Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images

Xianglin He [a], Lin Yang [a,b,*], Anqi Li [a], Lei Zhang [a], Feixue Shen [a], Yanyan Cai [a], Chenhu Zhou [a,b]

[a] *School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China*
[b] *State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, China*

ABSTRACT

It is important to predict the spatial distribution of SOC accurately for migrating carbon emission and sustainable soil management. Environmental variables influence the accuracy of SOC prediction with digital soil mapping (DSM) approaches. In addition to the commonly-used natural predictors, remote sensing variables have been recently used in DSM. However, it is still challenging which variables are effective to predict SOC in farmland. Although phenological parameters have been recently used to indicate human activities that affect SOC in farmland, there are few studies that employ the phenological parameters in SOC prediction. Therefore, this study investigates the feasibility of SOC prediction with the phenological parameters and numerous remote sensing variables extracted from Sentinel-2 at high temporal and spatial resolutions. From 34 Sentinel-2 time series images from 2018 to 2019, 17 phenological parameters were extracted for Xuanzhou, Anhui Province using a dynamic threshold method. Furthermore, fifteen remote sensing predictors comprised of vegetation indices, bright-related indices, and moisture indices were generated from the Sentinel-2 images. The phenological parameters and remote sensing variables were combined with natural variables to predict SOC contents at the surface soil layer using random forest. The results showed that the auxiliary parameters, i.e., the phenological parameters and remote sensing predictors, enhanced the predictability of SOC with an increase in R^2 by 171% and a decrease in RMSE by 14%. This study also identified relatively more important auxiliary parameters for the SOC prediction: the largest data value for the fitted function during the season (a6), rate of increase at the beginning of the season (a8), large seasonal integral (a10), SATVI, and Band8. Therefore, this study verified that the phenological parameters and remote sensing predictors extracted from the Sentinel-2 EVI time series are effective for DSM in farmland.

1. Introduction

As the largest pool in the terrestrial ecosystem, soil carbon pool greatly affects the global carbon cycle (Lal, 2004; Wiesmeier et al., 2019; Hamzehpour et al., 2019). Soil organic carbon (SOC) plays an important role in determining soil quality and crop yield in farmland as it affects the biological, chemical, and physical processes in soils (McBratney et al., 2014; Lamichhane et al., 2019). Since the Industrial Revolution, however, SOC in farmland has significantly depleted by cultivation that induces emission into the atmosphere as an important source of atmospheric CO_2 (Lal, 2004). Therefore, it is crucial to understand the spatial distribution of SOC or carbon stock in farmland due to its critical role in global carbon dynamics. Recently, many studies have paid attention to

the accurate prediction of the spatial distribution of farmland SOC (Craig et al., 2015; Karunaratne et al., 2014; Khaledian and Miller, 2020; Minasny et al., 2013; Köchy et al., 2015).

Digital soil mapping (DSM) has the advantages of high mapping efficiency and low cost, which can make up for the shortcomings of traditional mapping methods,Therefore, DSM is widely used to estimate the spatial distribution of the SOC content (Camera et al., 2016; Jeong et al., 2017; Mcbratney and Santos, 2003; ; . DSM produces SOC maps mainly by the relationships between SOC and its environmental covariates with various approaches, such as geostatistical models (Xiangtian et al., 2020; Yang et al., 2017) or machine learning models (Sushil et al., 2019). The performance of SOC prediction can be improved by not only developing advanced prediction models but also incorporating more

---

* Corresponding author.
  E-mail address: yanglin@nju.edu.cn (L. Yang).

auxiliary environmental variables impacting the SOC spatial variation (Ottoy et al., 2017).

Numerous environmental variables, including variables indicating climate, organisms, relief, parent lithology and land use, have been developed for SOC mapping (Gao et al., 2018; Sushil et al., 2019). As advances in the remote sensing (RS) technology, RS data has been increasingly used for SOC mapping (Vaudour et al., 2013; Mohammadi et al., 2016). Compared to traditional predictors, the available remote sensing bands data with high spatial resolution indirectly reflect the soil information (Vaudour et al., 2013; Forkuor et al., 2017). The commonly-used remote sensing predictors is mainly vegetation index, and few studies use bright-related indices (such as redness index) (Hu et al., 2020; Marques et al., 2020), and moisture indices (such as land surface water index) (Zhou et al. 2016; Semeraro et al. 2019). However, the complex mechanism of the soil-vegetation systems makes it challenging what remote sensing variables are indicative to evaluate SOC in farmland (Fabio et al., 2018; Fathololoumi et al., 2020; Asa et al., 2018). It is necessary to assess the utility of comprehensive remote sensing predictors (i.e., vegetation index, bright-related indices and moisture index) in SOC mapping in farmland(Chen et al., 2019).

In addition to the remote sensing variables, variables indicating human activities have been received attention for SOC mapping (Grunwald et al., 2011; Yang et al., 2020a). Human activities (fertilization, crop rotation, etc.) have a significant impact on SOC. For example, reasonable agricultural management measures may lead to an increase in the SOC content in farmland while decreasing by irrational agricultural management (Chen et al., 2018; Yang et al., 2020a). A few studies also concluded that the impact of these human activities on the SOC content in farmland may exceed those of the normal natural factors (Tong et al., 2014; Li et al., 2017b). However, it is difficult to investigate and quantify these human activities or their impact (Chen et al., 2020). The shortage of environmental variables that can characterize the impact of human activities on SOC limits the SOC mapping accuracy in farming areas.

In recent years, several studies have been conducted on developing predictors to describe the impacts of human activity on SOC (Ratnayake et al., 2016; Nandan et al., 2019; Yang et al., 2020a; Xue et al., 2019). Nandan et al.,(2019) found that returning straw to the field can increase the content of soil organic carbon in farmland and increase soil fertility. Yang et al. (2019) employed Fourier decomposed variables extracted from Normalized Differences Vegetation Index (NDVI) time series data with crop rotation to map SOC in croplands. Yang et al. (2020) also extracted crop phenological parameters from HJ-1 A/B NDVI time series images to assess the impacts of agricultural activities in crop lands. Although the phenological parameters have been successfully incorporated to estimate SOC in farmland, the mapping performance of extracting phenological parameters from other vegetation indices and other remote sensing data sources needs further study. Sentinel-2 has a potential to be an effective data source as it offers exceptional perspectives on land with a combination of wide coverage (swath width of 290 km) and minimum five-day global revisit-time (with twin satellites in orbit) (Van et al., 2014; Grinand et al., 2017). Furthermore, Sentinel-2 provides higher spatial resolution images (20 m) and a smaller system error of atmospheric correction. NDVI is suitable for monitoring the early stage of crop growth, yet enhanced vegetation index (EVI) which has a better sensitivity to high biomass conditions (Vijith and Dodge-Wan, 2020; Wang et al., 2003; Fraga et al., 2014), may have a better ability for monitoring the growth status of crops. Therefore, this study employs the EVI time series from sentinel-2 images to study the effectiveness of phenological parameters in SOC mapping.

