

RESEARCH

Open Access



Improving soil organic carbon mapping in farmlands using machine learning models and complex cropping system information

Jianxiong Ou^{1,2}, Zihao Wu^{1,3,4,5*}, Qingwu Yan^{1,2,3,4}, Xiangyang Feng^{1,3} and Zilong Zhao^{1,3}

Abstract

Obtaining accurate spatial maps of soil organic carbon (SOC) in farmlands is crucial for assessing soil quality and achieving precision agriculture. The cropping system is an important factor that affects the soil carbon cycle in farmlands, and different agricultural managements under different cropping systems lead to spatial heterogeneity of SOC. However, current research often ignores differences in the main controlling factors of SOC under different cropping systems, especially when the cropping pattern is complex, which is not conducive to farmland zoning management. This study aims to (i) obtain the spatial distribution map of six cropping systems by using multi-phase HJ-CCD satellite images; (ii) explore the stratified heterogeneous relationship between SOC and environmental variables under different cropping systems by using the Cubist model; and (iii) predict the spatial map of SOC. The Xiantao, Tianmen, and Qianjiang cities, which are the core agricultural areas of the Jiangnan Plain, were selected as the study area. Results showed that the SOC content in rice–wheat rotation was the highest among the six cropping systems. The Cubist model outperformed random forest, ordinary kriging, and multiple linear regression in SOC mapping. The results of the Cubist model showed that cropping system, climate, soil attributes, and vegetation index were important influencing factors of SOC in farmlands. The main controlling factors of SOC under different cropping systems were different. Specifically, summer crop types had a greater influence on spatial variations in SOC than winter crops. Paddy–upland rotation was more affected by river distance and NDVI, while upland–upland rotation was more affected by irrigation-related factors. This work highlights the differentiated main controlling factors of SOC under different cropping systems and provides data support for farmland zoning management. The Cubist model can improve the prediction accuracy of SOC under complex cropping systems.

Keywords Soil organic carbon, Digital soil mapping, Complex cropping system, Cubist model, Winter wheat and rape

*Correspondence:

Zihao Wu

6247@cumt.edu.cn

¹ Observation and Research Station of Ministry of Education for Resource Exhausted Mining Area Land Restoration and Ecological Succession, China University of Mining and Technology, Xuzhou 221116, China

² School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China

³ School of Public Policy and Management, China University of Mining and Technology, Xuzhou 221116, China

⁴ Research Center for Transformation Development and Rural Revitalization of Resource-Based Cities in China, China University of Mining and Technology, Xuzhou 221116, China

⁵ Research Center for Land Use and Ecological Security Governance in Mining Area, China University of Mining and Technology, Xuzhou 221116, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

Soil organic carbon (SOC) in farmland plays a pivotal role in both soil fertility and vegetation growth [1, 2]. The spatial heterogeneity of SOC in farmland is notably pronounced, influenced by diverse factors including climate, topography, soil properties, and human activities [3]. These factors collectively determine the input of SOC in farmland to a significant extent. Obtaining accurate spatial map of SOC in farmlands is conducive by the fact that it facilitates monitoring the changes over time, playing a crucial role in assessing farmland soil quality [4–6].

The current studies often use easy-to-measure environmental variables in prediction of SOC spatial distribution. The philosophy of utilizing environmental variables was established after introducing *SCORPAN* model [7]. It includes soil properties, climate, organisms, topography, parent material, time factor and spatial position [8–10].

In recent years, the accuracy of farmland SOC digital mapping has been remarkably improved by combining farmland planting and management systems, including cropping system, crop type, multiple cropping index, stubble index, and distance to irrigation canals [5, 11–15]. However, existing studies only address typical cropping systems in certain regions, where the systems are relatively straightforward [16–18]. Examples are traditional rice–wheat and wheat–corn rotation systems. Complex cropping systems will exacerbate the spatial heterogeneity of farmland SOC, which brings challenges to SOC spatial prediction.

Soil-landscape prediction models have been numerously performed multiple linear regression (MLR). It is commonly used algorithm in predicting SOC [19]. The model explores the joint influence of multiple variables and is simple, intuitive, and highly interpretable [20]. However, it is limited to linear relationship assumptions and low prediction accuracy [21, 22]. Commonly used machine learning algorithms, such as artificial neural network, support vector machine (SVM), random forest (RF) [23–25], has a high prediction accuracy and can reflect the relative importance of variables. However, these algorithms cannot reveal the relationship between SOC and environmental variables and have poor interpretability [26–29].

The Cubist model is a rule-based predictive model, each rule is associated with a MLR sub-model [30, 31]. This rule and model matching completes the shortcomings of a single model, thereby improving the predictive accuracy of the model. More importantly, it fits nonlinear relationships in the form of stratified linear regression [32–34]. Therefore, the use of Cubist can not only reveal the stratified linear relationship between SOC and environmental variables, but can also contribute to acquiring a high-precision SOC spatial distribution map.

In this study, spatial distribution maps of local cropping systems were derived using multi-period “Environment and Disaster Monitoring and Forecasting Small Satellite Constellation System” (abbreviated as HJ-CCD) images in Tianmen, Qianjiang, and Xiantao cities situated in the Jiangnan Plain. Subsequently, the Cubist model was employed to investigate the stratified linear relationship between SOC and environmental variables within the region. The primary controlling factors influencing SOC under various cropping systems were identified, leading to the creation of a spatial distribution map for SOC.

Materials and methods

Study area and soil samples

The study area is located in the hinterland of the Jiangnan Plain, Hubei Province, China (112° 29′–113° 49′ E, 30° 04′–30° 54′ N). The land was formed by the alluvial deposits of the Yangtze and Han rivers, with a total area of ~ 7133 km². The region has a typical subtropical monsoon climate, which is warm and humid, and has an average annual rainfall of 1135 mm and an average annual temperature of 17.3 °C. Rainfall in the area is mainly concentrated in the summer, accounting for 70% of the total annual precipitation. The terrain is flat, with an average elevation of 30 m above sea level, and the water level drop is small, making the area prone to flooding. The predominant soil in this area is fluvisols, with a small amount of gleysols also present. Rivers crisscross this area, and the soil is fertile, creating favorable natural conditions for crop growth. On this fertile land, crop cultivation is mainly carried out using biannual or triple-season planting methods, making it one of the important grain production areas in Hubei Province.

For this research, a total of 12,041 soil samples were collected from agricultural fields in 2015 (Fig. 1). Due to insufficient preparation of auxiliary variables in the preliminary stage and considering the challenges of accessing certain areas, we opted for a relatively simple random sampling method. The sampling depth was 30 cm. The samples were air-dried in the laboratory. After removing plant roots and gravel, they were pulverized and sieved through a 20-mesh nylon sieve. Soil organic carbon content was determined using Walkley–Black wet oxidation method [35].

Acquisition of environment variables

Based on the *SCORPAN* model [7], 17 environmental variables were selected considering four aspects: soil properties, climate, organism, and spatial position (Table 1). The study area is located in a plain, the terrain undulation is small, and the correlation with SOC is low, so topographic factors are not included in the model. Additionally, residuals in the study area did not exhibit

