

RESEARCH

Open Access



Prediction of soil organic carbon stock combining Sentinel-1 and Sentinel-2 images in the Zoige Plateau, the northeastern Qinghai-Tibet Plateau

Junjie Lei¹, Changli Zeng¹, Lv Zhang², Xiaogang Wang³, Chanhua Ma³, Tao Zhou⁴, Benjamin Laffitte¹, Ke Luo⁴, Zhihan Yang⁴ and Xiaolu Tang^{1,5*}

Abstract

Background Soil organic carbon (SOC) is a critical component of the global carbon cycle, and an accurate estimate of regional SOC stock (SOCS) would significantly improve our understanding of SOC sequestration and cycles. Zoige Plateau, locating in the northeastern Qinghai-Tibet Plateau, has the largest alpine marsh wetland worldwide and exhibits a high sensitivity to climate fluctuations. Despite an increasing use of optical remote sensing in predicting regional SOCS, optical remote sensing has obvious limitations in the Zoige Plateau due to highly cloudy weather, and knowledge of on the spatial patterns of SOCS is limited. Therefore, in the current study, the spatial distributions of SOCS within 100 cm were predicted using an XGBoost model—a machine learning approach, by integrating Sentinel-1, Sentinel-2 and field observations in the Zoige Plateau.

Results The results showed that SOC content exhibited vertical distribution patterns within 100 cm, with the highest SOC content in topsoil. The tenfold cross-validation approach showed that XGBoost model satisfactorily predicted the spatial patterns of SOCS with a model efficiency of 0.59 and a root mean standard error of 95.2 Mg ha⁻¹. Predicted SOCS showed a distinct spatial heterogeneity in the Zoige Plateau, with an average of 355.7 ± 123.1 Mg ha⁻¹ within 100 cm and totaled 0.27 × 10⁹ Mg carbon.

Conclusions High SOC content in topsoil highlights the high risks of significant carbon loss from topsoil due to human activities in the Zoige Plateau. Combining Sentinel-1 and Sentinel-2 satisfactorily predicted SOCS using the XGBoost model, which demonstrates the importance of selecting modeling approaches and satellite images to improve efficiency in predicting SOCS distribution at a fine spatial resolution of 10 m. Furthermore, the study emphasizes the potential of radar (Sentinel-1) in developing SOCS mapping, with the newly developed fine-resolution mapping having important applications in land management, ecological restoration, and protection efforts in the Zoige Plateau.

Keywords SOC, Vegetation index, Texture, XGBoost, Sentinel-1, Sentinel-2

*Correspondence:

Xiaolu Tang

lxtt2010@163.com

Full list of author information is available at the end of the article



© 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), which is approximately 2 to 2.5 times higher than the carbon stocks of other biosphere compartments (atmosphere and vegetation) within the 1 m depth range (Lal 2004), represents the largest carbon reservoir of about 1500 Pg C (1 Pg = 10^{15} g) in terrestrial ecosystems globally (Jin et al. 2017). Thus, even minor changes in SOC reservoirs and emissions can significantly impact atmospheric CO₂ concentrations, positively feedback global climate change. Human-induced extreme climate events disrupt the equilibrium of SOC exchange between the atmosphere and soil, making SOC a significant source of excess atmospheric CO₂ emissions (Dash et al. 2019). Although a large number of studies predict SOC distribution at site, regional and global scales (Hengl et al. 2017; Tang et al. 2017), large variabilities still exist in amount and distribution due to different modeling approaches, data sources and spatial resolutions. Global estimates of SOC have greatly advanced our understanding of the amount, distribution and driving mechanisms at a global scale; however, low spatial resolution (> 250 m) of global SOC have constrained their use in small scales that require fine spatial resolution of SOC (< 10 m). Therefore, the development of rapid, non-destructive, and cost-effective detection methods in predicting SOC is crucial for implementing effective soil management measures and optimizing soil utilization to maintain healthy soil ecosystems (Marchant et al. 2015; Six and Paustian 2014).

Remote sensing technology is an effective approach to study the spatial patterns of SOC stock (SOCS) while reducing cost of field sampling and laboratory analysis (Zhou et al. 2023a). Remote sensing-based SOCS is commonly predicted based on the quantitative relationship between field SOCS observations and vegetation indices (Zhou et al. 2020), or spectral bands to understand the spatial patterns of SOCS (Zhou et al. 2021). Increasing availability of satellite sensors, e.g., multispectral and radar, has enriched land surface information for SOCS prediction, and have been validated for mapping SOCS in different ecosystems (Castaldi et al. 2019b; Gholizadeh et al. 2018a; Zhou et al. 2021). The majority of studies have relied on optical satellite sensors, and the commonly derived variables included surface reflectance and vegetation indices (Cao and Ling 2021; Castaldi et al. 2019b). For example, Castaldi et al. (2019b) used Sentinel-2 predicted SOCS in a cropland, while Geng et al. (2024) applied Landsat 9 to assess the spatial variation of SOC in northeast China. However, optical satellites (e.g., Sentinel-2 and Landsat images) suffered from cloud cover and limited penetration depth, which restricts their ability to detect SOC below the soil surface, constraining their applications in SOCS prediction. Therefore, relying

solely on optical sensors might be insufficient for accurately estimating SOC (Tang 2013). On the other hand, radar images possess the ability to penetrate clouds and can provide imagery all-weather, day-and-night supply of imagery of Earth's surface (Ren et al. 2021). Radar sensors (e.g., Sentinel-1) can also obtain valuable information concerning the vegetation canopy and soil. Thus, combining optical satellite images and radar data is regarded as an innovative approach (Zhou et al. 2020; Zhou et al. 2023a). Zhou et al. (2020) leveraged data from Sentinel-1 and Sentinel-2 satellites to assess the potential of various machine learning algorithms in predicting SOCS. Their findings indicate that, compared to individual sensors, a multi-source sensor approach demonstrates greater accuracy in SOCS prediction. Therefore, combining Sentinel-1 and 2 images with a spatial resolution of 10 m offers a nice probability to predict the spatial patterns of SOCS.

In addition to using various remote sensing image data, it is also crucial to select appropriate algorithms to improve the accuracy of predicting SOCS (Ghatasheh et al. 2022). Commonly used machine learning algorithms include Classification and Regression tree, Random Forest, Support Vector Machine, XGBoost (Lei 2019). For instance, Rentschler et al. (2019) explored the potential of Random Forest and Support Vector Machine to predict the spatial distribution of SOCS in horizontal and vertical soil layers and revealed that the exponential depth function combined with the Random Forest method improved the depiction of the spatial distribution of SOCS. Meanwhile, Gebauer et al. (2019) used accelerated regression tree and differential evolution algorithm to predict SOCS of the dry forest ecosystem in southwest Ecuador and found the beneficial impact of the differential evolution algorithm on improving the precision of predictions. Therefore, the application of multiple remote sensing data and machine learning algorithms has large potentials to enhance the predictive accuracy of SOCS.