Therefore, this study extracted phenological parameters and comprehensive remote sensing predictors from Sentinel-2 images. The main objectives of this study are: 1) to verify the utility of phenological parameters extracted from the Sentinel-2 EVI time series in Predicting SOC in farmland, 2) to identify the effective remote sensing predictors that improve the prediction accuracy of SOC in farmland, and 3) to evaluate the superiority of the SOC prediction driven by the phenological parameters and remote sensing predictors compared with the natural predictors.

2. Materials and methods

2.1. Study area and sampling

Xuancheng City (30°33′~31°19′N, 118°28′~119°4′E) and Langxi County (30°49′~31°18′N, 118°59′~119°23′E) in Anhui Province (Fig. 1). China are located in the transition zone between the southeast hills and the middle and lower reaches of Yangtze River. Cultivated lands in xuancheng and langxi were selected as a study domain that covers an area of 2552 km^2. The southern part of the area is comprised of mainly mountains with an altitude between 500 m and 800 m. The middle part includes mostly hilly areas with an altitude between 50 and 500 m, and the northern part is mostly a plain with an altitude below 50 m. The study area has an annual average precipitation of 1429 mm under a subtropical monsoon climate. The main soil type in the study area is red soil while the main cultivated soil is paddy soil widely distributed in plains and hilly areas, accounting for 17.3% in this area.

From the spring of 2018 to the fall of 2019, a comprehensive survey has been conducted by interviewing farmers and visiting soil fertilizer stations for collecting agricultural management measures such as crop varieties, planting methods, planting time, maturity time, rotation methods, fertilization status, and straw treatment methods. There are mainly two tillage methods in this area: subsoiling and rotary tillage. The depth of subsoiling is till to 25 cm, and the rotary tillage is deep to 15 cm. Crops generally are fertilized twice during a growing season. Before the crops are planted, a basic fertilizer is added (basic fertilizer application: compound fertilizer is applied for about 375 kg/hm^2 while 150 kg/hm^2 for urea fertilizer) (Zhang et al., 2019).

The main rotation types in the study area are rice–wheat, rice-rape, single rice, and little rice-tobacco. The types of rice are determined by maturity time, i.e., early rice (sowing in april and harvesting in June), late rice (sowing in july and harvesting in late October), and middle rice between the early rice and the late rice periods. There are two ways to sow rice: direct seeding and seedling throwing. If the seeding time is sufficient, direct seeding is usually adopted while seedling throwing for other cases. Wheat is sown in November after rice and harvested in early June of the following year. The sowing times of oilseed in oilseed rice-rape rotation and wheat in rice–wheat rotation are slightly different as oilseed rape usually matures a week before wheat. The crop straws in this study area are returning to the field after harvest.

In the fall of 2018, a total of 62 sampling points were collected in the study area through a random sampling strategy. For each location, five soil samples at a depth of 0–20 cm were collected from the four corners and the center in a 10 m × 10 m square area (Zhou et al., 2019).

2.2. Overview of the methodology

The flowchart is shown in Fig. 2. First, we generated the natural predictors, phenological parameters and three types of remote sensing predictors (including vegetation index, bright-related indices and moisture index). We grouped the different categories of predictors into four groups to verify the validity of phenological parameters and remote sensing predictors in SOC mapping. The four groups are, 1) only the normal predictors, 2) normal predictors and the remote sensing predictors, 3) normal predictors and phenological predictors, and 4) normal predictors, phenological predictors, and remote sensing predictors. The importance of environmental variables in each group was then calculated for SOC mapping, and the best combination of predictors in each group was obtained to compare their prediction accuracies.
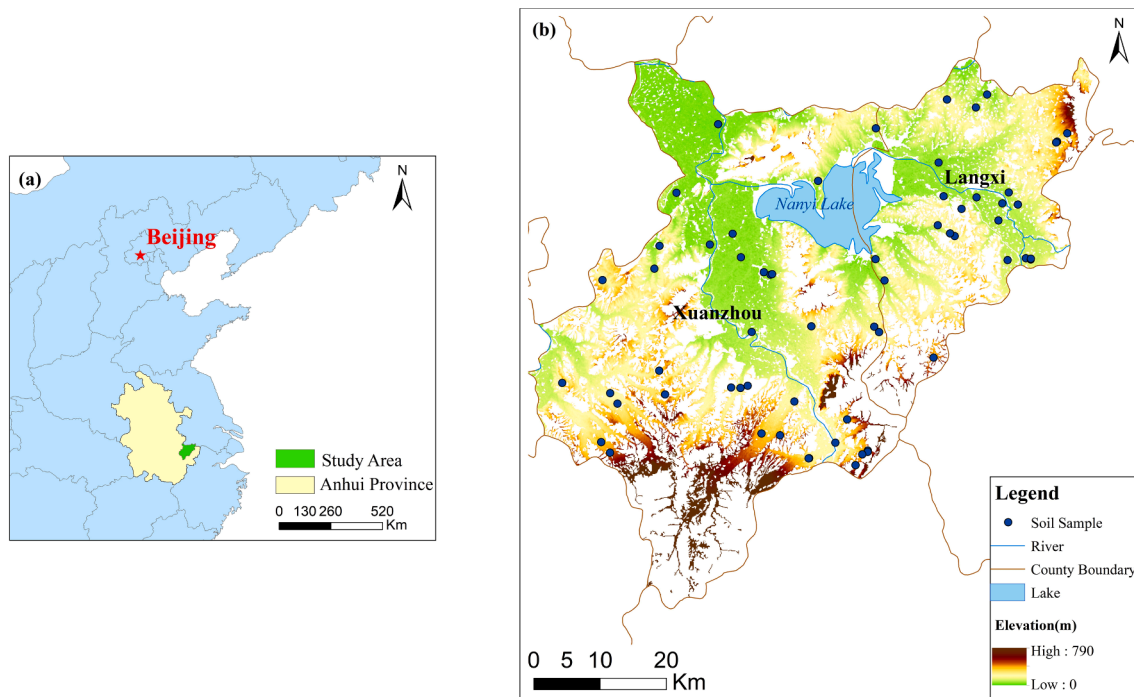
Fig. 1. The study area and soil samples.