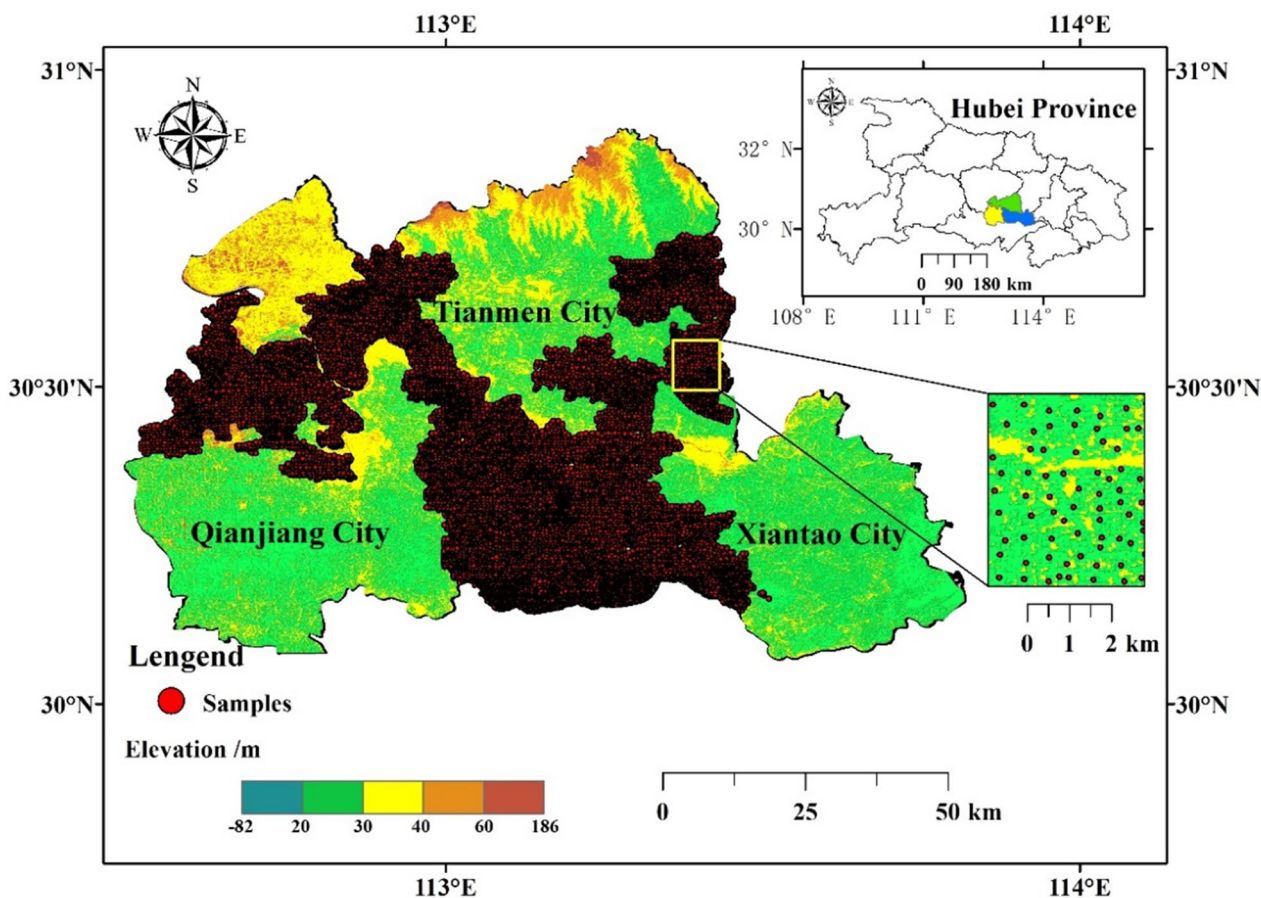


Fig. 1 Location of the study area and sampling points

Table 1 Influencing factors of SOC

Category	Environment variable	Resolution/scale	Source
Soil properties	Soil types	250 m	Soil grids
	Clay content	250 m	Soil grids
	Silt content	250 m	Soil grids
	pH	250 m	Soil grids
	Total nitrogen, TN	90 m	National Earth System Science Data Center
	Total phosphorus, TP	90 m	National Earth System Science Data Center
	Total potassium, TK	90 m	National Earth System Science Data Center
Climate	Mean annual temperature, MAT	30 m	Chengdu Institute of Mountain Hazards and Environment, Chinese Academy of Sciences
	Mean annual precipitation, MAP	30 m	Chengdu Institute of Mountain Hazards and Environment, Chinese Academy of Sciences
Organism	Annual NDVI maximum, NDVI _{max}	30 m	Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
	Annual NDVI average, NDVI _{mean}	1 km	Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
Spatial position	Administrative district (AD)		Based on the boundaries of county-level administrative districts
	Distance from the nearest irrigated canal		Based on land use map
	Distance from the nearest rural settlement		
	Distance from the nearest pond		
	Distance from the nearest river		

spatial correlation, and as a result, no further processing was applied to the residuals. The formula for the SCORPAN model is as follows:

$$S_a \text{ or } S_c = f(s, c, o, r, p, a, n) + \varepsilon,$$

where S_a or S_c represents soil properties or soil type; s represents other soil information at the same point; c represents climatic factors; o represents biological factors; r represents topographic and geomorphological features; p represents soil parent material or lithological characteristics; a represents the time factor of soil formation; n represents spatial location; and ε represents the residuals with spatial autocorrelation.

Acquisition of spatial distribution map of cropping system

The spatial distribution map of the cropping system is based on time series and is calculated from summer and winter crops. By investigating the growth cycles of winter rapeseed and winter wheat, we discovered that the flowering period of winter rapeseed occurs in March to April, while this time coincides with the jointing stage of winter wheat. During this period, winter wheat and winter rapeseed have distinct differences. Therefore, we selected an image (ID: 2365654) from late March taken by the HJ-CCD with no cloud cover. Environmental satellites, characterized by high temporal and spatial resolution, have a spatial resolution of 30 m and a temporal resolution of 2 days, which surpasses what is achievable by other images. They are often specifically utilized for monitoring environmental changes and assessing disaster risks. This environmental satellite image serves as the foundation for obtaining a distribution map of winter crops. From the true-color remote sensing image observation (Fig. 2), the winter rapeseed flowers appear yellow–green, which distinguishes the yellow–green regions as winter rapeseed fields. Winter wheat, being in its jointing stage, shows rapid leaf area growth and appears as a deep green color on the image. Fallow land is depicted in shades of pink–purple. A combination of standard false-color composite images was employed to enhance vegetation characteristics and improve interpretational accuracy. Under the standard false-color composite image, winter rapeseed is depicted as a pale pink color, winter wheat is represented in red, and fallow land appears as a dark green color. The two vertical viewing tools in ENVI software was used to compare the standard false-color composite image and the true-color image to establish areas of interest. The supervised classification method using SVM was applied to classify the images. The classified results were then compared with the validation samples. The overall accuracy and kappa coefficient of the classification were as high as 91.28% and 0.86, respectively. The distribution map of summer crops was

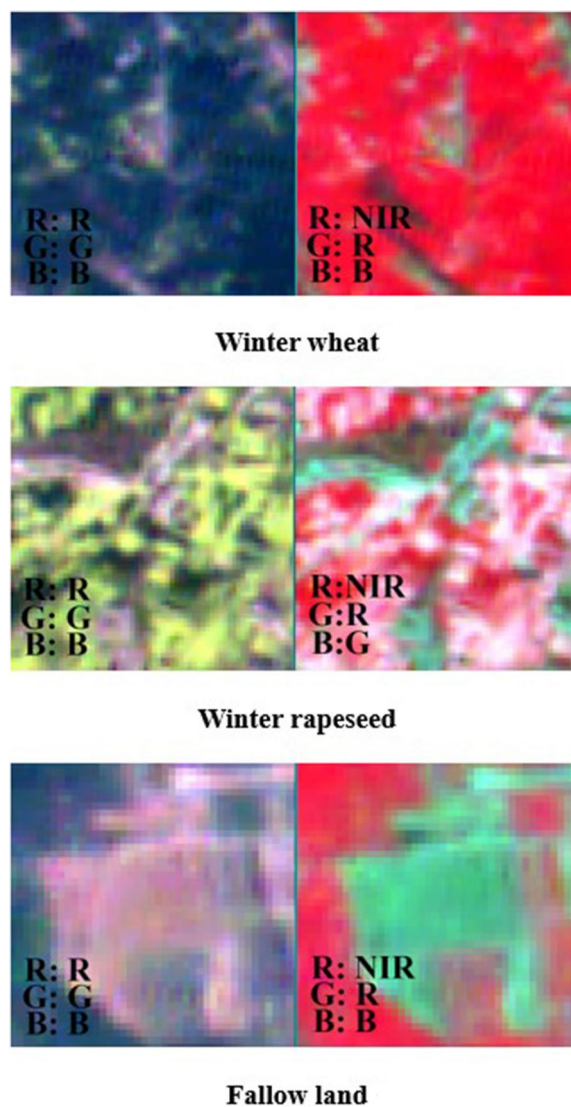


Fig. 2 Remote sensing images of winter crops in true color and standard false color

created using 30 m resolution land use data from Hubei Province, China. The paddy fields and dry land areas were extracted from the land use types to determine the geographic distribution of summer crops. The cropping system distribution map was obtained using the ArcGIS 10.8 platform. Initially, the distribution maps for summer and winter crops were imported into the ArcGIS interface. The two raster images were separately reclassified. The final map was generated using the raster calculator tool (Fig. 3).

Acquisition of other environment variables

Soil types, clay content, silt content, and pH are from the Soil Grids website (<https://www.soilgrids.org/>) with a