The Zoige Plateau, located in the northeastern part of the Qinghai-Tibet Plateau, has the largest alpine marsh wetland worldwide, owing to its unique climatic and hydrological conditions, as well as its topographic and soil characteristics (Ma 2013). These wetlands are highly sensitive to climate change and anthropogenic disturbances. Since the 1970s, climate change, artificial drainage, peat extraction and over livestock grazing has reduced wetland areas by more than 30% in the Zoige Plateau (Wu et al. 2011), which may have a significant impact on SOCS. Therefore, estimating SOCS in the Zoige Plateau has been a long-term focus at both site and regional scales. At a site scale, Ye et al. (2023) assessed topsoil (0–10 cm) SOC content and SOCS changes following different vegetation restoration

practices, and Chen et al. (2014) measured SOC content and carbon age down to 6 m at seven sites, while Cao and Ling (2021) predicted the spatial patterns of SOCS within a depth of 0–30 cm using the stepwise regression method combined with vegetation indices, texture characteristics, and topographic features from GF-1 image. Moreover, a recent study estimated the spatial patterns of SOCS within 100 cm using kriging approach in Zoige and Hongyuan County (Ma et al. 2016); however, the relative coarse resolution of 1 km and the lack of field validation of kriging models might constrain the application to land management at a small scale, requiring an urgent SOCS mapping at a fine resolution across the Zoige Plateau.

To fill this substantial knowledge gap, we combined Sentinel-1 and Sentinel-2 to predict the spatial patterns of SOCS using XGBoost, a machine learning approach in the Zoige Plateau. The main objectives were to: (1) Explore the vertical distribution characteristics of SOCS down to 100 cm soil profile; (2) Develop machine learning based model for SOCS prediction; (3) Investigate the spatial distributions of SOCS; and (4) Quantify total SOCS across the Zoige Plateau.

Materials and methods

Study area

The Zoige Plateau, located in the upper reaches of the Yellow River in the northeastern Tibet Plateau, covers an area of approximately 12,000 km² and has an elevation range of 3400 to 3900 m above sea level (Li et al. 2020). Zoige has a plateau cold temperate zone, characterized by sub-cold climatic conditions, high precipitation, and humidity levels, with an annual average precipitation of around 600–800 mm (Jin et al. 2020). The region has an annual average temperature of approximately 1 °C, with the coldest month in January and the warmest month in July. Zoige experiences a long winter from October to April of the following year and a short summer from May to September (Qiu et al. 2009). The vegetation types are alpine meadow and marsh meadow, dominated by *Kobresia kansuensis*, *Carex muliensis*, *Carex lasiocarpa*, *Kobresia setchwanensis*, *Blysmus sinocompressus* and *Elymus nutans*, etc.

Field observations

Before conducting field observations, a pre-selection of field observational sites was conducted based on grassland distributions, vegetation coverages and elevations

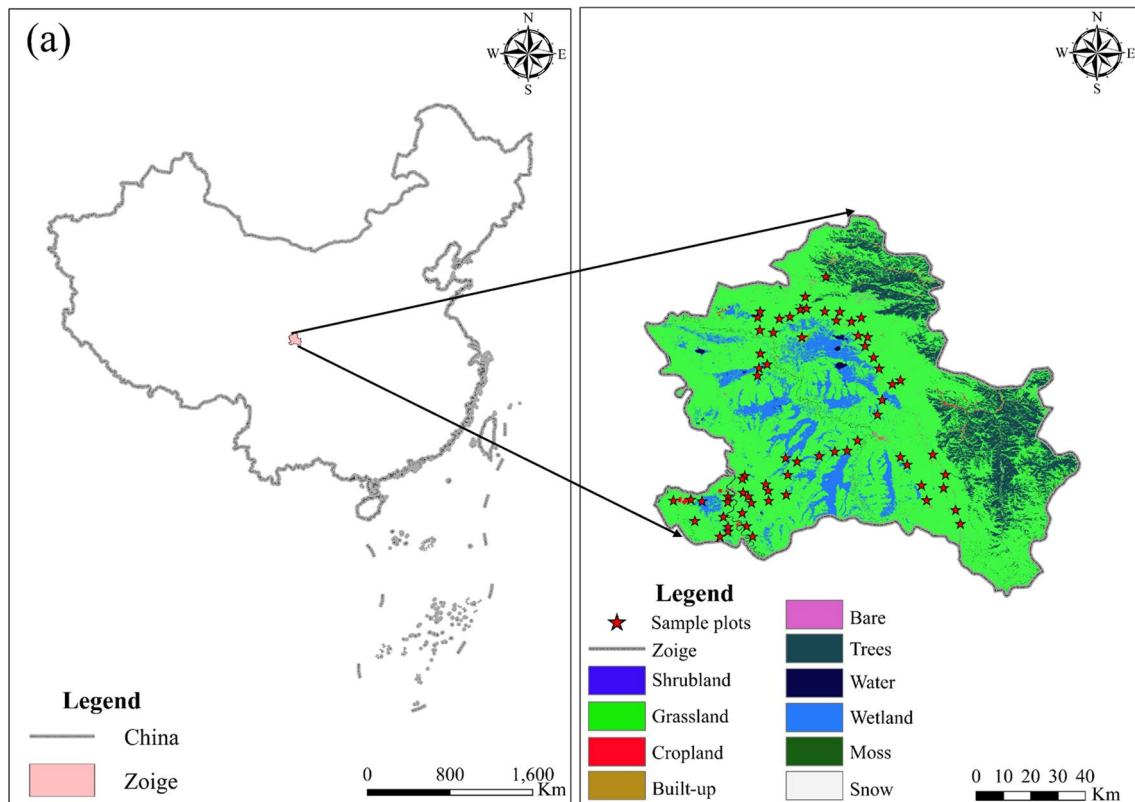


Fig. 1 Location of the study area. The land cover data originated from the European Space Agency’s (ESA) WorldCover 10 m 2020 product (Venter et al. 2022). The cartographic reference number GS(2019)1822

to ensure that our field sites had a broad coverage of the whole Zoige Plateau. Finally, a total of 59 sampling plots (Fig. 1) were selected to ensure their representativeness and the access across the Zoige Plateau in 2019. The general information of vegetation types, vegetation coverages and elevations of field sites were shown in Table 1. Within each site, a square plot with a 1 m × 1 m was set. Soil samples were collected at four layers: 0–10, 10–30, 30–50, and 50–100 cm, which resulted in 236 soil samples in total. The fresh weight of each soil sample was measured using a balance, and the samples were placed in labeled bags. The bags were then sealed and arranged in sequential order in storage baskets for subsequent laboratory analysis involving the removal of roots, stones, and debris. Then, the soil samples were air-dried and finely grounded through a 0.15 mm sieve, and SOC content was measured. Intact soils were collected using a soil ring knife to determine soil bulk density and sand content for each soil sample. SOC content was analyzed using the potassium dichromate heating method (Knicker et al. 2007), then, SOCS (Mg ha⁻¹) was calculated (Huang et al. 2013):

$$SOCS = \sum_{i=1}^{n=k} BD_i \times SOC_i \times D_i \times (1 - S_i) \times 0.1 \quad (1)$$

where BD_i is the soil bulk density of layer i (g cm⁻³); SOC_i is the SOC content of layer i (g kg⁻¹); S_i is sand content of layer i (%); D_i is the soil thickness of layer i (cm).

The Kruskal–Wallis test was employed to examine the significance of the differences in SOC content and SOCS distribution across soil depths.

