena.2021.105572.

Zhang, Z., Ding, J., Zhu, C., Wang, J., Ge, X., Li, X., Han, L., Chen, X., Wang, J., 2023. Historical and future variation of soil organic carbon in China. Geoderma 436, 116557. http://dx.doi.org/10.1016/j.geoderma.2023.116557.

Zhang, H., Ouyang, Z., Jiang, P., Li, M., Zhao, X., 2022b. Spatial distribution patterns and influencing factors of soil carbon, phosphorus, and C:P ratio on farmlands in southeastern China. CATENA 216, 106409. http://dx.doi.org/10.1016/j.catena.2022.106409.

Zhang, L., Yang, L., Ma, Y., Zhu, A.-X., Wei, R., Liu, J., Greve, M.H., Zhou, C., 2025. Regional-scale soil carbon predictions can be enhanced by transferring global-scale soil–environment relationships. Geoderma 461, 117466. http://dx.doi.org/10.1016/j.geoderma.2025.117466.

Zhao, M.-S., Rossiter, D.G., Li, D.-C., Zhao, Y.-G., Liu, F., Zhang, G.-L., 2014. Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index. Ecol. Indic. 39, 120–133. http://dx.doi.org/10.1016/j.ecolind.2013.12.015.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci. Total Environ. 729, 138244. http://dx.doi.org/10.1016/j.scitotenv.2020.138244.

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A., 2021. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Sci. Total Environ. 755, 142661. http://dx.doi.org/10.1016/j.scitotenv.2020.142661.