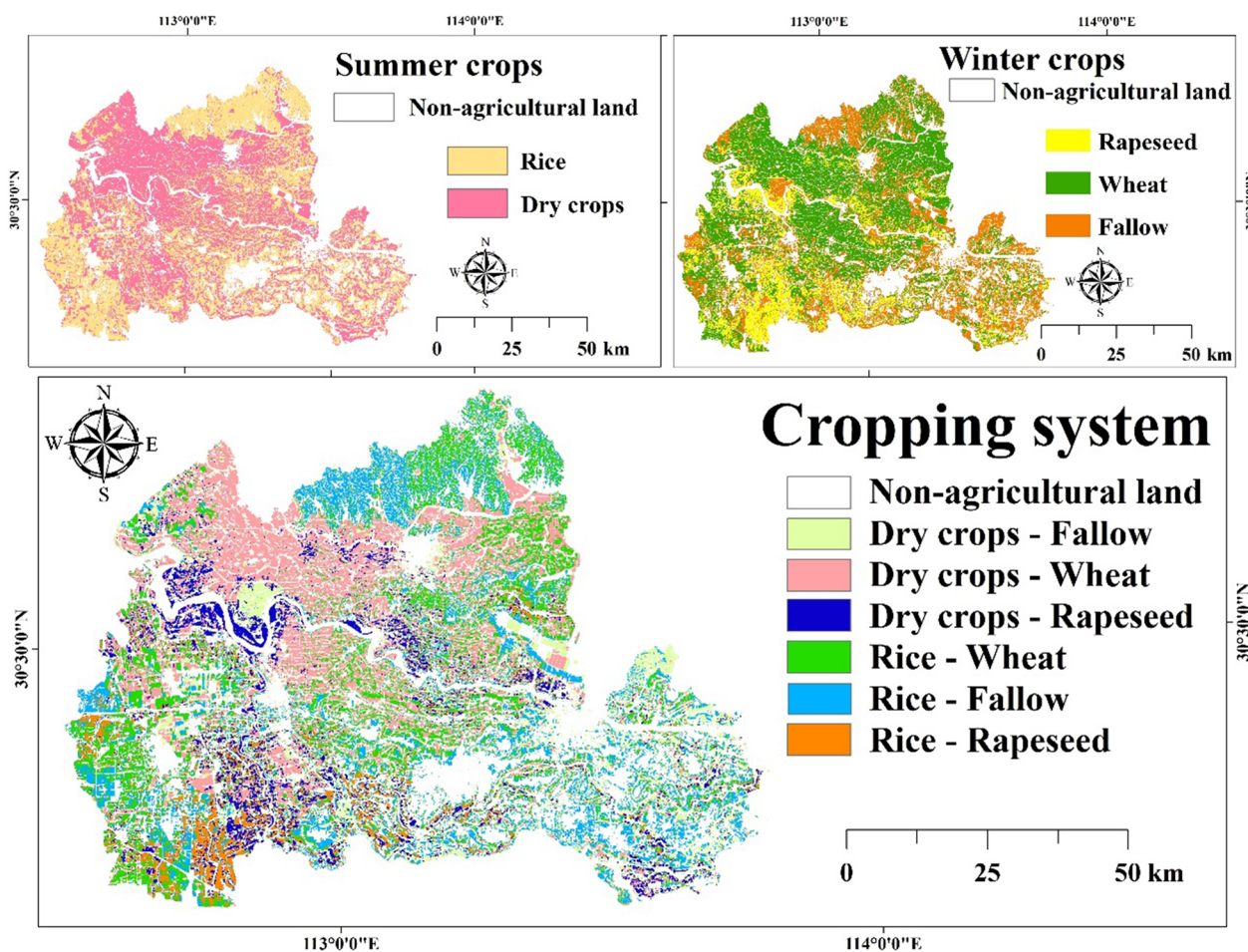


Fig. 3 Spatial distribution of cropping systems

resolution of 250 m. The spatial distribution map of the total contents of nitrogen, phosphorus, and potassium content with a resolution of 90 m was obtained from the National Earth System Science Data Center (<http://soil.geodata.cn/ztsj.html>). The average annual temperature and average annual rainfall data were obtained from the Chengdu Institute of Mountain Hazards and Environment (<https://mp.weixin.qq.com/s/FPBT39rBDGzXe9sdunO-9Q>), Chinese Academy of Sciences (Fig. 4). These meteorological data consist of a 30-year average from 1991 to 2020, and its resolution is 30 m. The maximum resolution of the annual normalized vegetation index is 30 m, and the average resolution of the average annual normalized vegetation index is 1 km, both from the Data Center for Resources and Environmental Sciences (<http://www.resdc.cn/DOI>), Chinese Academy of Sciences. Dis_IC, Dis_River, Dis_RS, and Dis_Pond were calculated on the ArcGIS 10.8 platform by using the 'Near' tool in the 'ArcToolbox' (Table 1). Administrative district data were obtained by filtering the county-level

administrative boundaries through the attribute table by using ArcGIS 10.8.

Spatial prediction model

Ordinary kriging

Kriging is a geostatistical interpolation technique that is an optimal linear unbiased spatial interpolation method [36, 37]. This interpolation method is characterized by the introduction of a semivariogram when estimating the interpolation coefficients to measure the spatial correlation of sample data with distances. Several kriging methods are used in the interpolation formula, and ordinary kriging (OK) is the most commonly used method in resource reserve estimation in kriging method valuation [38]. The estimation formula for OK method is as follows:

$$Z^*(x) = \sum_{i=1}^n \lambda_i Z(x_i),$$

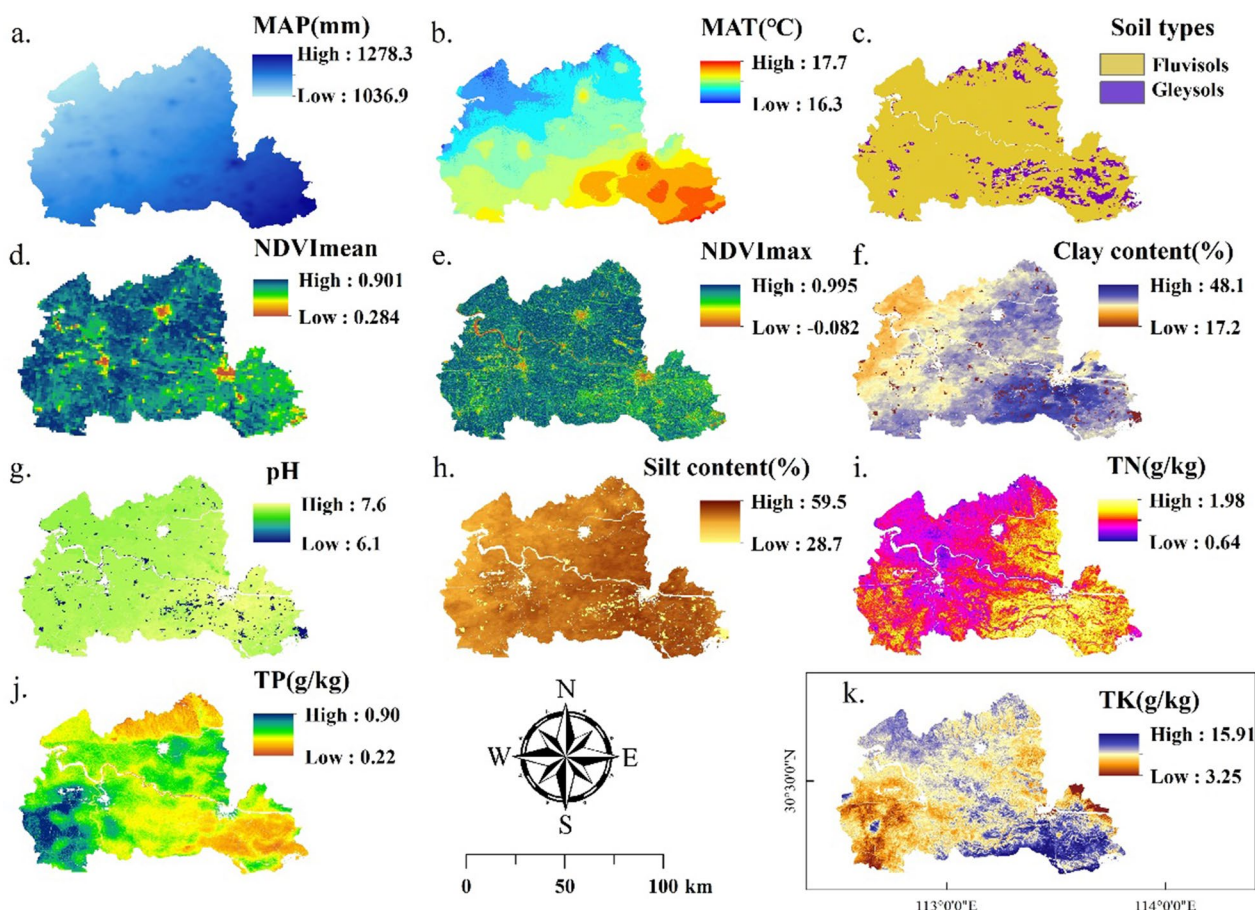


Fig. 4 Spatial distribution of environmental variables. **a** MAP: mean annual precipitation. **b** MAT: mean annual temperature. **c** Soil types. **d** NDVImean: annual NDVI average. **e** NDVImax: annual NDVI maximum. **f** Clay content. **g** pH. **h** Silt content. **i** TN. **j** TP, and **k** TK

where $Z^*(x)$ is the value of the point to be estimated, $Z(x_i)$ is the observation at location x_i , and λ_i is the weight factor of $Z(x_i)$.

Random forest

Random forest (RF), one of the most popular machine learning algorithms, employs the concept of ensemble learning to integrate multiple decision trees. RF can handle not only classification tasks, but also address regression problems; it is now widely applied in estimating agricultural yields [39]. The algorithm performs multiple random selections on the sample set through the specified number of features and the number of decision trees; each new set of samples after random selection corresponds to a decision tree [40, 41]. Subsequently, the results of each decision tree are voted, and that with the most votes is the result of RF. In this study, RF was implemented using the ‘randomForest’ package in R [40]. RF has two key parameters, namely “mtry” and “ntree”, where “mtry” represents the number of features chosen for each tree, and “ntree” represents the total number of

decision trees in the ensemble. In this study, the default parameters were found to be optimal, with the parameters set as follows: ntree = 500, mtry = 5.

Cubist model

Cubist comes from Quinlan’s M5 model decision tree algorithm [30]. The Cubist model splits the dataset by establishing several rules. The principle of splitting is that the prediction error of the model and the dependent variable is minimized, and each subset of data after being split is simulated; a linear regression model of each subset is then established separately [42]. Similar to RF models, the Cubist models employ an ensemble learning strategy, in which the final prediction result is equal to the weighted average of all model tree predictions.