Remote sensing data and preprocessing

Sentinel-1 is an all-weather radar imaging system, which was developed by the European Commission and the European Space Agency for the Copernicus Global Earth Observation Project. The Sentinel-1 imaging system operates in the C-band and has four imaging modes, providing technical support for long-term monitoring of a given region due to its dual polarization capability, short revisit period, and fast product production (Plank 2014). In this study, two remote sensing images of Sentinel-1 IW GRD acquired on September 18, 2019 with a spatial resolution of 10 m were selected. The ESA software SNAP was used to preprocess the Sentinel-1 images, which included orbit correction, thermal noise removal, radiometric calibration, speckle filtering, and terrain correction. Finally, the VV and VH polarization backscatter coefficients were obtained.

Sentinel-2 is a high-resolution multispectral imaging satellite that carries a Multispectral Imager (MSI) at a height of 786 km with 13 spectral bands. The Sentinel-2 image data is unique in having three specialized red-edge bands, which makes it particularly effective for monitoring vegetation health information. Four remote sensing images of Sentinel-2 L2A were selected on September 18, 2019.

In this study, we selected the VV and VH bands of Sentinel-1 radar data and their corresponding 16 texture indices, as well as B1, B2, B3, B4, B5, B6, B7, B8, B8A, B10, B11, and B12 bands of Sentinel-2 data and their corresponding 96 texture indices (Table 2) as variables. In addition, we calculated eight vegetation indices resulting in a total of 134 variables. Previous studies have

Table 1 General characteristics of observational plots

	Dominant species				
	<i>Carex muliensis</i> , <i>Caltha scaposa</i> , <i>Juncus concinnus</i>	<i>Blymus sinocom</i> , <i>Kobresia capillifolia</i> , <i>Chamaesium thalistrif</i>	<i>Kobresia tibetica</i> , <i>Kobresia capillifolia</i> , <i>Carex ennrvis</i>	<i>Kobresia tibetica</i> , <i>Trollius ranunculoides</i> , <i>Carex muliensis</i>	<i>Elymus nutans</i> , <i>Deyeuxia sylvatica</i> , <i>Festuca chinensis</i>
Elevation (m)					
Maximum	3577	3146	3514	3595	3608
Minimum	3428	3549	3429	3248	3437
Mean	3491	3442	3464	3477	3487
SD	45.47	95	28.2	95.1	50.5
Vegetation cover (%)					
Maximum	98	88	86	84	83
Minimum	88	71	77	78	75
Mean	91	85	84	82	81
SD	4	5	3	3	3

SD, standard deviation

Table 2 Calculations of texture and vegetation indices according to Haralick (1979); Rouse Jr et al. (1974)

Texture index	Formula	Vegetation index	Formula
Mean	$\sum_{i,j=0}^{N-1} iP_{ij}$	NDVI	$\frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$
Variance	$\sum_{i,j=0}^{N-1} iP_{ij}(i - Mean)^2$	EVI	$\frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6.0\rho_{RED} - 7.5\rho_{BLUE} + 1} \times 2.5$
Homogeneity	$\sum_{i,j=0}^{N-1} i \frac{P_{ij}}{1+(i-j)^2}$	RVI	$\frac{\rho_{NIR}}{\rho_{RED}}$
Contrast	$\sum_{i,j=0}^{N-1} iP_{ij}(i - j)^2$	DVI	$\rho_{NIR} - \rho_{RED}$
Dissimilarity	$\sum_{i,j=0}^{N-1} iP_{ij} i - j $	NGBDI	$\frac{\rho_{GREEN} - \rho_{BLUE}}{\rho_{GREEN} + \rho_{BLUE}}$
Entropy	$\sum_{i,j=0}^{N-1} iP_{ij}(-\ln P_{ij})$	MSR	$\frac{\rho_{NIR}}{\rho_{RED}} - 1 / \sqrt{\frac{\rho_{NIR}}{\rho_{RED}} + 1}$
Second moment	$\sum_{i,j=0}^{N-1} iP_{ij}^2$	MSI	ρ_{SWIR} / ρ_{NIR}
Correlation	$\sum_{i,j=0}^{N-1} iP_{ij} \left[\frac{(i-Mean)(j-Mean)}{\sqrt{VA_i VA_j}} \right]$	VARI	$\frac{\rho_{GREEN} - \rho_{RED}}{\rho_{GREEN} + \rho_{RED} - \rho_{BLUE}}$

P_{ij} , row i and column j of the gray level co-occurrence matrix; ρ , the corresponding band of S2 (Sentinel-2) image

suggested that vegetation and texture indices based on band reflectivity can help improve the estimation accuracy of SOCS (Wang et al. 2019b).

Feature selection

Based on the coordinates of each site, texture and vegetation indices were extracted. Before feature selection, SOCS within 100 cm was summed of four layers (0–10, 10–30, 30–50, and 50–100 cm). In order to reduce the calculating cost and improve the modeling efficiency, the gradient boosting algorithm was used for feature selection. One advantage of gradient boosting algorithms is that they allow for obtaining the importance score of each attribute relatively easily after creating the boosting trees. In general, the importance score measures the value of a feature in the decision tree construction of the model. The more an attribute is used to construct decision trees in the model, the higher its importance. The attribute importance is calculated by computing and ranking each attribute in the dataset. In a single decision tree, attribute importance is calculated by the quantity that improves the performance metric by each attribute split point, weighted by the nodes it is responsible for and the number of times it is recorded. That is, the greater the improvement of the performance metric by an attribute at a split point (closer to the root node), the greater the weight assigned to it, and the more important the attribute is as it is selected by more boosting trees. The performance metric can be the Gini purity for selecting split plots or other scoring functions. Finally, the results of each attribute in all boosting trees are weighted and

summed before being averaged to obtain an importance score (Bentéjac et al. 2021; Mayr et al. 2014).

During the feature selection process, the XGBoost model parameters were configured as follows: max_depth=1, eta=8/10, silent=1, objective='reg:linear', nround=150, nthread=2, verbosity=0, etc. Optimal model parameter selection involved iteratively training each parameter value with the optimal seed number, ultimately determining the optimal model parameters. The xgb.importance function was then employed to select the top 15 most important variables from the dependent variables. Subsequently, utilizing the recursive feature elimination principle, the model underwent tenfold cross validation training, initially starting with 15 variables and iteratively discarding the least important variable in each cycle. Finally, the results indicated that the highest efficiency was achieved when six variables were involved in model training (Demir and Şahin 2022; Zhang et al. 2022).

XGBoost model

In this study, we selected the XGBoost model as the core algorithm, which was motivated by its effectiveness in feature selection. XGBoost initially proposed by Tianqi Chen is derived from one of the boosting algorithms and its core idea was to combine classification and regression trees (CART) to form a robust classifier (Chen and Guestrin 2016). This was improved on Gradient Boosting Decision Tree (GBDT), making it more versatile and powerful. Within the XGBoost framework, we assessed the performance and significance of the initial feature

set by constructing decision trees for regression estimation. This approach facilitated the derivation of importance scores for feature variables. Moreover, XGBoost demonstrated considerable computational acceleration, thereby enhancing model efficiency (Zheng et al. 2017). As a result, XGBoost has gained widespread application in various fields due to its high accuracy, parallel processing and portability, stability and lack of overfitting (Chen et al. 2019).