2.3. Natural predictors

The main natural environmental factors, including climate, terrain and parent materials, impact the spatial distribution of SOC. This study employed the following natural predictors: annual mean precipitation, annual mean temperature, elevation, slope, planform curvature, profile curvature, topographic wetness index (TWI), and parent lithology. All information of the natural predictors has a raster data format which is projected into the same coordinate reference system as WGS84-UTM. More detailed information is found in the following sections (2.2.1, 2.22, and 2.2.3). As each data have different spatial resolutions, we resampled the raster data into a 90 m resolution using a nearest neighbor resampling method (Seidel et al., 2018) for a consistent data structure.

2.3.1. Climatic variables

The annual mean temperature and annual mean precipitation of the study area in 2015 (http://www.resdc.cn) were employed in this study as climatic variables (Landré et al., 2020). Based on the daily observation data at bout 2,400 meteorological stations across the country, spatial maps (1 km resolution) of the national annual mean temperature and annual mean precipitation were produced by an interpolation technique (ANUSPLIN interpolation) which is a tool for analyzing and interpolating multivariate data using a smooth spline function (Hutchinson, 1998). The units for the precipitation and temperature are 0.1 mm and 0.1 °C, respectively.

2.3.2. Topographic variables

The topographic variables were derived from the digital elevation model (DEM) data generated from the Shuttle Radar Topographic Mission (SRTM) at 90 m resolution (http://srtm.csi.cgiar.org/srtmdata). The topographic variables were elevation, slope, planform curvature, profile curvature, and topographic wetness index (TWI) (Qin et al., 2007). A terrain analysis software, DMapper (https://3d-mapper.com), was adopted.

2.3.3. Parent lithology

The parent lithology maps in this study area were generated from 1: 500,000 Chinese geological maps. There are eight parent lithology types

in this area: pyroclastic rocks, shale, sandstone, conglomerate, granite and granodiorite, limestone, quaternary clay-silt-gravel, quaternary vermicule boulder and grave clay (Grimm et al., 2008; Taghizadeh-Mehrjardi et al., 2014).

2.4. Remote sensing predictors

Remote sensing images mainly provide various soil information, e.g., mineral composition, soil moisture, organic carbon content, and soil texture, corresponding to different bands or combinations of bands. Therefore, multiple remote sensing predictors were generated to reflect the above soil information (Table 1), such as vegetation indices (sensitive to changes in organic matter content), moisture indices (sensitive to soil moisture) and bright-related indices (sensitive to soil texture) (Ben-Dor et al., 2002; Jin et al., 2017; Asa et al., 2018; Wang et al., 2018). With Sentinel-2 images in 2018, 18 remote sensing predictors were generated using the formulas as shown in Table 1 (Jin et al., 2017; Zhang et al., 2019). In order to reduce data redundancy and avoid the image cloud pollution (Ratnayake et al., 2016), crop growth information was extracted from the key time points (February 26, June 6, and November 23) in 2018. All of the remote sensing predictors were generated on the three key time points.

2.5. Sentinel-2 time series data and pre-processing

The Sentinel-2 satellite carries a multispectral sensor (MSI) at an altitude of 786 km, which can cover 13 spectral bands and a width of 290 km. The revisit cycle is 10 days for one satellite and 5 days for the other two satellites. The bands of Sentinel-2 have different spatial resolutions (i.e. 10 m, 20 m and 6 0 m) from visible light and near-infrared to short-wave infrared (Castaldi et al., 2019). As only Sentinel-2 provides optical image data with three bands in the red-edge range, it is very effective for monitoring vegetation health information (Zhou et al., 2020).

During the period from 12 January 2018 to 11 June 2019, thirty-four Sentinel-2 images, Level-1C product (Top of Atmosphere (TOA) reflectance), with less than 10% cloud contamination were selected to cover the two seasons of crop growth in the study area (Table 2), which are
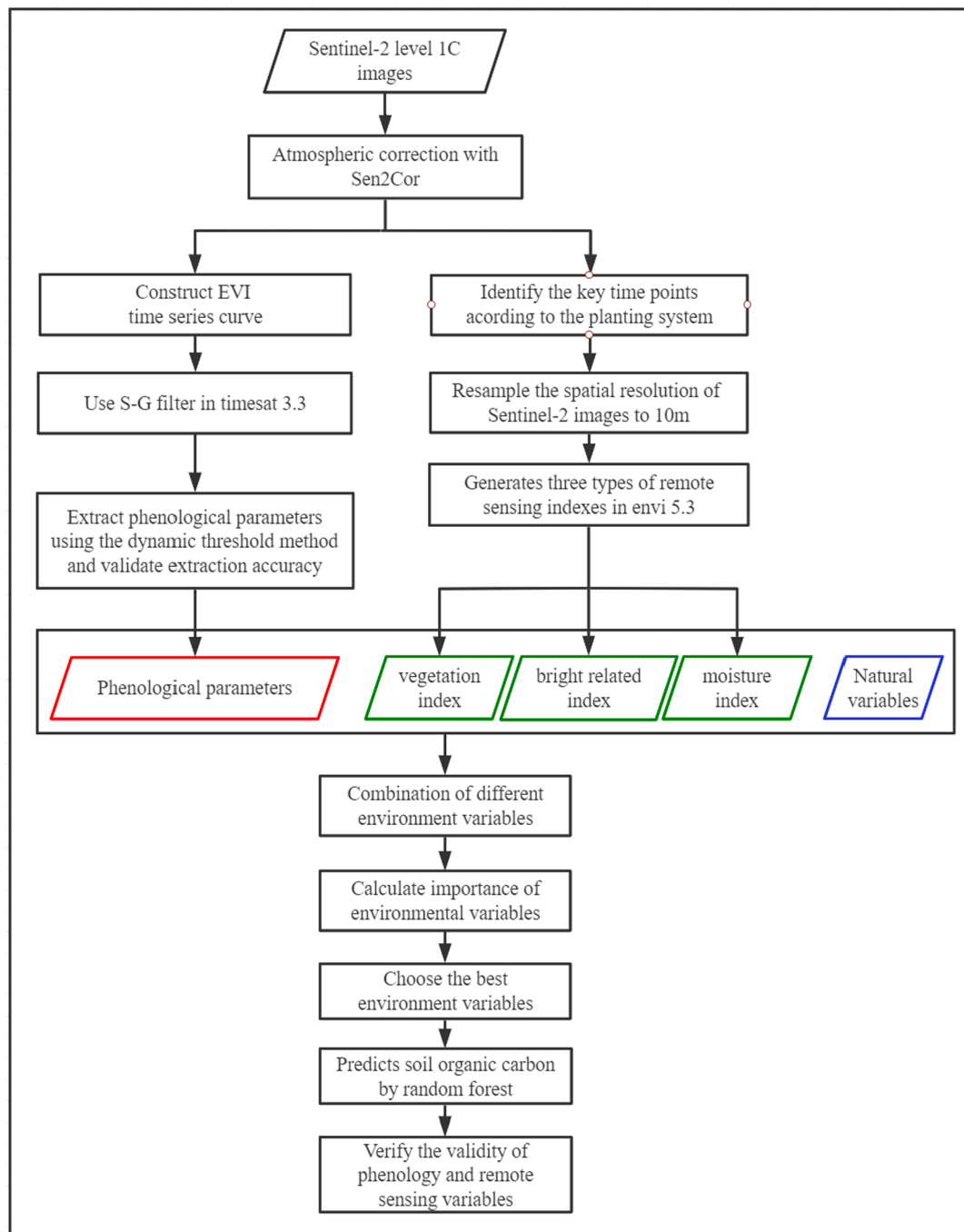
Fig. 2. The flow chart of extracting phenological parameters and remote sensing variables for SOC mapping.