The two important parameters that need to be set when using the Cubist model are “committee” and “neighbors”. The “committees” parameter signifies the number of decision trees used during model construction, while the “neighbors” parameter is exclusive to the prediction phase, representing the number of neighbors to consider

during prediction. The “committees” parameter can be set between 1 and 100, while “neighbors” range from 0 to 9. When predicting a specific point, the final prediction is the sum of the predicted value and the average of the “neighbors” residuals. In the Cubist model, the model can calculate the proportion of independent variables used in all “committee” to showcase the importance of independent variables in the calibration process. These proportions are utilized for rule formulation and model establishment. These proportions include the conditional contribution rate (%) and the modeling contribution rate (%).

The calibration process for Cubist in this study was implemented through the “Cubist” package in R [43]. After validation, the optimal parameter values of “committee” and “neighbors” are 100 and 9, respectively.

Model validation

In this study, 12,041 sampling points were randomly divided into calibration set and validation set, of which 80% sampling points were used for calibration ($n=9633$) and 20% sampling points were used for validation ($n=2408$).

We conducted tenfold cross-validation on the calibration set, employing this cross-validation method to enhance the credibility of the model results. In order to better reflect the performance differences between models, we performed external validation on the validation set. Evaluation metrics such as root mean square error (RMSE), coefficient of determination (R^2), and Lin’s concordance correlation coefficient (LCCC) were chosen for assessment. RMSE values reflect the degree of deviation between the actual and observed values. R^2 verifies the fitting degree of the model to the data, while LCCC considers the accuracy, consistency, and bias of the model. A powerful model typically exhibits low RMSE values as well as R^2 and LCCC values close to 1. The formula for calculating the four evaluation indicators is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n (o_i - P_i)^2},$$

$$R^2 = 1 - \frac{\sum_1^n (o_i - P_i)^2}{\sum_1^n (o_i - \bar{O})^2},$$

$$LCCC = \frac{2rS_O S_P}{S_O^2 + S_P^2 + (\bar{O} - \bar{P})^2},$$

where n is the number of validation samples, o_i is the observation at sample point i , P_i is the predicted value at sample point i , \bar{O} is the average of the observations, \bar{P} is the average of the predicted values, R is the Pearson correlation coefficient between the observed and predicted values, S_O is the standard deviation of the observation, and S_P is the standard deviation of the predicted value.

Results

Descriptive statistics

The descriptive statistics of the sample point data include minimum, maximum, mean, standard deviation, skewness, and kurtosis (Table 2). The SOC varies between 0.60 and 54.20 g/kg, the mean of the divided calibration set and the validation set is not different from the total data set, and the standard deviation of the divided data set is similar to that of the total sample set. The skewness and kurtosis of the dataset are 1.063 and 2.083, respectively. A skewness of 1.063 indicates that the data are generally right-skewed, indicating the existence of some extreme values in the data; a kurtosis of 2.083 indicates that the data distribution is flatter than the normal distribution, indicating few extreme values. The RStudio platform was used to test the normal distribution of K-S, and the results confirmed that the data satisfied the normal distribution.

Soil properties in different cropping systems

This study employed SPSS 24 software for one-way ANOVA to explore variations in SOC content and soil properties under different cropping systems. The SOC content under paddy–upland rotation is ~ 16 g/kg, and that under upland–upland rotation is ~ 12 g/kg. The SOC content in paddy field is significantly higher than that in dry land (Table 3). Duncan’s post hoc test following the one-way ANOVA demonstrated significant differences in SOC content among different cropping systems. Rice–wheat rotation has the highest SOC content. As the TN content increases, the SOC content also increases, suggesting the positive correlation between SOC content

Table 2 Descriptive statistics for soil organic carbon samples

Samples	Number	Min (g/kg)	Max (g/kg)	Mean (g/kg)	SDE	Skewness	Kurtosis
All data	12,041	0.600	54.200	14.190	0.564	1.063	2.083
Calibration set	9633	0.600	54.200	14.180	0.564	1.085	2.164
Validation set	2408	1.200	52.400	14.240	0.560	0.976	1.755

SDE standard deviation of the error

Table 3 Soil properties under different cropping systems

Cropping system	Mean and significance tests				
	SOC (g/kg)	TN (g/kg)	TP (g/kg)	TK (g/kg)	pH
Dry crops–wheat	12.633c	1.197d	0.402a	10.927c	6.936e
Dry crops–fallow	12.193d	1.221c	0.389c	11.105b	7.027c
Dry crops–rapeseed	11.013e	1.169e	0.389c	10.826d	6.951d
Rice–wheat	16.694a	1.251b	0.395b	10.911c	7.027c
Rice–fallow	16.621a	1.287a	0.369d	11.486a	7.089a
Rice–rape-seed	16.162b	1.221c	0.387c	10.916c	7.065b

Duncan’s multiple range test: values with same superscript letters are insignificantly different ($p > 0.05$)

SOC soil organic carbon, TN total nitrogen, TP total phosphorus, TK total potassium

and total nitrogen content. Significant differences in TN content under different cropping systems, with the highest TN content in rice–fallow rotation and the lowest TN content in dry crops–rape rotation. pH varies among the different cropping systems in the study area, with the highest pH value in rice–fallow rotation and the lowest pH value in dry crop–wheat rotation; the pH of the high SOC value area is weakly alkaline.

Cubist model results

Analysis of main controlling factors

Based on the analysis of the calibration process of the Cubist model and the contribution of each environmental variable, variations in conditional contribution were

particularly pronounced. The contribution rates of CS and MAP were high, reaching 85% and 46%, respectively (Fig. 5). The SOC varies under different CS and MAP conditions, making CS and MAP as critical stratifying variables. The top five contributors to the modeling in terms of contribution rates are MAP, TN, Silt, Clay, and NDVImax. Climate and soil attributes were the most influential variables. The importance of MAP was particularly substantial, accounting for 83% of the influence. Prior research highlighted the significance of climate on a global scale [44]. However, this study demonstrates its equally important role at a local level, with precipitation being more significant than temperature in this area. Regarding soil nutrients, the importance of TN content stood out, whereas the importance of TP and TK contents was lower. The influence of pH was found to be the weakest. Soil texture emerged as a key factor influencing SOC, with clay and silt contents playing significant roles in this study. In terms of remote sensing data, NDVImax emerged as a critical variable, while NDVImean’s importance was weaker. Among distance-related factors, the importance of water-related factors such as Dis_IC, Dis_Pond, and Dis_River was higher than that of Dis_RS.

Stratified linear model results

The stratified rules of the Cubist model and the linear regression results for each layer are presented in Fig. 6. Zoning rules were primarily based on summer crop type, precipitation, and NDVImax. Rule 1 corresponds to the cultivation of summer dry crops. The majority of this rule is located in Tianmen City and Qianjiang City, while Xiantao City mostly consists of scattered cultivation. Rule 1 is influenced by irrigation channels and

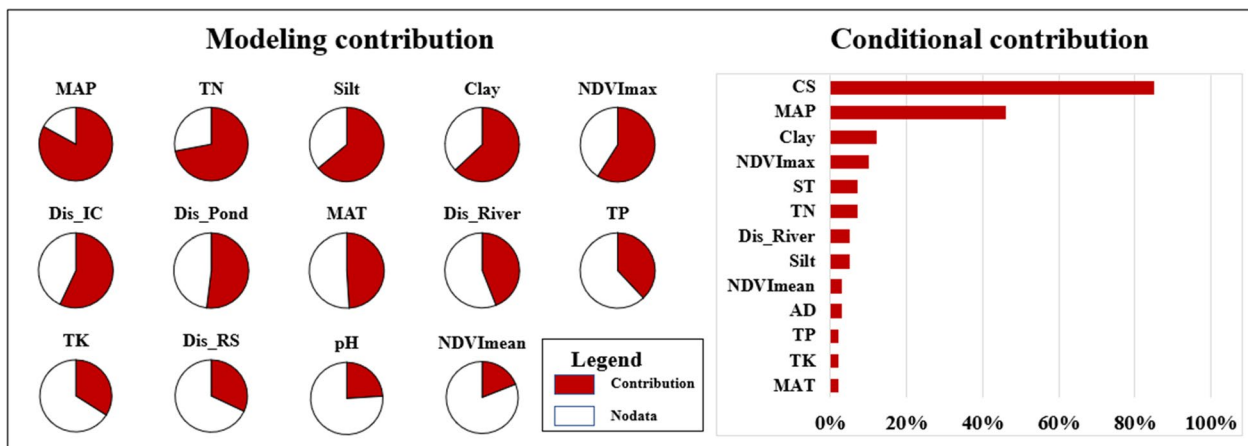
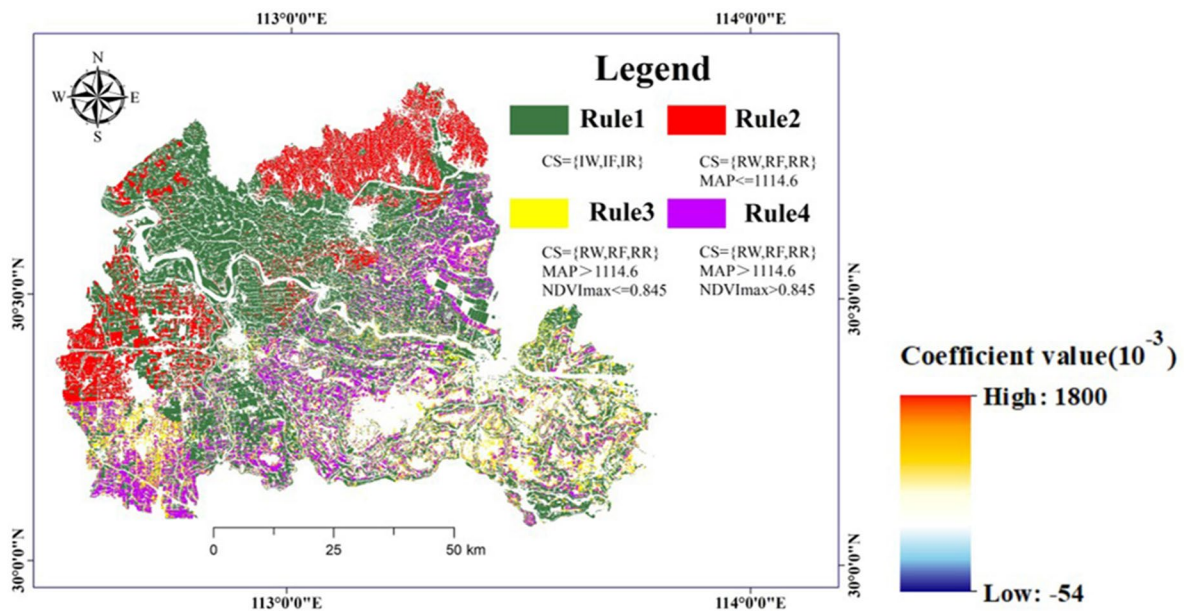