XGBoost has several parameters, and the following are some of the most critical parameters used in this study: (1) Gamma—a minimum loss reduction required to make a further partition on the tree’s leaf nodes; (2) Min_child_weight—the sum of minimum leaf node instance weights; (3) Max_depth—the maximum depth of a single tree; (4) Subsample—the proportion of random samples per tree; (5) nrounds—the maximum boosting iterations; and (6) Eta—controls the learning rate: when adding the contribution of each tree to the current estimate, scale by a factor of $0 < \eta < 1$. The lower the eta value, the higher the nrounds value. A lower eta value means that the model is more robust to overfitting, but the computation speed is slow (Chen et al. 2019).

After the feature selection, the XGBoost model was used to predict the spatial distribution of SOCS. The XGBoost model was trained using the *caret* software package in R, with defined parameters such as *method* = “cv”, *number* = 10, *savePredictions* = ‘final’, etc. During the model training, the best model was selected automatically by *caret*, then *raster::predict* function was applied to predict the spatial pattern of SOCS.

Model efficiency

The model accuracy was evaluated based on tenfold cross-validation. Its principle was to randomly divide the whole dataset into ten nearly equal-sized parts and iteratively use nine of them for training and the remaining one for validation. The accuracy of each validation dataset was used as the evaluation criterion (Fushiki 2011). The final validation result was obtained by averaging the outcomes of tenfold cross validation (Singh and Panda 2011). The model prediction accuracy is evaluated using the root mean square error (*RMSE*, formula 2) and *R*-squared (R^2 , formula 2 and 3). The *RMSE* value reflects the relative dispersion between the predicted value and the observed value, while R^2 indicates the closeness between the predicted value and the observed value. The R^2 value ranges from 0 to 1, and the closer it is to 1, the smaller the *RMSE* value will be.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \tag{2}$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \bar{O}_i)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2} \tag{3}$$

where n represents the number of samples, P_i and O_i represent the predicted and observed SOCS, respectively.

Results

Vertical distribution SOC and SOCS

SOC content decreased with increasing soil depth (Fig. 2a) and exhibited a clear vertical spatial distribution pattern. Mean SOC content decreased from 73.2 g kg⁻¹ for 0–10 cm to 33.9 g kg⁻¹ for 50–100 cm with a weighted average (by depth) of 41.9 g kg⁻¹ within 100 cm. Kruskal–Wallis test showed that soil depth had a significant impact on SOC content ($p < 0.01$). Mean SOC content at 0–10 cm was significantly different with 10–30 cm, 30–50 cm, and 50–100 cm, while no significant difference was observed between the 30–50 cm and 50–100 cm.

SOCS demonstrated different patterns compared to SOC content (Fig. 2b), which tended to increase with soil depth (except 10–30 cm) due to different soil depths. Mean SOCS was 48.8 ± 17.5 Mg ha⁻¹ for 0–10 cm, 80.1 ± 27.0 Mg ha⁻¹ for 10–30 cm, 65.6 ± 33.4 Mg ha⁻¹ for 30–50 cm, and 131.5 ± 90.3 Mg ha⁻¹ for 50–100 cm, respectively. The Kruskal–Wallis test revealed that soil depth had a significant impact on SOCS ($p < 0.01$). Mean SOCS differed significantly among the 0–10, 10–30, 30–50 and 50–100 cm.

Spatial modeling

The gradient boosting algorithm was employed to perform variable selection for SOCS, resulting in the identification of 15 important variables (Fig. 3). After feature selection, the top six variables were chosen to construct the model. Correlation_SAR_VV emerged as the most crucial variable in the model, followed by S2_B12 and Homogeneity_S2_B4 (see Fig. 4, 5).

Spatial distributions

The spatial patterns of SOCS were generally heterogeneous (Fig. 6). The study revealed that areas SOCS near wetlands, forests, and rivers were notably higher, whereas those situated at a longer distance from these areas exhibited lower SOCS levels. Spatially, soils with ample moisture demonstrated high SOCS, while SOCS tended to be low with low soil moisture. Predicted SOCS within ranged from 75 to 660 Mg ha⁻¹ with relative high SOCS

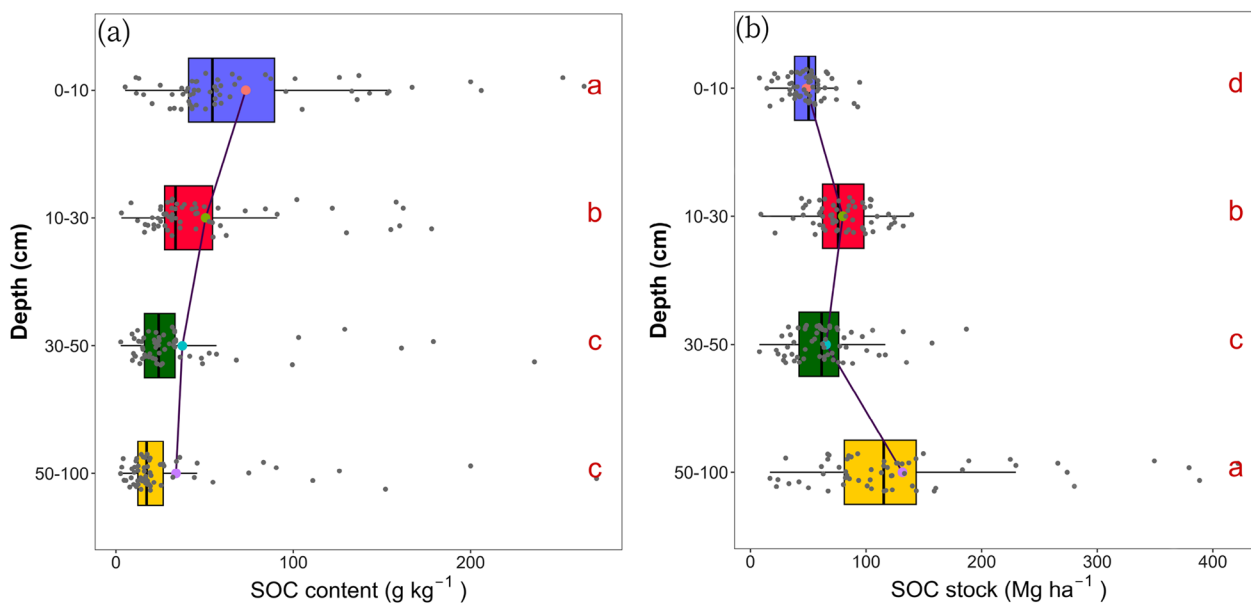


Fig. 2 **a** Relationship between SOC content and soil depth. **b** Relationship between SOCS and soil depth. The letters 'a', 'b', 'c' and 'd' represent the significance of differences, while the line graph represents the average values of different content depths and the grey dots represent the distribution of the values