available at the Copernicus open access hub. In addition, the downloaded images were atmospherically corrected to obtain a Level-2A production (Bottom of Atmosphere (BOA) reflectance) using the Sen2-Cor processor (v.2.5.5), a plugin incorporated into the Sentinel Application Platform (SNAP) software. The standard Sen2Cor rural aerosol mode was selected for all images. The atmospherically corrected images were spatially resampled to 10 m to retain the information of Sentinel-2 data to the greatest extent (Castaldi et al., 2019).

Both of NDVI and EVI have been commonly used to generate phenological parameters (Peng et al., 2017; Testa et al., 2018). NDVI is usually saturated in high vegetation coverage areas due to its defect in eliminating atmospheric noise (Kaufman and Tanre, 1992; Li et al., 2010). In addition, NDVI does not consider the impact of background soil noise, which may have an impact on the monitoring quality of crop

growth information. However, EVI usually has a good atmospheric correction effect on the original data, resulting in avoiding the saturation problem of NDVI(Liu et al., 2017). EVI also has an advantage that reduces the effects of aerosol and soil background (Wang et al., 2003). Therefore, this study used the EVI time series constructed by the Sentinel-2 B2, B4, and B8 bands for monitoring the growth of crops in the study area.

The EVI time series curve has usually fluctuated due to the effects of cloud pollution and noise, causing misrepresentations on crop growth information and subsequently affecting the reliability of phenological parameters (Testa et al., 2018). Therefore, this study applied a moving window weighted average algorithm, Savitzky-Golay filter, that smoothes the EVI time series value with a least-squares fitting (Jönsson and Eklundh, 2004; Li et al., 2017a; Mercier et al., 2020). In this study,

Table 1
The detailed descriptions of the remote sensing predictors.

Classes	Abbreviation	Remote sensing Predictors	Formula based on Sentinel-2	Reference
Vegetation index	NDVI	Normalized Differences Vegetation Index	$\dfrac{B8-B4}{B8+B4}$	(Júnior et al. 2019)
	GNDVI	Green Normalized Difference Vegetation Index	$\dfrac{B8-B3}{B8+B3}$	(Zhou et al. 2016)
	TVI	Transformed Vegetation Index	$\left(\dfrac{B8-B4}{B8+B4}+0.5\right)^{1/2}\times100$	(Bagheri, 2020)
	EVI	Enhanced Vegetation Index	$2.5\times\dfrac{B8-B4}{B8+6\times B4-7.52\times B2+1}$	(Rezzouk et al. 2019)
	SATVI	Soil Adjusted Total Vegetation Index	$\dfrac{B11-B4}{B11+B4+1}\times(1+1)-\dfrac{B12}{2}$	(Villarreal et al. 2016)
	SAVI	Soil Adjusted Vegetation Index	$\dfrac{(B8-B4)\times(1+0.5)}{B8-B4+0.5}$	(Venancio et al. 2019)
	GRVI	Green-Red Vegetation Index	$\dfrac{B3-B4}{B3+B4}$	(Tucker, 1979)
	RVI	RatioVegetationIndex	$\dfrac{B8}{B4}$	(Ren and Zhou, 2019)
	MSAVI2	Second Modified Soil Adjusted Vegetation Index	$\dfrac{2\times B8+1-\sqrt{(2\times B8+1)^2-8\times(B8-B4)}}{2}$	(Mahmood et al. 2016)
Bright-related index	BI	Brightness Index	$\dfrac{\sqrt{(B4\times B4)+(B3\times B3)}}{2}$	(Escadafal, 1989)
	BI2	The Second Brightness Index	$\dfrac{\sqrt{(B4\times B4)+(B3\times B3)+(B8\times B8)}}{2}$	(Escadafal, 1989)
	RI	Redness Index	$\dfrac{B4\times B4}{B2\times B3\times B3\times B3}$	(Mathieu et al. 1998)
	CI	Colour Index	$\dfrac{B4-B3}{B4+B3}$	(Pouget et al. 1991)
Moisture index	LSWI	Land Surface Water Index	$\dfrac{B8-B11}{B8+B11}$	(Semeraro et al. 2019)
	MSI	Moisture Stress Index	$\dfrac{B11}{B8}$	(Rock et al. 1985)
	B6	Vegetation red band	Sentinel-2 Band6resolution: 10 m, central wavelength: 740 nm	(Vaudour et al. 2019)
	B8	Near infrared band	Sentinel-2 Band8resolution: 20 m, central wavelength: 842 nm	(Castaldi et al., 2019)
	B11	Short wave infrared	Sentinel-2 Band11resolution: 10 m, central wavelength: 1610 nm	(Castaldi et al. 2016)

Table 2
The date of the Sentinel-2 Time Series data in 2018 and 2019 (images with * were in 2019).

NO.	1	2	3	4	5	6	7	8	9	10	11	12
Date	01–12	02–11	02–26	03–13	03–23	03–28	04–07	04–17	04–27	06–06	06–11	07–26
DOY	12	42	57	72	82	87	97	107	117	157	162	207
NO.	13	14	15	16	17	18	19	20	21	22	23	24
Date	07–31	08–05	08–20	08–30	09–04	10–04	10–19	10–24	10–29	11–23	12–13	12–18
DOY	212	217	232	242	247	277	292	297	302	327	347	352
NO.	25	26*	27*	28*	29*	30*	31*	32*	33*	34*		
Date	12–23	01–17	01–22	01–27	02–06	03–13	04–07	04–17	05–22	06–11		
DOY	357	382	387	392	402	437	462	472	507	527		

we adopted the Savitzky-Golay filter with two parameters, the moving window width = 1 and the order of the polynomial fitting = 2.