Fig. 5 Cubist importance ranking chart. MAP mean annual precipitation, Silt silt content, Clay clay content, NDVImax annual NDVI maximum, Dis_IC distance from the nearest irrigated canal, Dis_Pond distance from the nearest pond, MAT mean annual temperature, Dis_River distance from the nearest river, Dis_RS distance from the nearest rural settlement, NDVImean annual NDVI average, CS cropping system, ST soil types, AD administrative district



Rule 1	0	14	0	9	9	0	340	-2	130	0	-54	0	-1	1
Rule 2	1	1	0	10	0	1	1580	0	70	1800	0	1	-1	0
Rule 3	0	0	0	13	0	0	1120	-12	100	0	0	0	0	0
Rule 4	-16	10	0	7	-14	0	0	0	1780	0	0	-26	0	0
	Clay	Silt	pH	TN	TP	TK	MAT	MAP	NDVI _{max}	NDVI _{mean}	Dis_IC	Dis_River	Dis_RS	Dis_Pond

Fig. 6 Cubist modeling zoning plot and zoning coefficient plot. *RW* rotation of rice with winter wheat, *RF* rotation of rice with fallow land, *RR* rotation of rice with winter rapeseed, *DW* rotation of dry crops with winter wheat, *DF* rotation of dry crops with fallow land, *DR* rotation of dry crops with winter rapeseed, *Silt* silt content, *Clay* clay content, *MAP* mean annual precipitation, *MAT* mean annual temperature, *Dis_IC* distance from the nearest irrigated canal, *Dis_Pond* distance from the nearest pond, *Dis_River* distance from the nearest river, *Dis_RS* distance from the nearest rural settlement

ponds, whereas Rules 2–4 remain unaffected. Rule 2 is mainly situated in the northern part of Tianmen City and the western part of Qianjiang City. As inferred from the stratified rules, Rules 1 and 2 are also influenced by *Dis_RS*. Rule 3 is predominantly distributed in the eastern part of Xiantao City and the southwestern part of Qianjiang City. The classification of Rules 3 and 4 is primarily based on precipitation, with Rule 3 is more influenced by precipitation than Rule 1. However, Rule 3 is not affected by soil texture. Under the premise of precipitation-based division, Rules 3 and 4 further partition zones based on *NDVI_{max}*. Compared with Rule 3, Rule 4 is significantly influenced by *NDVI_{max}*. These findings indicate the stratified heterogeneity between SOC and environmental variables. Moreover, main controlling factors vary significantly under different cropping systems.

Model evaluation results

RMSE, *R²*, and *LCCC* were used to evaluate the prediction accuracy of the models (Table 4). The Cubist model has the highest *R²* (0.292), followed closely by the RF model (0.263), then the OK model (0.211), while the MLR model has the lowest *R²* (0.207). The Cubist model has the lowest overall deviation level, while the MLR model exhibits the highest deviation, indicating less accurate predictions by the MLR model. When observing the *LCCC* values, the Cubist model still performs the best (0.482). These findings suggest that the Cubist model is optimal, followed by the RF and OK models, while the MLR model demonstrates the weakest performance in the study area.

SOC spatial distribution map

By observing the spatial distribution map of SOC predicted through the Cubist model (Fig. 7), an overall trend

Table 4 Comparison of accuracy among OK, MLR, RF, and Cubist predictive models

	OK	MLR	RF	Cubist
RMSE (g/kg)	0.497	0.499	0.481	0.479
R ²	0.211	0.207	0.263	0.292
LCCC	0.380	0.342	0.431	0.482

OK ordinary kriging, MLR multiple linear regression, RF random forest, RMSE root mean square error, R² coefficient of determination, LCCC Lin's concordance correlation coefficient

can be found in the spatial distribution of SOC in the research area, with higher values in the south and north and lower values in the east and west. Numerous localized high-SOC regions exist in the eastern part. These high-SOC regions are primarily situated in the southern, northern, southeastern, and northeastern parts of the research area, with a few also found in the southwest. The average SOC content in the study area is 13.5 g/kg, ranging from 2.7 g/kg to 34.17 g/kg. These high-SOC areas share a common characteristic: they are all located near

ponds, with this feature being particularly pronounced in the southern and northeastern parts of the study area. Based on analysis of Figs. 2 and 3, the paddy fields have significantly higher SOC values than the dry land, and regions with higher MAP values, such as the eastern part of the study area, exhibit lower SOC content. The spatial distribution maps of clay and silt exhibit similar trends to the SOC spatial distribution map, with areas having higher clay content corresponding to higher SOC content. When analyzing the distribution map of TN in conjunction with SOC, a high TN content corresponds to a high SOC content, consistent with the analysis in Table 3.

Discussion

Relationship between SOC and environmental variables

The results from the Cubist model indicate that the spatial distribution of SOC is influenced by cropping system, climate, soil nutrients, soil texture, and vegetation. The cropping system has the greatest conditional contribution rate.

The climate makes very important contributions in terms of modeling contribution rate and conditional

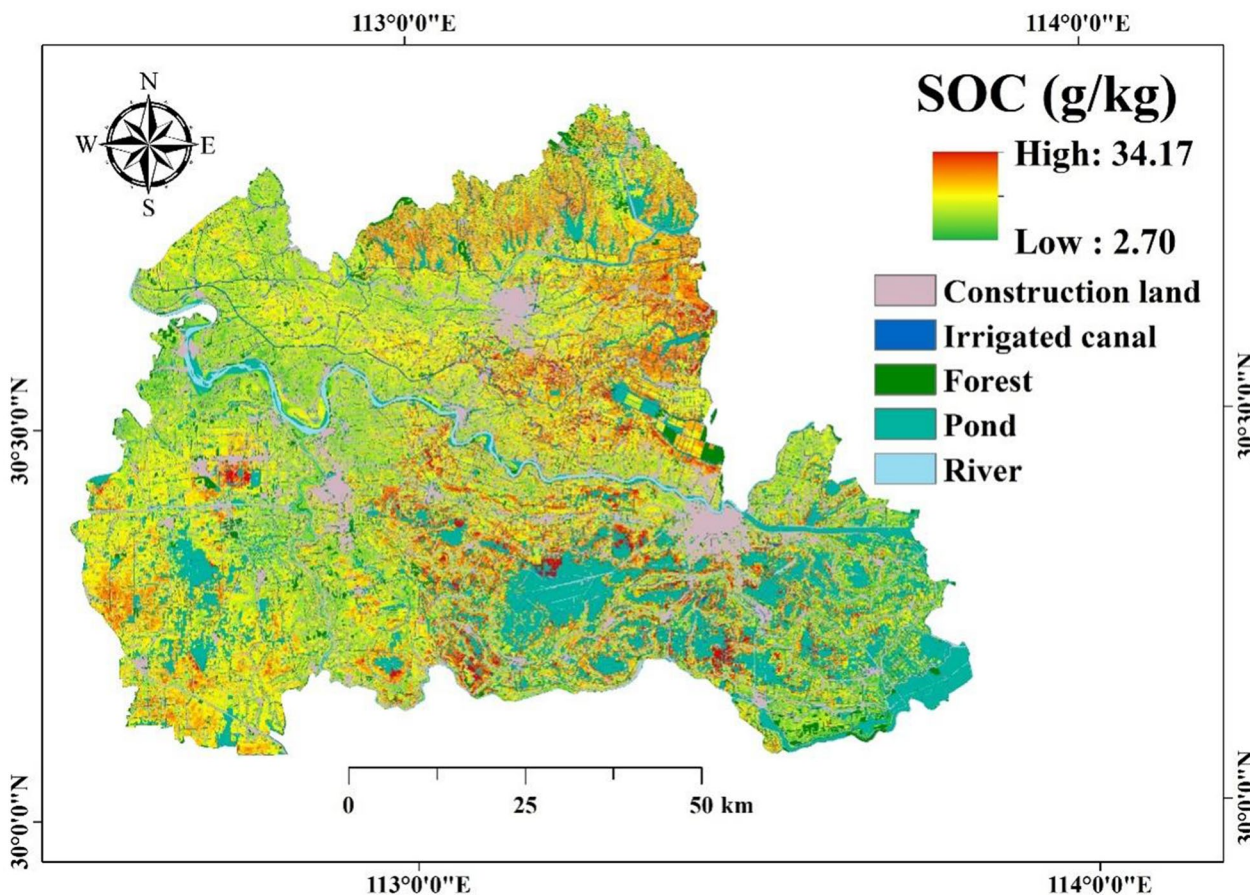


Fig. 7 SOC spatial distribution map via the Cubist model

contribution rate. Numerous large-scale studies found that under natural conditions, SOC is significantly positively correlated with precipitation and negatively correlated with temperature [45–48]. Increasing the temperatures can enhance microbial activity, thereby accelerating the mineralization rate of SOC [49, 50]. However, this result contrasts our study findings, which could be attributed to the relatively small scope of our study area (predominantly consists of cultivated land with well-developed irrigation facilities). In such cases, higher temperatures could enhance vegetation photosynthesis, leading to increased input of crop residues, thereby favoring SOC accumulation. Excessive rainfall might potentially result in soil erosion and leaching of SOC, leading to its volatilization [51]. Given the study area's location between the Yangtze River and the Han River, excessive rainfall could result in the erosion of soil particles carrying SOC.