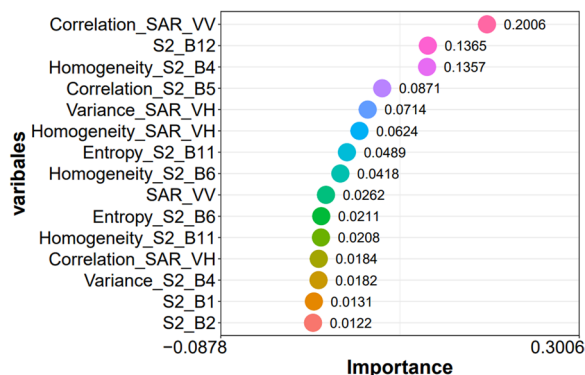


Fig. 3 Ranking of variable importance

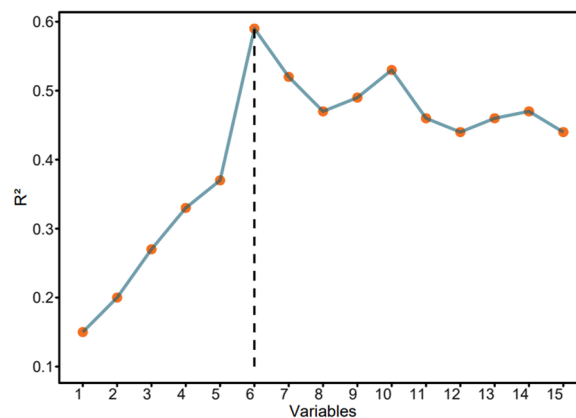


Fig. 4 The impact of increasing the number of variables on R^2 . The results of tenfold cross-validation showed that XGBoost could satisfactorily predict SOCS with a model efficiency of 0.59 with $RMSE$ of 95.24 Mg ha^{-1} . XGBoost tended to overestimate SOCS in areas with low SOCS and underestimate SOCS in areas with high SOCS

value in grassland. Average SOCS within 100 cm was 355.7 Mg ha^{-1} , totaled $0.27 \times 10^9 \text{ Mg}$ carbon across the Zoige Plateau.

Discussion

Vertical variations of SOC content

SOC content showed a decreasing trend with increasing soil depth and significant differences were found among 0–10 cm, 10–30 cm, and 30–50 cm depths ($p < 0.01$, Fig. 2a). These findings are consistent with previous studies conducted by Wei et al. (2023) and Fan et al. (2018). Such result was mainly related to carbon input from vegetation roots (both through root exudates and root mortality) and litter, soil leaching and microbial activities

(Feida et al. 2016). The alpine grassland ecosystem of Zoige was dominated by herbaceous and shrub vegetation, e.g., Cyperaceae and Poaceae, and the majority (86–95%) of root biomass was distributed in topmost 30 cm, with over 75% of vegetation roots concentrating within 10 cm (Li et al. 2004). As the increases of soil depth, the decreasing vegetation roots and oxygen limited carbon input and microbial activities in subsoil layers (Gomes et al. 2019). Therefore, SOC content still experienced a

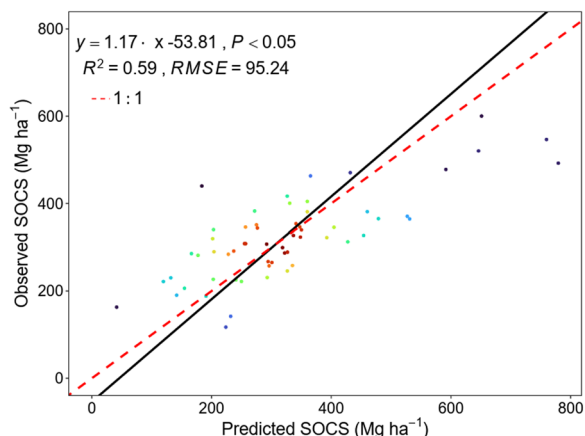


Fig. 5 The correlation between predicted and observed SOCS according to tenfold cross-validation

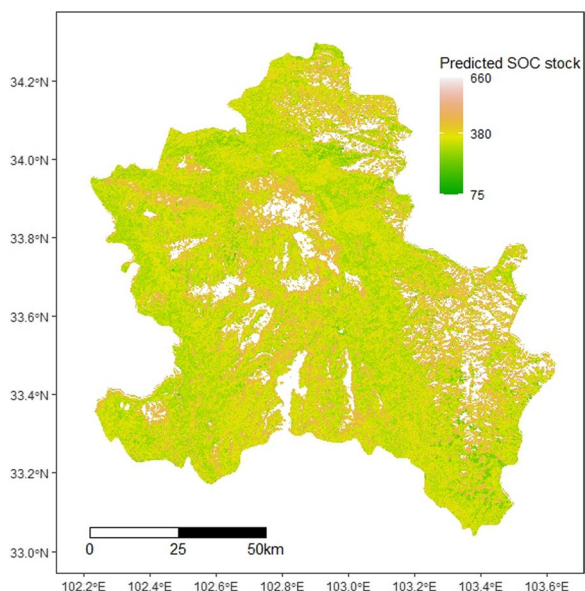


Fig. 6 Spatial distribution of SOCS (Mg ha⁻¹)

decreasing trend from 30–50 cm to 50–100 cm, although the difference in SOC content between 30–50 cm and 50–100 cm was not statistically significant ($p > 0.05$). Although SOCS showed an increasing trend along soil depth, it was related to the different depths of each soil layer. If the depth of each layer was consistent, there was still a decreasing trend with the increasing soil depth with high SOCS in topsoil and low SOCS in subsoil. On the other hand, higher SOC in topsoil layer indicated high risks of large amounts of carbon loss from topsoil when significant human activities occur in the Zoige Plateau. Our results further highlight the urgent soil protection from over grazing and sandification under warming climate.

SOC content was 73.2, 50.6, 37.4 and 33.9 g kg⁻¹ for 0–10 cm, 10–30 cm, 30–50 cm and 50–100 cm, respectively, which were generally higher than mean SOC content from 0–100 cm reported in previous studies, with 23.31 g kg⁻¹ in Heihe River Basin (Wei et al. 2023), 12.87 g kg⁻¹ in the Loess Plateau (Yu et al. 2019) and 12.09 g kg⁻¹ in the Tuojiang River Basin (Wang et al. 2023). High SOC content was mainly resulted from low temperature in the Zoige Plateau. For example, mean temperature was 1 °C with the lowest temperature of – 10.5 to – 7.9 °C in January and the highest temperature of 10.9 to 11.4 °C in July. The low temperature contributed to reduced soil microbial activity and low organic matter decomposition, creating favoring carbon accumulation (Bai et al. 2013; Gao et al. 2007).

Even in the alpine meadows of Zoige, where the soil depth ranges from 50 to 100 cm, the average SOC content (33.9 g kg⁻¹) is significantly higher than the average surface SOC content (14.3 g kg⁻¹) in cultivated soils across China (Li et al. 2022). When compared to other plain ecosystems, SOC content in the alpine meadows of Zoige is consistently high (Cai et al. 2013). For instance, in the Luya mountain typical forest, the surface SOC content can reach up to 29.93 g kg⁻¹, but it declines to near-zero levels at a depth of 100 cm (Wu et al. 2011). In the grassland ecosystems of the Loess Plateau, the surface SOC content is highest at 8.45 g kg⁻¹, while it drops to only 0.99 g kg⁻¹ at a depth of 100 cm (Cheng et al. 2012). Some studies suggest that the distribution of SOC in Zoige can extend as deep as 4 m (Cai et al. 2013). These findings highlight the unique characteristics of SOC distribution in the Zoige Plateau compared to other ecosystems and act as a critical carbon pool in terrestrial ecosystems.