2.6. The dynamic threshold method to extract phenological parameters

Adopting a dynamic threshold method , phenological parameters were extracted (Jönsson and Eklundh, 2004; Cao et al., 2018). Compared to a fixed threshold method, the dynamic threshold method not only takes into account the change in a range of the vegetation index time series but also, to a certain extent, eliminates the effects of vegetation types and soil background on phenological parameters (Cao et al., 2018).

Dynamic thresholds are defined by the EVI base level (a4) and the EVI maximum value (a6) within a length of the season (a3). The start of the season (a1) occurs when the left part of the fitted vegetation index curve has reached a specified range (a7) of the amplitude (between a4 to a6), counted from the base level. The end of the season (a2) is defined similarly on the right side of the curve as shown in Fig. 3. The dynamic threshold method was conducted by software timesat3.3 (Jönsson and Eklundh, 2004) to extract 17 phenological parameters during the growing season of crops. More detailed descriptions on the 17 phenological parameters are found in Fig. 3 and Table 3. As combinations of phenological parameters in multiple seasons for one rotation type may
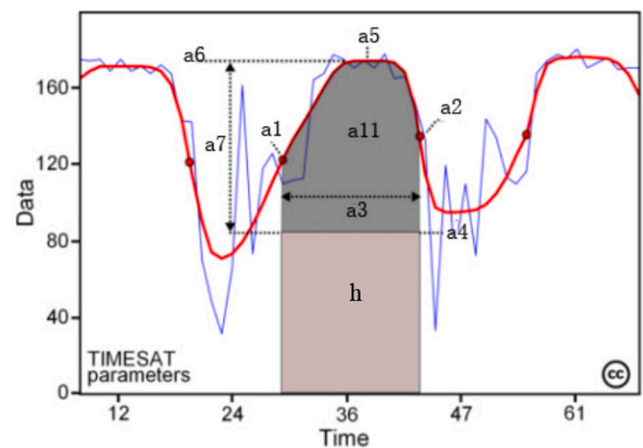

Fig. 3. Illustration of phenological parameters (a1 to a11) , revised according to (Jönsson and Eklundh, 2004).

Table 3
17 phenological parameters extracted from a single growing season of crops using timesat 3.3.

No.	Parameters	Description	Unit
a1	Starting time of a season	The time when data increases to a user-defined value	Day of year
a2	Ending time of a season	The time when data falls to the user-defined value	Day of year
a3	Length of a season	Difference between time for the end of the season and time for the start of the season	Number of days
a4	Base level	The average of the minimum values on the left and right sides of the curve	EVI unit
a5	Mid-time of a season	The average of the two times corresponding to the data at 80% on the left and right sides of the curve	Day of year
a6	Maximum value of a season	Maximum data in a growing season curve	EVI unit
a7	Seasonal amplitude	Difference between the largest data value for the fitted function during the season and the base level	EVI unit
a8	Increase rate at the beginning of a season	The ratio of the data difference of 80% and 20% on the left side of the curve to the corresponding time difference	(EVI unit) /(time unit)
a9	Decrease rate at the end of a season	The ratio of the data difference of 80% and 20% on the right side of the curve to the corresponding time difference	(EVI unit) /(time unit)
a10	Large seasonal integral	The integral of the data of curve function in length of the season (h + a11).	(EVI unit) * (time unit)
a11	Small seasonal integral	The integral of the difference between the curve function data and the base level value in the length of the season	(EVI unit) * (time unit)
a12	Value at the starting time	The data corresponding to the curve function at time for the start of the season	EVI unit
a13	Value at the ending time	The data corresponding to the curve function at time for the end of the season	EVI unit
c14	Base seasonal integral	Difference between a10 and a11 in the same growing season.	(EVI unit) * (time unit)
c15	Complement of small integrals	Subtract a11 from Product of a3 and a7.	(EVI unit) * (time unit)
c16	Triangle season integral	The average of growing time length (a3) and growing season amplitude (a7)	(EVI unit) * (time unit)
c17	Time average of integral	Integral of the function over growth time divided by growth time (a3)	(EVI unit)

also reflect characteristics that represent crop growth, these composite phenological indices can be potentially effective indicators for crop growth and soil. In this study, four composite phenological indices (c14, c15, c16, c17) were generated from the 13 phenological parameters during two seasons. As the single season rice is not unified with other rotation, the first season phenological parameters were mainly used for SOC prediction (Zeng et al., 2019).

2.7. Random forest for SOC prediction

We used a machine learning method, Random Forest (RF), to evaluate the importance of environmental variables and to predict SOC content (Yang et al., 2016; Marcela et al., 2016). RF integrates multiple decision trees that reflect the random selection of samples and features (Breiman et al., 2001). Random forests generally converge to lower generalization errors as the number of decision trees increases. In the RF model, the data are classified into in-bag data and out-of-bag (OOB) data. The "in-bag" samples selected by the bootstrap approach were used for model training (Grimm et al., 2008; Heung et al., 2014) while

the OOB samples were used to estimate general errors for validating a developed model.

RF has been actively applied as an effective model for predicting soil properties or soil types (Grimm et al., 2008; Yang et al., 2016) and for DSM (Wang et al., 2018) as it can evaluate the importance of individual features in classification problems. RF mainly needs two parameters to develop a prediction model: the number of regression trees (ntree) and the number of randomly selected variables at each node (mtry).

In this study, we adopted a recursive feature elimination (RFE) algorithm based on RF to select the most accurate combination of variables for each group of environmental covariates. RFE identifies a subset of environmental covariates that significantly contributes to target variables, avoids the redundancy of the subset of variables, and subsequently produce the best predictive performance. The procedures of using RFE algorithm are as follows:

a. Train a random forest.
b. Calculate the relative importance of candidates among all of the environmental variables.
c. Eliminate the least important variables based on their ranking of importance.
d. Repeat steps 1 to 3 until a subset of the best predictive performance variables is found.

Environmental variable importance is recalculated in each cycle, the ranking of the variables is consistent with their importance in the current forest. After each cycle of eliminate the least important variables, the variable importance and the prediction accuracy of the SOC content will change (Gazzola and Jeong, 2019; Genuer et al., 2010; Shi et al., 2018).

The accuracy using the most accurate combination of variables for each group was compared to validate how effectiveness of adding the pheonological parameters or/and remote sensing variables into the natural variables.