Soil nutrients and texture are important influencing factors of SOC, and nitrogen to a certain extent promotes the sequestration of SOC. An increase in nitrogen content can promote the production of plant biomass, thereby facilitating the input of carbon into the soil [52, 53]. In the present study, we revealed that an elevated nitrogen content augments the sequestration capacity of SOC. This outcome aligns harmoniously with the conclusions drawn from the majority of antecedent research endeavors [54–56]. Clay and silt have the capability to adsorb organic carbon onto their surfaces, thereby impeding the microbial decomposition of SOC [57, 58]. Additionally, a high clay content enhances soil aggregate stability, leading to a reduction in the mineralization rate of SOC. Consequently, soils characterized by a clayey texture often exhibit elevated levels of organic carbon [59, 60], consistent with our findings.

The vegetation index serves as a crucial metric for assessing vegetation productivity and health status [61–64]. A higher vegetation index, on one hand, indicates improved vegetation growth and higher soil nutrient content. On the other hand, it signifies elevated input of plant residues [65, 66]. A positive correlation exists between the vegetation index and SOC, consistent with the outcomes of the present research.

Planting wheat in winter has the highest SOC content, followed by fallow and planting rape. Winter wheat cultivation during the winter season increases tillage intensity, which may lead to soil structure disruption, hindering the accumulation of SOC [67, 68]. Therefore, the local practice of straw returning has been adopted. This approach can directly increase the content of SOC while improving soil structure and enhancing soil fertility [69–71]. This finding may explain why the SOC content in winter wheat is higher than that in fallow land. At the

same time, fallow has a higher SOC content than rapeseed, which may be related to tillage intensity [72, 73]. In most areas of the study region, a double cropping rice–rape rotation is practiced, and this high-intensity planting often leads to soil nutrient loss. This finding may be a key reason for the low SOC content in rapeseed fields. The carbon sequestration capacity of the rice–wheat rotation system is higher than those in other rotation systems. The above analysis is an important aspect explaining why rice–wheat rotation soil has a higher SOC content. On the other hand, this can be attributed to the benefits of alternate wetting and drying. Such a rotation stimulates carbon and nitrogen cycling, leading to an increase in SOC [74, 75].

Stratified heterogeneous relationship analysis of SOC and impact factor

The Cubist model has the smallest RMSE and possesses the highest R^2 and LCCC values, making it the optimal model for the study area. This rule-based Cubist model effectively reveals the stratified heterogeneity between SOC and environmental variables. Cropping system stands out as a prominent stratifying variable within the Cubist model, with a substantial conditional contribution rate of 85%. The stratification rules indicate varying main controlling factors under different cropping systems.

The differences in cropping systems lead to significant disparities in SOC (Table 3). However, the influence of winter crops appears to be less pronounced, while summer crops predominantly determine the input of SOC. The stratified outcomes from the Cubist model validate this observation. The results highlight that paddy fields exhibit higher SOC content than drylands. This discrepancy could be attributed to the prolonged flooding of paddy fields, leading to poor soil aeration and suppressed decomposition of SOC [76, 77]. Consequently, SOC content in paddy fields tends to be higher than in well-aerated drylands, consistent with previous research outcomes.

The stratified results of the Cubist model also partly reflect a characteristic. During summer, when dry crops are cultivated (upland–upland rotation), they are influenced by factors, such as ponds and irrigation channels, which are water-related. However, when rice is cultivated during the summer (paddy–upland rotation), there is no impact from ponds or irrigation channels. This finding might be closely tied to the land use types. In dry land rotations, the soil moisture content is low, making upland–upland rotations highly sensitive to water. In paddy field rotations, the soil remains flooded year-round, maintaining a high moisture content; thus, it is less responsive to subtle water changes. Regardless of whether it is upland–upland rotation or paddy–upland

rotation, both are influenced by ‘distance from rural residential areas.’ This finding could be attributed to the proximity to residential areas, which facilitates fertilization and irrigation, thereby promoting crop growth and the accumulation of organic carbon. Hence, geographical location to some extent influences SOC dynamics.

Rules 3 and 4 have precipitation exceeding 1114.6 mm and are paddy–upland rotations; the distinction lies in their different vegetation indices. Rule 4 is significantly influenced by the vegetation index compared with Rule 3. The vegetation index is a crucial indicator of vegetation coverage and health. With $NDVI_{max} > 0.845$ in Rule 4, this result indicates a very healthy vegetation state. Under conditions of abundant precipitation, Rule 4 can greatly enhance the input of SOC.

These results underscore a significant relationship between changes in SOC content and different cropping systems. The control of SOC is closely associated with human agricultural management. The main controlling factors vary under different cropping systems. Tailoring agricultural management practices based on the main controlling factors for different regions might contribute to increased crop yields.

Limitations

In this study, the influence of factors such as cropping system on SOC is explored, and mapping is realized based on the quantitative relationship between the factors. The R^2 of the model is only 29.2%, which may be limited by the following points. The year of soil attributes is inconsistent with the sampling year, soil nutrient data are predicted from 2010 to 2018, and soil texture data are predicted in 2019.

The study area falls within China’s significant grain and material production region, the Jiangnan Plain, characterized by frequent human activities. Human activity indicators related to SOC, such as fertilization amount, land ownership, and methods of straw returning, are challenging to quantify spatially and do not effectively reflect changes in SOC [78]. Moreover, the study area has a per capita arable land area of 1.48 mu, with varying cultivation and management practices among different landowners, leading to substantial random errors.

Conclusion

This study investigated the main controlling factors under a complex cropping system and employed Cubist, RF, OK, and MLR models to predict the spatial distribution of SOC. Overall, the SOC content in the study area ranged from 2.70 g/kg to 34.17 g/kg, with the highest content observed in rice–wheat rotation. The Cubist model outperformed other models, indicating its feasibility in explaining SOC variations under intricate cropping

systems. Cropping system, MAP, TN, clay content, silt content and $NDVI_{max}$ as the main controlling factors for farmland SOC, highlighting that lower rainfall, higher soil attributes, and increased vegetation cover contribute to SOC accumulation. The main controlling factors for SOC differed significantly across various cropping systems. Summer crops exhibited a more pronounced impact on SOC spatial variation compared with winter crops. For paddy–upland rotation, factors such as river distance and NDVI played a key role; for upland–upland rotation, irrigation-related factors were more influential. This finding underscores the need for a greater focus on cultivating summer crops and implementing appropriate planting density in paddy–upland rotations as well as considering irrigation factors in upland–upland rotations. This work reveals the variations in main controlling factors of SOC under different cropping systems and highlights the significance of field zoning management.

Abbreviations

SOC	Soil organic carbon
CS	Cropping system
MLR	Multiple linear regression
SVM	Support vector machine
RF	Random forest
OK	Ordinary kriging
MAP	Mean annual precipitation
MAT	Mean annual temperature
ST	Soil types
Clay	Clay content
Silt	Silt content
TN	Total nitrogen
TP	Total phosphorus
TK	Total potassium
$NDVI_{max}$	Annual NDVI maximum
$NDVI_{mean}$	Annual NDVI average
AD	Administrative district
Dis_IC	Distance from the nearest irrigated canal
Dis_RS	Distance from the nearest rural settlement
Dis_River	Distance from the nearest river
Dis_Pond	Distance from the nearest pond
RMSE	Root mean square error
R^2	Coefficient of determination
LCCC	Lin’s concordance correlation coefficient
RW	Rotation of rice with winter wheat
RF	Rotation of rice with fallow land
RR	Rotation of rice with winter rapeseed
DW	Rotation of dry crops with winter wheat
DF	Rotation of dry crops with fallow land
DR	Rotation of dry crops with winter rapeseed
SDE	Standard deviation of the error

Acknowledgements

This research was supported by We are grateful for the data support from the Hubei Geological Survey and Soil SubCenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China.