Model and method selection for SOCS modeling

Our results indicated that the overall performance of the XGBoost method was favorable, consistent with previous findings (Wang et al. 2019a; Yu et al. 2020). Zhou et al. (2020a) demonstrated that machine learning algorithms based on boosting methods exhibited significantly superior predictive performance for SOCS compared to random forest (RF) and support vector regression (SVR). This could be attributed to the iterative nature of boosting algorithms, which progressively enhanced prediction accuracy by iteratively improving upon previous results. Moreover, the incorporation of regularization terms in the XGBoost method based on tree complexity reduced model variance, prevented overfitting, and further enhanced prediction accuracy. These factors consistently contributed to superior predictive outcomes obtained with the XGBoost method when compared to traditional boosted regression trees (BRT) and RF methods

in numerous prior studies. However, these conclusions differ from those reported by Mahmoudzadeh et al. (2020), who argued that RF outperformed other methods in terms of predictive performance. This indicates that the uncertainties inherent in XGBoost and other machine learning prediction techniques were frequently influenced by various factors, including the abundance of SOCS in the study region, the choice of environmental variables, and modeling inaccuracies. There was no universally standardized prediction approach that guarantees the optimal performance of predictive models (Gomes et al. 2019; Zhou et al. 2020; Zhou et al. 2023b). Therefore, it was crucial to conduct extensive analysis and testing to determine the optimal prediction method. Furthermore, considering time efficiency aspects, the parallel learning capability supported by the XGBoost method leads to relatively faster model execution speed compared to BRT and RF methods (Chen and Guestrin 2016). Additionally, the findings demonstrated high efficiency exhibited by the XGBoost method in predicting SOCS in the Zoige Plateau.

Besides, the selection of remote sensing imagery played a crucial role in determining the modeling accuracy of SOCS (Gholizadeh et al. 2018b; Zhou et al. 2020). Through comparing various satellite images in different areas, Castaldi et al. (2019a) found that the choice of remote sensing data and study area has varying degrees of influence on prediction accuracy. Earlier studies on SOCS prediction predominantly relied on a single type of sensor, such as Landsat or MODIS. For instance, Vaudour et al. (2013) and Gholizadeh et al. (2018b) examined the potential of Sentinel-2 optical data in predicting soil properties. This is attributed to the strong correlation between soil properties and vegetation cover, where vegetation indices are able to capture variations in soil properties, especially SOCS (Gholizadeh et al. 2018b).

In the present study, the combined use of Sentinel-1 and Sentinel-2 images demonstrated that Correlation_SAR_VV from S1 (Sentinel-1) was the most important variable in predicting SOCS (Fig. 3), because Sentinel-1 imagery can effectively predict SOCS by capturing short-term vegetation changes. Similarly, Yang and Guo (2019) reported a significant correlation between backscattering coefficients obtained from Sentinel-1 images and SOCS. The model construction primarily focused on S1 and S2 band reflectance and their derived indices, while indices like NDVI were not selected by the XGBoost model. Although this finding contradicts the previous belief that NDVI is an important indicator for predicting SOCS, it aligned with the findings of Zhou et al. (2020a). Importantly, the texture index of the Sentinel-1 VV band was identified as the most significant variable, implying that Synthetic Aperture Radar (SAR) data can enhance the

modeling accuracy in predicting SOCS, because S1 and S2 reflectance and derived indices carry more essential information for SOCS than vegetation indices. These results demonstrated that combining optical and radar sensors can effectively improve the modeling accuracy for modeling SOCS, particularly in regions that face cloud cover challenges like the Zoige Plateau. However, this finding contradicts the report by Shafizadeh-Moghadam et al. (2022), who stated that the inclusion of S1 data does not improve the performance of any learning model. This discrepancy may arise from different choices of machine learning algorithms and the significant influence of various types and combinations of environmental variables on the selection of important variables.

Spatial distributions of SOCS

Our study reveals that grassland soils located in wetlands, riverbanks, and forest edges exhibit high levels of SOCS (Fig. 6), which may be attributed to the high soil moisture content in this region. Previous studies have demonstrated a positive correlation between soil moisture content and SOCS (Cheng et al. 2006; Gao et al. 2007; Li et al. 2007), which is consistent with the findings of in the Aba grasslands (Yang et al. 2014). There are several possible reasons for this finding. Firstly, in the surface layer, optimal soil moisture content can influence the uptake and utilization of organic matter and other nutrients by plants (Yu et al. 2019). Secondly, above-ground vegetation and root biomass tend to increase in response to abundant soil moisture (Cong et al. 2016). Lastly, as soil moisture levels rise to optimal thresholds, the decomposition rates of surface litter and shallow-root fine roots accelerate, thereby facilitating the accumulation of SOCS. However, for the subsoil layer, under favorable thermal conditions, soil respiration rates were notably reduced by soil moisture content. Meanwhile, soil respiration decreased under too low or too high soil moisture. In regions typified by high-altitude meadows adjacent to wetlands, riverbanks, and forest boundaries, characterized by sustained high soil moisture levels, soil respiration rates exhibit comparatively slower kinetics compared to lowland grasslands. Consequently, this environment fosters a greater accumulation of deep SOCS (Li et al. 2018; Suh et al. 2009).

A recent study further found that decreased soil water content was the direct reason for SOC degradation due to the decline of carbon input from vegetation (Dong et al. 2021). Understanding the relationship between soil moisture content and SOCS had significant implications for land management and carbon sequestration strategies. Because it was difficult to understand SOCS in a short term, soil water content change (easier to measure than SOCS) would be an important indicator for predicting

SOCS in the Zoige Plateau. Since the 1950s, due to intensive human disturbance (e.g., drainage) and climate change, the Zoige Plateau suffered from a significant loss of wetland (Wu et al. 2011; Xiang et al. 2009), which had significantly increased soil respiration and decreased SOCS (Bai et al. 2013). In terms of global climate change, conservation efforts targeting towards preserving and restoring wetland areas with high soil moisture content are necessary to preserve regional carbon sequestration service and carbon budget. Therefore, the newly developed SOCS product with a spatial resolution of 10 m has important implications for informed land management and ecological restoration.

Conclusion

In this study, we investigated the spatial and vertical distribution characteristics of SOCS in the Zoige Plateau using Sentinel-1 and Sentinel-2 combining field observations. Our results showed that SOC content had a significant vertical distribution and was generally higher than that of other areas due to high altitude, low temperature and soil microbial activities. The XGBoost algorithm integrating Sentinel-1 and Sentinel-2 images provided satisfactory modeling efficiency of 0.59 in SOCS, which was relatively higher compared to several other studies that used only single satellite image, highlighting the importance of model and satellite images in SOCS prediction. The predicted SOCS displayed a remarkable spatial heterogeneity, and newly developed SOCS map with a fine spatial resolution of 10 m would have important applications in land management, ecological restoration and protection in the Zoige Plateau.