2.8. Assessment of prediction results

A leave-one-out cross-validation was adopted to train the random forest model and validate the prediction results. We used the following measures, the mean absolute error (MAE), root mean squared error (RMSE), coefficient of determination (R^2) and Lin's concordance correlation coefficient (LCCC) (Chen et al., 2019), to evaluate the prediction performance of the random forest models with the four groups of environmental variables. The formulas of the validation indices are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |P_i - O_i| \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2} \tag{5}$$

$$LCCC = \frac{2r\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\bar{P} - \bar{O})^2} \tag{6}$$

where P_i and O_i are the predicted and observed SOC values at the time i, respectively; n is the number of total samples; $\bar{P}$ and $\bar{O}$ are the mean values of the predicted and observed SOC, respectively; r is the Pearson correlation coefficient between the predicted and observed SOC; σ_P and σ_O are the standard deviations of predicted and observed SOC, respectively.

2.9. Uncertainty analysis

For SOC mapping, random forest using the optimal combination of environmental variables was run a hundred times and the average of the 100 times was used as the final prediction. We calculated the standard deviation of each raster pixel based on the 100 SOC maps generated and used the spatial variation of standard deviation to represent the prediction uncertainty (wang et al., 2020; zhou et al., 2021).

3. Results

3.1. Descriptive statistics of the topsoil organic carbon content

The statistics of the organic carbon content in topsoil for the different types of crop rotations are presented in Table 4. The highest SOC content appeared in rice–wheat rotation, followed by Double rice while relatively lower SOC content in rice-rape. Such result indicates that SOC is influenced by fertilization and management in response to crop rotation types. The lowest SOC content occurred in the other crop rotations that include two sampling points with rice-tobacco rotation and one with corn as a crop.

3.2. The extracted phenological parameters

The 17 phenological parameters of the crop in a growing season were extracted for the study area. Table 5 shows the 17 phenological parameters of rice in the rice-rape rotation in 2018, Table 6 is the statistical data of the 17 phenological parameters. The first 13 phenological parameters (a1 to a13) in Table 5 were extracted by the dynamic threshold method while the last four parameters (c14-c17) are the composite phenological indices. For the rice, sowing in early June and harvesting in mid-October, the accuracy of the extracted phenological parameters were validated by the field survey of crop phenology observation data and the absolute error of start and end time of the season was about 5 days. This indicates that the accuracy of the extracted phenological parameters was satisfactory to be used for the SOC prediction and assessment. The standard deviations of a10 and a11, characterizing the growth process of crops, were 35.63% and 47.44%, respectively, indicating a large heterogeneity in rice growth within the study area.

The spatial distribution of three phenological parameters (a6, a8, and a10) is shown in Fig. 4 as examples. The three parameters show different spatial patterns. The spatial distribution map of the EVI maximum value of a season is shown in Fig. 4.a. The upper left corner of the study area has a lower a6 (the maximum EVI value of the first season), which is due to lots of ponds for breeding crabs transformed from farming areas exist in this area. There are some other areas with lower maximum EVI which probably due to the different crop rotation, such as rice-rape, or the abandoned arable land. Fig. 4.b is the increase rate at the beginning of a season from seedling to maturity of crops (a8). The greener the area in the figure, the faster the increase rate. The standard deviation of the increasing rate is 0.08. Fig. 4.c is the integral of EVI of crops in growth time (a10). The standard deviation of a10 is 0.31, which shows different characteristics caused by different crops/crop rotation.

3.3. The relative importance of the environmental variables in each group

The relative importance of environmental variables may vary with groups investigated in this study. The environment variables selected by the RFE approach for each group were shown in Table 7. Fig. 5 shows the ranking of the importance of the variables listed in Table 7. In the first group, the temperature was ranked first, followed by TWI. Among the phenological parameters and remote sensing predictors, near-infrared (Band8), short-wave infrared (Band11), Soil Adjusted Total Vegetation Index (SATVI), the maximum EVI (a6), the increase rate at the beginning of the season (a8), large seasonal integral (a10) were relatively ranked higher, indicating that the remote sensing predictors and phenological parameters selected in this study play an important role in DSM.

3.4. The prediction accuracies of different environment variable groups

The performance measures obtained by the cross-validation for the four groups are presented in Table 8. The predicted and measured values of the SOC in the four environmental variable combinations are shown in Fig. 6. Although the R^2 of all the groups are not high (the points in the scatterplots are not very close to the 1:1 line in Fig. 6.), the R^2 gradually increased with the addition of remote sensing variables and phenological parameters. Considering group 1 as a base-line, the validation results show that the phenological variables and the remote sensing predictors provide added-values in DSM. In particular, the phenological parameters showed a slightly more contribution to improving the prediction accuracy. Moreover, group 4 that combines the phenological variables and the remote sensing predictors performed the best among the four groups. Such results indicate that, along with the natural variables, incorporating auxiliary variables (i.g., remote sensing predictors and the three phenological variables can enhance the performance in the SOC prediction. Compared with the first group (only natural variables), group 4 improved RMSE, MAE, R^2, and LCCC by 13%, 15%, 171%, 52%, respectively. Besides, we noted from Fig. 6 that random forest overestimated the low SOC content and underestimated the high SOC content. This is mainly due to the small number of samples with high and low SOC content.

3.5. The predicted map of SOC and its uncertainty

We generated 100 SOC maps with the optimal environment variables in group 4, and the average of the 100 predictions was used as the final result (Fig. 7.a). The standard deviation of the 100 predicted SOC content at each raster pixel was calculated to express the prediction uncertainty, shown in Fig. 7.b. It shows that the areas with wheat-rice and double rice have higher predicted SOC content while other areas have lower predicted SOC. This shows that the SOC is related with the crop rotations because crop rotations determine the returning straw volume, fertilization quantity, and tillage style(Edmondson et al., 2014; Grimm et al., 2008; Zeng et al., 2019). The average standard deviation value of the farming area is 0.17, showing a good stability of the model. The higher standard deviation is observed in the upper left corner, some areas in the middle plain valley, and areas in narrow valleys in south hills with very few samples. This map could guide the future sampling to reduce uncertainty.

Table 4
Descriptive statistics of SOC content for different crop rotations.

Crop rotations	Sample number	SOC min (g/kg)	SOC max (g/kg)	SOC mean (g/kg)	Standard deviation	Coefficient of variation (%)
Single cropping rice	28	7.30	20.35	13.97	3.29	23.52
Rice -wheat	13	11.05	18.62	14.85	2.48	16.71
Double rice	10	4.59	20.65	14.17	4.49	31.68
Rice-rape	8	9.19	21.92	13.00	3.93	30.20
other	3	8.89	13.31	10.65	2.34	22.00

Table 5
The17 phenological parameters of rice in rice-rape rotation.