Author contributions

Conceptualization, J.O. and Z.W.; methodology, J.O., data analysis, J.O.; formal analysis, J.O.; investigation, J.O.; writing—original draft, J.O.; software, J.O. and Z.Z.; writing—review and editing, Z.W.; supervision, Z.W. and Q.Y.; funding acquisition, Z.W. and Q.Y.; project administration, Z.W. and Q.Y.; validation, Q.Y. and Z.Z.; visualization, J.O. and X.F.; data curation, J.O. and X.F. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 42201447), the Graduate Innovation Program of China University of Mining and Technology (2023WLJCRCZL246), the Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX23_1319), the Fundamental Research Funds for the Central Universities (Grant No. 2022-11278) and the Third Comprehensive Scientific Expedition to Xinjiang in China—Geological Hazards and Ecological Environment Investigation of the National Major Energy Channel on the North Slope of Tianshan Mountains (Grant No. 2022xjkk1004).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Institutional review board statement

Not applicable.

Informed consent

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 5 December 2023 Accepted: 14 April 2024

Published online: 21 April 2024

References

- Scholten T, Goebes P, Kühn P et al (2017) On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—a study from SE China. *J Plant Ecol* 10:111–127. <https://doi.org/10.1093/jpe/rtw065>
- Viscarra Rossel RA et al (2016) Baseline estimates of soil organic carbon by proximal sensing: comparing design-based, model-assisted and model-based inference. *Geoderma* 265:152–163. <https://doi.org/10.1016/j.geoderma.2015.11.016>
- Lal R (2004) Soil carbon sequestration impacts on global climate change and food security. *Science* 304:1623–1627. <https://doi.org/10.1126/science.1097396>
- Li M, Han X, Du S, Li L-J (2019) Profile stock of soil organic carbon and distribution in croplands of Northeast China. *CATENA* 174:285–292. <https://doi.org/10.1016/j.catena.2018.11.027>
- Wu Z, Chen Y, Yang Z et al (2022) Mapping soil organic carbon in low-relief farmlands based on stratified heterogeneous relationship. *Remote Sensing* 14:3575–3575. <https://doi.org/10.3390/rs14153575>
- Navidi MN, Seyedmohammadi J (2022) Mapping and spatial analysis of soil chemical effective properties to manage precise nutrition and environment protection. *Int J Environ Anal Chem* 102:1948–1961. <https://doi.org/10.1080/03067319.2020.1746775>
- McBratney AB, Mendonça Santos ML, Minasny B (2003) On digital soil mapping. *Geoderma* 117:3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Luo Z, Feng W, Luo Y et al (2017) Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions. *Glob Change Biol* 23:4430–4439. <https://doi.org/10.1111/gcb.13767>
- Wiesmeier M, Urbanski L, Hobbey E et al (2019) Soil organic carbon storage as a key function of soils—a review of drivers and indicators at various scales. *Geoderma* 333:149–162. <https://doi.org/10.1016/j.geoderma.2018.07.026>
- Xiang Y, Li Y, Liu Y et al (2022) Factors shaping soil organic carbon stocks in grass covered orchards across China: a meta-analysis. *Sci Total Environ* 807:150632. <https://doi.org/10.1016/j.scitotenv.2021.150632>
- Jarecki MK, Lal R (2003) Crop management for soil carbon sequestration. *Crit Rev Plant Sci* 22:471–502. <https://doi.org/10.1080/713608318>
- Li J, Wen Y, Li X et al (2018) Soil labile organic carbon fractions and soil organic carbon stocks as affected by long-term organic and mineral fertilization regimes in the North China Plain. *Soil Tillage Res* 175:281–290. <https://doi.org/10.1016/j.still.2017.08.008>
- Zhang Y, Li X, Gregorich EG et al (2018) No-tillage with continuous maize cropping enhances soil aggregation and organic carbon storage in Northeast China. *Geoderma* 330:204–211. <https://doi.org/10.1016/j.geoderma.2018.05.037>
- Vidojević DD, Manojlović MS, Đorđević AR, et al. Correlations between soil organic carbon, land use and soil type in Serbia: КОРЕЛАЦИЈА ИЗМЕЂУ ОРГАНСКОГ УГЉЕНИКА У ЗЕМЉИ, КОРИШЋЕЊА ЗЕМЉИШТА И ВРСТЕ ОБРАДЛИВОГ ЗЕМЉИШТА У СРБИЈИ. *Matica Srpska J Nat Sci.* 2020;9–18. <https://doi.org/10.2298/ZMSPN2038009V>
- Hashakimana L, Tessema T, Niyitanga F et al (2023) Comparative analysis of monocropping and mixed cropping systems on selected soil properties, soil organic carbon stocks, and simulated maize yields in drought-hotspot regions of Rwanda. *Heliyon* 9:e19041. <https://doi.org/10.1016/j.heliyon.2023.e19041>
- Hu Q, Liu T, Ding H et al (2022) Application rates of nitrogen fertilizers change the pattern of soil organic carbon fractions in a rice–wheat rotation system in China. *Agr Ecosyst Environ* 338:108081. <https://doi.org/10.1016/j.agee.2022.108081>
- Lu J, Zhang W, Li Y et al (2023) Effects of reduced tillage with stubble remaining and nitrogen application on soil aggregation, soil organic carbon and grain yield in maize–wheat rotation system. *Eur J Agron* 149:126920. <https://doi.org/10.1016/j.eja.2023.126920>
- Aumtong S, Chotamonsak C, Glomchinda T (2023) Study of the interaction of dissolved organic carbon, available nutrients, and clay content driving soil carbon storage in the rice rotation cropping system in Northern Thailand. *Agronomy* 13:142. <https://doi.org/10.3390/agronomy13010142>
- Lamichhane S, Kumar L, Wilson B (2019) Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review. *Geoderma* 352:395–413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
- Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis. *Geoderma* 146:102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>
- Zhang H, Wu P, Yin A et al (2017) Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: a comparison of multiple linear regressions and the random forest model. *Sci Total Environ* 592:704–713. <https://doi.org/10.1016/j.scitotenv.2017.02.146>
- Guo L, Fu P, Shi T et al (2020) Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images. *Soil Tillage Res* 196:104477. <https://doi.org/10.1016/j.still.2019.104477>
- Munnaf MA, Mouazen AM (2022) Removal of external influences from on-line vis–NIR spectra for predicting soil organic carbon using machine learning. *CATENA* 211:106015. <https://doi.org/10.1016/j.catena.2022.106015>
- Yang J, Fan J, Lan Z et al (2023) Improved surface soil organic carbon mapping of SoilGrids250m using sentinel-2 spectral images in the Qinghai-Tibetan Plateau. *Remote Sens* 15:114. <https://doi.org/10.3390/rs15010114>
- Martín-López JM, Verchot LV, Martius C, da Silva M (2023) Modeling the spatial distribution of soil organic carbon and carbon stocks in the casanare flooded Savannas of the Colombian Llanos. *Wetlands* 43:65. <https://doi.org/10.1007/s13157-023-01705-3>
- Bui EN, Henderson BL, Viergever K (2006) Knowledge discovery from models of soil properties developed through data mining. *Ecol Model* 191:431–446. <https://doi.org/10.1016/j.ecolmodel.2005.05.021>
- Lacoste M, Minasny B, McBratney A et al (2014) High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma* 213:296–311. <https://doi.org/10.1016/j.geoderma.2013.07.002>
- Adhikari K, Hartemink AE (2015) Digital mapping of topsoil carbon content and changes in the driftless area of Wisconsin, USA. *Soil Sci Soc Am J* 79:155–164. <https://doi.org/10.2136/sssaj2014.09.0392>