Abbreviations

SOC	Soil organic carbon
SOCS	Soil organic carbon stock
S1	Sentinel-1
S2	Sentinel-2
XGBoost	EXtreme Gradient Boosting
RF	Random Forest
BRT	Boosted Regression Tree
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
RVI	Ratio Vegetation Index
DVI	Difference Vegetation Index
NGBDI	Normalized Green–Blue Difference Index
MSR	Modified Red Edge Simple Ratio Index
MSI	Modified Soil-adjusted Vegetation Index
VARI	Visible Atmospherically Resistant Index

Acknowledgements

The study was supported by the Demonstration Project of Carbon Sink Measurement and Value Realization Path after Ecological Restoration of Mountain and Water in Zoige (51000024Y000010980721), the Everest Scientific Research Program of Chengdu University of Technology (80000-2023ZF11410).

Author contributions

Junjie Lei: study design, field work, data entry, data analysis and interpretation, writing of original draft. Changli Zeng, Lv Zhang, Xiaogang Wang and

Chanhua Ma: study design, field work, soil data analysis and interpretation. Tao Zhou, Benjamin Laffitte, Ke Luo and Zhihan Yang: interpretation, review of drafts. Xiaolu Tang: field work, soil data analysis and interpretation, review and editing of drafts, review and editing of subsequent drafts. All authors read and approved the final manuscript.

Availability of data and materials

The data generated and analyzed during the present study are available from the corresponding author.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China. ²Sichuan Academy of Geological Survey, Chengdu 610072, China. ³Sichuan Institute of Nuclear Geological Survey, Chengdu 610061, China. ⁴College of Earth Sciences, Chengdu University of Technology, Chengdu 610095, China. ⁵Tianfu Yongxing Laboratory, Chengdu 610213, China.

Received: 13 January 2024 Accepted: 15 April 2024

Published online: 01 May 2024

References

- Bai J, Lu Q, Zhao Q, Wang J, Ouyang H (2013) Effects of alpine wetland landscapes on regional climate on the Zoige Plateau of China. *Adv Meteorol* 2013:972430. <https://doi.org/10.1155/2013/972430>
- Bentéjac C, Csörgő A, Martínez-Muñoz G (2021) A comparative analysis of gradient boosting algorithms. *Artif Intell Rev* 54:1937–1967
- Cai Q, Guo Z, Hu Q, Wu G (2013) Vertical distribution of soil organic carbon and carbon storage under different hydrologic conditions in Zoigé alpine *Kobresia* meadows wetland. *Scientia Silvae Sinicae* 49:9–16
- Cao B, Ling C (2021) Estimation of aboveground biomass and soil organic carbon density of Zoige Alpine Wetland based on GF-1 remote sensing data. *Remote Sens Technol Appl* 36:229–236
- Castaldi F, Chabrillat S, Don A, van Wesemael B (2019a) Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: an approach to reduce soil moisture and crop residue effects. *Remote Sensing* 11:2121
- Castaldi F, Hueni A, Chabrillat S et al (2019b) Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS J Photogramm Remote Sens* 147:267–282
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In: Chen T (ed) *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. University of Washington, New York, p 785
- Chen H, Yang G, Peng CH et al (2014) The carbon stock of alpine peatlands on the Qinghai-Tibetan Plateau during the Holocene and their future fate. *Quat Sci Rev* 95:151–158
- Chen M, Liu Q, Chen S et al (2019) XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. *IEEE Access* 7:13149–13158
- Cheng J, Wan H, Hu X, Zhao Y (2006) Accumulation and decomposition of litter in the semiarid enclosed grassland. *Acta Ecol Sin* 26:1207–1212
- Cheng J, Cheng J, Yang X, Liu W, Chen F (2012) Spatial distribution of carbon density in grassland vegetation of the Loess Plateau of China. *Acta Ecol Sin* 32:226–237
- Cong J, Wang X, Liu X, Zhang Y (2016) The distribution variation and key influencing factors of soil organic carbon of natural deciduous broadleaf forests along the latitudinal gradient. *Acta Ecol Sin* 36:333–339