NO.	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	c14	c15	c16	c17
1	158.90	276.40	117.50	0.09	219.40	0.76	0.67	0.27	0.20	4.38	3.60	0.14	0.24	2.10	71.02	100.43	1.01
2	159.30	283.75	124.45	0.14	219.85	0.79	0.65	0.26	0.19	5.04	3.59	0.22	0.25	3.62	50.64	82.37	1.05
3	157.90	282.10	124.20	0.04	225.40	0.81	0.77	0.30	0.25	4.87	4.43	0.04	0.23	1.11	78.97	120.50	1.07
4	157.80	284.05	126.25	0.08	224.95	0.72	0.64	0.55	0.28	4.67	3.84	0.06	0.23	2.07	59.76	102.93	1.06
5	157.00	268.60	111.60	0.13	214.50	0.78	0.65	0.37	0.16	4.83	3.49	0.14	0.27	3.34	63.30	95.76	1.15
6	157.20	286.45	129.25	0.12	225.10	0.71	0.59	0.23	0.16	4.72	3.54	0.06	0.29	2.84	66.16	83.38	0.90
7	158.40	277.30	118.90	0.12	219.55	0.65	0.53	0.22	0.16	4.31	3.06	0.14	0.26	3.12	42.47	88.17	1.09
8	157.60	286.30	128.70	0.11	222.85	0.59	0.48	0.16	0.17	3.86	2.71	0.16	0.14	2.86	71.11	79.06	0.79

Table 6
Descriptive analysis of the 17 phenological parameters in Table 5.

	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	c14	c15	c16	c17
Min	61.2	198.0	103.2	0.0	149.4	0.6	0.4	0.1	0.0	1.7	1.0	0.0	0.1	0.5	6.3	25.3	0.4
Max	169.2	319.8	196.8	0.2	244.8	1.0	0.9	0.7	0.7	6.7	5.7	0.3	0.4	4.6	99.4	120.5	1.2
Mean	141.7	271.9	130.2	0.1	209.9	0.7	0.6	0.3	0.2	4.4	3.2	0.1	0.2	2.0	46.7	62.0	0.8
Std	28.1	35.0	16.3	0.0	27.6	0.1	0.1	0.2	0.2	1.2	1.2	0.1	0.1	0.9	24.3	26.6	0.2
Cv (%)	19.9	12.9	12.5	33.0	13.1	12.3	18.9	55.8	75.0	26.3	36.1	50.6	25.3	44.3	52.2	42.9	28.5

Note:min: minimum, max : maximum, std : standard deviation, cv : Coefficient of variation.
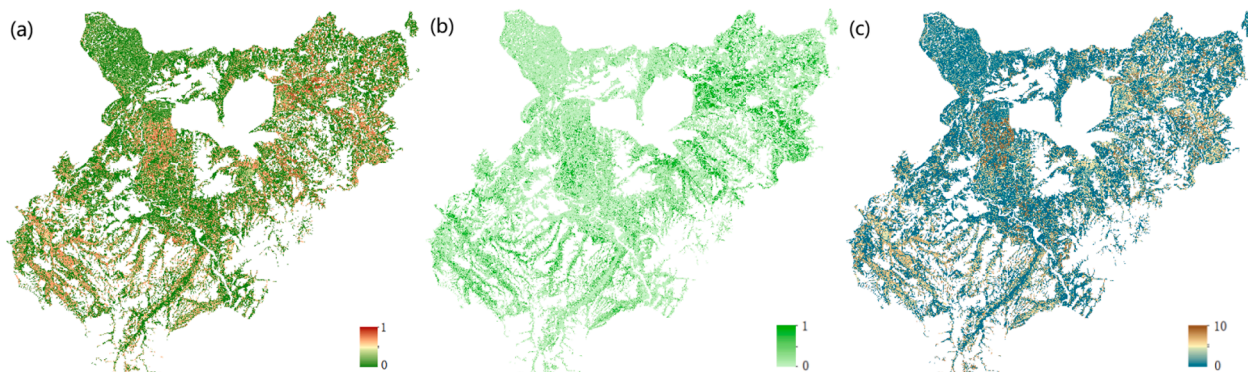

Fig. 4. The spatial distribution map of phenological parameters and field survey pictures (a): a6, EVI maximum value of a season. (b): a8, Increase rate at the beginning of a season (c): a10, Large seasonal integral.

Table 7
Environment variables selected in this study for optimal cartographic accuracy.

Group	Environment variables
1	temperature, twi, elevation, slope, profile curvature, parent lithology
2	temperature, twi, elevation, slope, M6band11, M11band11, M6band8, M11SATVI
3	temperature, elevation, slope, profile curvature, a3, a6, a8, c15
4	temperature, slope, elevation, M6band11, M11band11, M6band8, M11SATVI, a6, a8, a10

Note:The number after 'M' in names of the remote sensing predictors represents the month in 2018.

4. Discussion

4.1. Applicability using remote sensing predictors and phenological predictors in mapping SOC

It was proved that the remote sensing predictors, SATVI, Band8, and Band11, played an important role in improving the SOC prediction. Among them, SATVI with a parameter (L) that normalizes the effect of bare soil is sensitive to both green and senescent vegetations. In particular, SATVI can effectively reflect the change in green coverage of crops (Villarreal et al., 2016; Campos et al., 2018). Sentinel-2 Band8 (near-infrared) and Band11 (short-wave infrared) are sensitive to soil moisture. Near Infrared is used not only to effectively monitor crop growth and soil moisture but also to distinguish between healthy and diseased crops (Van et al., 2016). Short-wave infrared used to distinguish bare soil, water, and different crops has a good ability to penetrate clouds (Da Silveira et al., 2018; Zhou et al., 2020). Near-infrared and short-wave infrared bands have been used to extract vegetation growth, cover, and biomass and consequently have been selected as key variables for predicting SOC in previous studies. For example, Bian et al. (2019) found near-infrared and short-wave infrared bands are effective predictors for mapping SOC in the coastal areas of Northeast China (Van et al., 2016; Asa et al., 2018; Castaldi et al., 2019). This study also demonstrates that SATVI, Band8, Band11, and other remote sensing predictors are effective predictors for the SOC prediction in farmland.