29. Khaledian Y, Miller BA (2020) Selecting appropriate machine learning methods for digital soil mapping. *Appl Math Model* 81:401–418. <https://doi.org/10.1016/j.apm.2019.12.016>
30. Quinlan JR (1990) Decision trees and decision-making. *IEEE Trans Syst Man Cybern* 20:339–346. <https://doi.org/10.1109/21.52545>
31. Minasny B, McBratney AB (2008) Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom Intell Lab Syst* 94:72–79. <https://doi.org/10.1016/j.chemolab.2008.06.003>
32. Kuhn M, Weston S, Keefer C et al (2014) Cubist: rule- and instance-based regression modeling. R Package Version 00:18
33. Sun Y, Sun X, Wu Z et al (2023) Using a variety of machine learning approaches to predict and map topsoil pH of arable land on a regional scale. *Soil Sci Soc Am J* 87:613–630. <https://doi.org/10.1002/saj2.20525>
34. Yan Y, Li B, Rossel RV et al (2023) Optimal soil organic matter mapping using an ensemble model incorporating moderate resolution imaging spectroradiometer, portable X-ray fluorescence, and visible near-infrared data. *Comput Electron Agric* 210:107885. <https://doi.org/10.1016/j.compag.2023.107885>
35. Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci* 37:29
36. Goovaerts P. *Geostatistics for natural resources evaluation*. Applied Geo-statistics, 1997.
37. Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89:1–45. [https://doi.org/10.1016/S0016-7061\(98\)00078-0](https://doi.org/10.1016/S0016-7061(98)00078-0)
38. Jia Q, Li W, Che D (2020) A triangulated irregular network constrained ordinary Kriging method for three-dimensional modeling of faulted geological surfaces. *IEEE Access* 8:85179–85189. <https://doi.org/10.1109/ACCESS.2020.2993050>
39. Duarte E, Zagal E, Barrera JA et al (2022) Digital mapping of soil organic carbon stocks in the forest lands of Dominican Republic. *Eur J Remote Sens* 55:213–231. <https://doi.org/10.1080/22797254.2022.2045226>
40. Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
41. Sekulić A, Kilibarda M, Heuvelink GBM et al (2020) Random forest spatial interpolation. *Remote Sens* 12:1687–1687. <https://doi.org/10.3390/rs12101687>
42. Adhikari K, Hartemink AE, Minasny B et al (2014) Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE* 9:e105519. <https://doi.org/10.1371/journal.pone.0105519>
43. Quinlan JR. *Combining instance-based and model-based learning*. In: *Machine Learning Proceedings 1993*. Morgan Kaufmann, San Francisco (CA), 1993. pp 236–243
44. Adhikari K, Mishra U, Owens PR et al (2020) Importance and strength of environmental controllers of soil organic carbon changes with scale. *Geoderma* 375:114472. <https://doi.org/10.1016/j.geoderma.2020.114472>
45. Ou Y, Rousseau AN, Wang L, Yan B (2017) Spatio-temporal patterns of soil organic carbon and pH in relation to environmental factors—a case study of the Black Soil Region of Northeastern China. *Agric Ecosyst Environ* 245:22–31. <https://doi.org/10.1016/j.agee.2017.05.003>
46. Ayala Izurieta JE, Márquez CO, García VJ et al (2021) Multi-predictor mapping of soil organic carbon in the alpine tundra: a case study for the central Ecuadorian páramo. *Carbon Balance Manage* 16:1–19. <https://doi.org/10.1186/s13021-021-00195-2>
47. Sun Y, Ma J, Zhao W et al (2023) Digital mapping of soil organic carbon density in China using an ensemble model. *Environ Res* 231:116131. <https://doi.org/10.1016/j.envres.2023.116131>
48. Yang R-M, Huang L-M, Zhang X et al (2023) Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. *Geoderma* 429:116232. <https://doi.org/10.1016/j.geoderma.2022.116232>
49. Fang X, Zhou G, Qu C et al (2020) Translocating subtropical forest soils to a warmer region alters microbial communities and increases the decomposition of mineral-associated organic carbon. *Soil Biol Biochem* 142:107707. <https://doi.org/10.1016/j.soilbio.2020.107707>
50. Atourakai MRA, Tsozué D, Basga SD et al (2023) Soil organic carbon accumulation in dry tropical mountainous zone of Cameroon. *Arab J Geosci* 16:158. <https://doi.org/10.1007/s12517-023-11248-w>
51. Zhang Z, Zhan T, Li Y et al (2023) Soil organic carbon stock responded more sensitively to degradation in alpine meadows than in alpine steppes on the Qinghai-Tibetan Plateau. *Land Degrad Dev* 34:353–361. <https://doi.org/10.1002/ldr.4463>
52. Chen J, Luo Y, van Groenigen KJ et al (2018) A keystone microbial enzyme for nitrogen control of soil carbon storage. *Sci Adv* 4:eaaq689. <https://doi.org/10.1126/sciadv.aaq1689>
53. Ye C, Chen D, Hall SJ et al (2018) Reconciling multiple impacts of nitrogen enrichment on soil carbon: plant, microbial and geochemical controls. *Ecol Lett* 21:1162–1173. <https://doi.org/10.1111/ele.13083>
54. Wang H, Liu S, Song Z et al (2019) Introducing nitrogen-fixing tree species and mixing with *Pinus massoniana* alters and evenly distributes various chemical compositions of soil organic carbon in a planted forest in southern China. *For Ecol Manage* 449:117477. <https://doi.org/10.1016/j.foreco.2019.117477>
55. Bai J, Zong M, Li S et al (2021) Nitrogen, water content, phosphorus and active iron jointly regulate soil organic carbon in tropical acid red soil forest. *Eur J Soil Sci* 72:446–459. <https://doi.org/10.1111/ejss.12966>
56. Seyedmohammadi J, Navidi MN (2022) Applying fuzzy inference system and analytic network process based on GIS to determine land suitability potential for agricultural. *Environ Monit Assess* 194:712. <https://doi.org/10.1007/s10661-022-10327-x>
57. Zaffar M, Lu S-G (2015) Pore size distribution of clayey soils and its correlation with soil organic matter. *Pedosphere* 25:240–249. [https://doi.org/10.1016/S1002-0160\(15\)60009-1](https://doi.org/10.1016/S1002-0160(15)60009-1)
58. Poeplau C, Zopf D, Greiner B et al (2018) Why does mineral fertilization increase soil carbon stocks in temperate grasslands? *Agric Ecosyst Environ* 265:144–155. <https://doi.org/10.1016/j.agee.2018.06.003>
59. Six J, Conant RT, Paul EA, Paustian K (2002) Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. *Plant Soil* 241:155–176. <https://doi.org/10.1023/A:1016125726789>
60. Razafimbelo TM, Albrecht A, Oliver R et al (2008) Aggregate associated-C and physical protection in a tropical clayey soil under Malagasy conventional and no-tillage systems. *Soil Tillage Res* 98:140–149. <https://doi.org/10.1016/j.still.2007.10.012>
61. Li C, Li X (2019) Hazard rate and reversed hazard rate orders on extremes of heterogeneous and dependent random variables. *Statist Probab Lett* 146:104–111. <https://doi.org/10.1016/j.spl.2018.11.005>
62. Yang L, Song M, Zhu A-X et al (2019) Predicting soil organic carbon content in croplands using crop rotation and Fourier transform decomposed variables. *Geoderma* 340:289–302. <https://doi.org/10.1016/j.geoderma.2019.01.015>
63. Wang S, Zhuang Q, Zhou M et al (2023) Temporal and spatial changes in soil organic carbon and soil inorganic carbon stocks in the semi-arid area of northeast China. *Ecol Ind* 146:109776. <https://doi.org/10.1016/j.ecolind.2022.109776>
64. Seyedmohammadi J, Sarmadian F, Jafarzadeh AA, McDowell RW (2019) Integration of ANP and Fuzzy set techniques for land suitability assessment based on remote sensing and GIS for irrigated maize cultivation. *Arch Agron Soil Sci* 65:1063–1079. <https://doi.org/10.1080/03650340.2018.1549363>
65. Kuzyakov Y, Friedel JK, Stahr K (2000) Review of mechanisms and quantification of priming effects. *Soil Biol* 32:1485
66. Zhang Y, Xu X, Li Z et al (2021) Improvements in soil quality with vegetation succession in subtropical China karst. *Sci Total Environ* 775:145876. <https://doi.org/10.1016/j.scitotenv.2021.145876>
67. Seitz S, Goebeles P, Puerta VL et al (2018) Conservation tillage and organic farming reduce soil erosion. *Agron Sustain Dev* 39:4. <https://doi.org/10.1007/s13593-018-0545-z>
68. Nunes MR, Karlen DL, Veum KS et al (2020) Biological soil health indicators respond to tillage intensity: a US meta-analysis. *Geoderma* 369:114335. <https://doi.org/10.1016/j.geoderma.2020.114335>
69. Liu C, Lu M, Cui J et al (2014) Effects of straw carbon input on carbon dynamics in agricultural soils: a meta-analysis. *Glob Change Biol* 20:1366–1381. <https://doi.org/10.1111/gcb.12517>
70. Zhang X, Wang J, Feng X et al (2023) Effects of tillage on soil organic carbon and crop yield under straw return. *Agric Ecosyst Environ* 354:108543. <https://doi.org/10.1016/j.agee.2023.108543>
71. Zhu C, Zhong W, Han C et al (2023) Driving factors of soil organic carbon sequestration under straw returning across China's uplands. *J Environ Manage* 335:117590. <https://doi.org/10.1016/j.jenvman.2023.117590>

72. Haddaway NR, Hedlund K, Jackson LE et al (2017) How does tillage intensity affect soil organic carbon? A systematic review. *Environ Evid* 6:30. <https://doi.org/10.1186/s13750-017-0108-9>
73. Sandén T, Spiegel H, Stüger H-P et al (2018) European long-term field experiments: knowledge gained about alternative management practices. *Soil Use Manage* 34:167–176. <https://doi.org/10.1111/sum.12421>
74. Wang L, Huang G-Q, Sun D-P, Wang S-B (2018) Crop yield and soil nutrients under paddy-upland multiple cropping rotation systems. *Chin J Ecol* 37:3284–3290. <https://doi.org/10.13292/j.1000-4890.201811.028>
75. Wei L, Ge T, Zhu Z et al (2022) Paddy soils have a much higher microbial biomass content than upland soils: a review of the origin, mechanisms, and drivers. *Agric Ecosyst Environ* 326:107798. <https://doi.org/10.1016/j.agee.2021.107798>
76. Yan X, Zhou H, Zhu QH et al (2013) Carbon sequestration efficiency in paddy soil and upland soil under long-term fertilization in southern China. *Soil Tillage Res* 130:42–51. <https://doi.org/10.1016/j.still.2013.01.013>
77. Wu Z, Liu Y, Han Y et al (2021) Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data. *Sci Total Environ* 754:142120. <https://doi.org/10.1016/j.scitotenv.2020.142120>
78. Zhang Z, Zhang H, Xu E (2022) Enhancing the digital mapping accuracy of farmland soil organic carbon in arid areas using agricultural land use history. *J Clean Prod* 334:130232. <https://doi.org/10.1016/j.jclepro.2021.130232>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.