- Dash PK, Bhattacharyya P, Roy KS, Neogi S, Nayak AK (2019) Environmental constraints' sensitivity of soil organic carbon decomposition to temperature, management practices and climate change. *Ecol Indic* 107:105644
- Demir S, Şahin EK (2022) Liquefaction prediction with robust machine learning algorithms (SVM, RF, and XGBoost) supported by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing. *Environ Earth Sci* 81:459
- Dong L, Li J, Chen S et al (2021) Changes in soil organic carbon content and their causes during the degradation of alpine meadows in Zoige Wetland. *Chin J Plant Ecol* 45:507–515
- Fan H, Zhao W, Daryanto S et al (2018) Vertical distributions of soil organic carbon and its influencing factors under different land use types in the desert riparian zone of downstream Heihe River basin, China. *J Geophys Res Atmos* 123:7741–7753
- Feida S, Hui L, Ye H (2016) Tge soil organic carbon storage and its spatial characteristics in an Aloine degraded grassland of Zoige, southwest China. *Chin J Grassl* 38:78–84
- Fushiki T (2011) Estimation of prediction error by using K-fold cross-validation. *Stat Comput* 21:137–146
- Gao J, Ou Y, Zhang F, Zhang C (2007) Characteristics of spatial distribution of soil organic carbon in Zoige wetland. *Ecol Environ* 16:1723–1727
- Gebauer A, Gómez VMB, Liess M (2019) Optimisation in machine learning: an application to topsoil organic stocks prediction in a dry forest ecosystem. *Geoderma* 354:113846
- Geng J, Tan Q, Lv J, Fang H (2024) Assessing spatial variations in soil organic carbon and C:N ratio in Northeast China's black soil region: insights from Landsat-9 satellite and crop growth information. *Soil Tillage Res* 235:105897
- Ghatasheh N, Altaharwa I, Aldebei K (2022) Modified genetic algorithm for feature selection and hyper parameter optimization: case of XGBoost in spam prediction. *IEEE Access* 10:84365–84383
- Gholizadeh A, Zizala D, Saberioon M, Boruvka L (2018a) Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens Environ* 218:89–103
- Gholizadeh A, Žizala D, Saberioon M, Borůvka L (2018b) Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens Environ* 218:89–103
- Gomes LC, Faria RM, de Souza E, Veloso GV, Schaefer CE, Fernandes Filho EI (2019) Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma* 340:337–350
- Haralick RM (1979) Statistical and structural approaches to texture. *Proc IEEE* 67:786–804
- Hengl T, de Jesus JM, Heuvelink GBM et al (2017) SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12:e0169748
- Huang T, Gao B, Christie P, Ju X (2013) Net global warming potential and greenhouse gas intensity in a double-cropping cereal rotation as affected by nitrogen and straw management. *Biogeosciences* 10(12):7897–7911
- Jin F, Yang H, Zhao QG (2017) Research progress of soil organic carbon storage and its influencing factors. *Soils* 32:11–17
- Jin X, Qiang H, Zhao L et al (2020) SPEI-based analysis of spatio-temporal variation characteristics for annual and seasonal drought in the Zoige Wetland, Southwest China from 1961 to 2016. *Theoret Appl Climatol* 139:711–725
- Knicker H, Müller P, Hilscher A (2007) How useful is chemical oxidation with dichromate for the determination of "black carbon" in fire-affected soils? *Geoderma* 142:178–196
- Lal R (2004) Soil carbon sequestration impacts on global climate change and food security. *Science* 304:1623–1627
- Lei XD (2019) Applications of machine learning algorithms in forest growth and yield prediction. *J Beijing For Univ* 41:23–36
- Li M, Dong Y, Geng Y, Qi Y (2004) Analyses of the correlation between the fluxes of CO₂ and the distribution of C & N in grassland soils. *Environ Sci* 25:7–11
- Li JH, Hou YL, Zhang SX, Li WJ, Xu DH, Knops JMH, Shi XM (2018) Fertilization with nitrogen and/or phosphorus lowers soil organic carbon sequestration in alpine meadows. *Land Degrad Dev* 29:1634–1641
- Li ZW, Gao P, Hu XY, Yi YJ, Pan BZ, You YC (2020) Coupled impact of decadal precipitation and evapotranspiration on peatland degradation in the Zoige basin, China. *Phys Geogr* 41:145–168
- Li H, Wu Y, Liu S et al (2022) Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Glob Change Biol* 28:3394–3410
- Li L, Xiao HA, Wu JS (2007) Decomposition and transformations of organic substrates in upland and paddy soils in red earth region. *Acta Pedol Sin* 44:669–674
- Ma K, Zhang Y, Tang S, Liu J (2016) Spatial distribution of soil organic carbon in the Zoige alpine wetland, northeastern Qinghai-Tibet plateau. *Catena* 144:102–108
- Ma Q (2013) Ecosystem carbon storage in Zoige alpine marsh of Southwest China. Chinese Academy of Forestry, Beijing
- Mahmoudzadeh H, Matinfar HR, Taghizadeh-Mehrjardi R, Kerry R (2020) Spatial prediction of soil organic carbon using machine learning techniques in western Iran. *Geoderma Reg* 21:e00260
- Marchant B, Villanneau E, Arrouays D, Saby N, Rawlins B (2015) Quantifying and mapping topsoil inorganic carbon concentrations and stocks: approaches tested in France. *Soil Use Manag* 31:29–38
- Mayr A, Binder H, Gefeller O, Schmid M (2014) The evolution of boosting algorithms. *Methods Inf Med* 53:419–427
- Plank S (2014) Rapid damage assessment by means of multi-temporal SAR—a comprehensive review and outlook to Sentinel-1. *Remote Sensing* 6:4870–4906
- Qiu P, Wu N, Luo P, Wang Z, Li M (2009) Analysis of dynamics and driving factors of wetland landscape in Zoige, Eastern Qinghai-Tibetan Plateau. *J Mt Sci* 6:42–55
- Ren Y, Li X, Yang X, Xu H (2021) Development of a dual-attention U-Net model for sea ice and open water classification on SAR images. *IEEE Geosci Remote Sens Lett* 19:1–5
- Rentschler T, Gries P, Behrens T et al (2019) Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi Province, PR China. *PLoS ONE* 14:e0220881
- Rouse Jr JW, Haas RH, Deering DW, Schell JA, Harlan JC (1974) Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. NASA Technical Reports Server
- Shafizadeh-Moghadam H, Minaei F, Talebi-khiyavi H, Xu T, Homaei M (2022) Synergetic use of multi-temporal Sentinel-1, Sentinel-2, NDVI, and topographic factors for estimating soil organic carbon. *Catena* 212:106077
- Singh G, Panda RK (2011) Daily sediment yield modeling with artificial neural network using 10-fold cross validation method: a small agricultural watershed, Kappari, India. *Int J Earth Sci Eng* 4:443–450
- Six J, Paustian K (2014) Aggregate-associated soil organic matter as an ecosystem property and a measurement tool. *Soil Biol Biochem* 68:A4–A9
- Suh S, Lee E, Lee J (2009) Temperature and moisture sensitivities of CO₂ efflux from lowland and alpine meadow soils. *J Plant Ecol* 2:225–231
- Tang X (2013) Estimation of forest aboveground biomass by integrating ICESat/GLAS waveform and TM data. University of Chinese Academy of Sciences Beijing, Beijing
- Tang X, Xia M, Pérez-Cruzado C, Guan F, Fan S (2017) Spatial distribution of soil organic carbon stock in Moso bamboo forests in subtropical China. *Sci Rep* 7:42640
- Vaudour E, Bel L, Gilliot JM et al (2013) Potential of SPOT multispectral satellite images for mapping topsoil organic carbon content over peri-urban croplands. *Soil Sci Soc Am J* 77:2122–2139
- Venter ZS, Barton DN, Chakraborty T, Simensen T, Singh G (2022) Global 10 m land use land cover datasets: a comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sensing* 14:4101
- Wang R, Lu S, Li Q (2019a) Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustain Cities Soc* 49:101623
- Wang X, Li Y, Gong X et al (2019b) Storage, pattern and driving factors of soil organic carbon in an ecologically fragile zone of northern China. *Geoderma* 343:155–165
- Wang Q, Le Noë J, Li QQ et al (2023) Incorporating agricultural practices in digital mapping improves prediction of cropland soil organic carbon content: the case of the Tuojiang River Basin. *J Environ Manage* 330:117203
- Wei L, Tian S, Lu Q et al (2023) Estimating soil organic carbon content of multiple soil horizons in the middle and upper reaches of the Heihe River Basin. *Catena* 234:107574

- Wu X, Guo J, Yang X, Tian X (2011) Soil organic carbon storage and profile inventory in the different vegetation types of Luya Mountain. *Acta Ecol Sin* 31:3009–3019
- Xiang S, Guo R, Wu N, Sun S (2009) Current status and future prospects of Zoige Marsh in Eastern Qinghai-Tibet Plateau. *Ecol Eng* 35:553–562
- Yang R-M, Guo W-W (2019) Modelling of soil organic carbon and bulk density in invaded coastal wetlands using Sentinel-1 imagery. *Int J Appl Earth Obs Geoinf* 82:101906
- Yang S, Li T, Gan Y (2014) Response of soil organic carbon storage in alpine meadow in Aba pastoral areas to different ways of using and degree. *Chin J Grassl* 6:12–17
- Ye CY, Li L, Zhao RY et al (2023) Soil organic carbon and its stability after vegetation restoration in Zoige grassland, eastern Qinghai-Tibet Plateau. *Restor Ecol* 31:e13896
- Yu H, Zha T, Zhang X, Ma L (2019) Vertical distribution and influencing factors of soil organic carbon in the Loess Plateau, China. *Sci Total Environ* 693:133632
- Yu X, Wang YH, Wu LF et al (2020) Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. *J Hydrol* 582:124293
- Zhang B, Zhang Y, Jiang X (2022) Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm. *Sci Rep* 12:9244
- Zheng H, Yuan J, Chen L (2017) Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies* 10:1168
- Zhou T, Geng YJ, Chen J et al (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
- Zhou T, Geng YJ, Ji C et al (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
- Zhou T, Geng YJ, Lv WH et al (2023a) Effects of optical and radar satellite observations within Google Earth Engine on soil organic carbon prediction models in Spain. *J Environ Manage* 338:117810
- Zhou T, Hou YT, Yang ZH et al (2023b) Reducing spatial resolution increased net primary productivity prediction of terrestrial ecosystems: a Random Forest approach. *Sci Total Environ* 897:165134

Publisher's Note

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