When incorporating phenological parameters into the SOC prediction by the random forest approach, the largest data value for the fitted function during the season (a6), the increase rate of at the beginning of the season (a8) and the large seasonal integral (a10) were selected as the important predictors, especially a8 ranked the first in group 3. Incorporating these auxiliary variables with the commonly-used natural predictors provided considerable added-values in the SOC prediction. Such results may be explained by biological mechanism aspects. There are mainly two stages in the growth process of crops: 1) vegetative growth stage and 2) reproductive growth stage (Akir, 2004). The vegetative growth stage represents the period from seed germination to young spike differentiation when roots, stems, and leaves are growing. During this stage, organic matters are accumulated in crops as a material
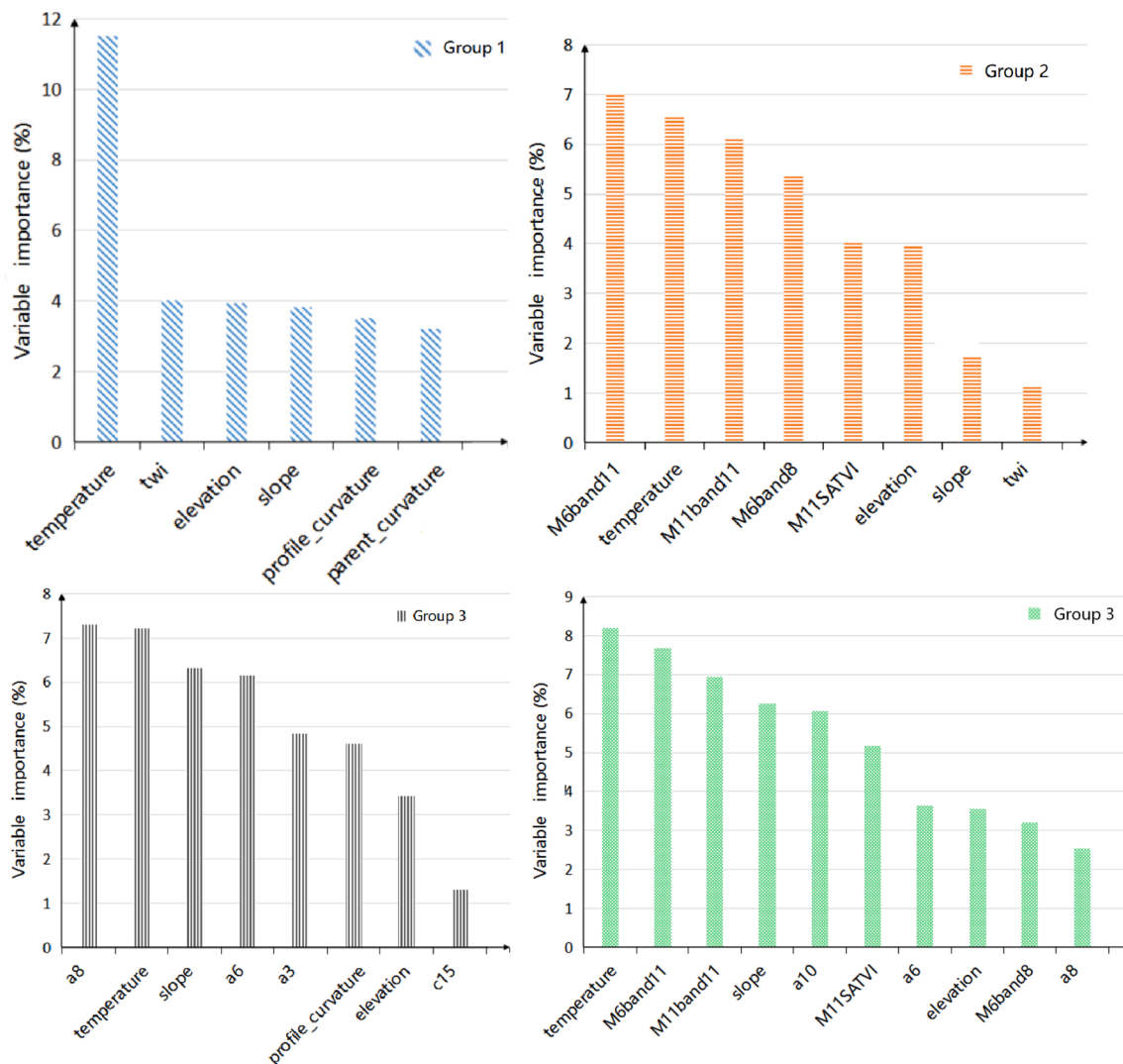
Fig. 5. Variable importance ranking of different environmental variable groups.

Table 8
Performance measures for the four groups investigated in this study.

NO.	Environment Variable Groups	MAE	RMSE	R^2	LCCC
1	Natural predictors	2.68	3.25	0.14	0.29
2	Natural predictors, remote sensing predictors	2.53	3.01	0.27	0.37
3	Natural predictors, phenological predictors	2.39	3.00	0.28	0.36
4	Natural predictors, remote sensing predictors, phenological predictors	2.29	2.83	0.38	0.44

basis for the reproductive growth stage. The young spike differentiation, when the stems and leaves of young spikes grow simultaneously, is a period of vegetative growth and reproductive growth(Xu et al., 2019). After heading, crops will flower, pollinate, fill, and bear fruits, which is the reason that this stage is called reproductive growth(Cheabu et al., 2018). The vegetative and reproductive growth of crops is affected by various factors such as moisture, fertilizer, soil, and variety. The rate of increase at the beginning of the season (a8) is significantly influenced during the vegetative growth period by the growth rate of crops that is related to the SOC content (Rudnick et al., 2017; Cheabu et al., 2018; Si et al., 2020). The more accumulation of SOC, the higher the crop growth rate. The largest data value for the fitted function during the season (a6) usually reflects the highest growth density of crops during the

reproductive growth period. Crop chlorophyll undergoes photosynthesis and absorbs red light to reflect near-infrared(Möller et al., 2017). Therefore, more near-infrared light indicates the crops in farmland absorb more red light, resulting in a higher crop growth rate densely throughout the growth cycle. Consequently, the growth density of crops includes information on the content of SOC. The large seasonal integral (a10) represents the integral of the EVI value over the growing time of the crop (Attila et al., 2018), which describes the life cycle characteristics of the vegetative and reproductive stages of crops(Jiang et al., 2018; Padilla et al., 2017; Saikia et al., 2019) .

It is noteworthy that the phenological parameters used in our study were extracted from EVI time seriers data while there are other ways to produce the phenological parameters such as unmanned aerial vehicle (UAV) (Yang et al., 2020b) images and digital camera monitoring (Alberton et al., 2017).

4.2. Model performance

Farmland is usually plain areas or with gentle terrain condition. Those widely-used predictors such as topographic variables may be too homogenous to map soil variations effectively (Liu et al., 2012; Zeng et al., 2017). Thus, it is often that the SOC mapping accuracy is not high in farmland (Nandan et al., 2019; Wu et al., 2021). In addition, the low sampling density is also a possible reason of the prediction accuracy , i.e.
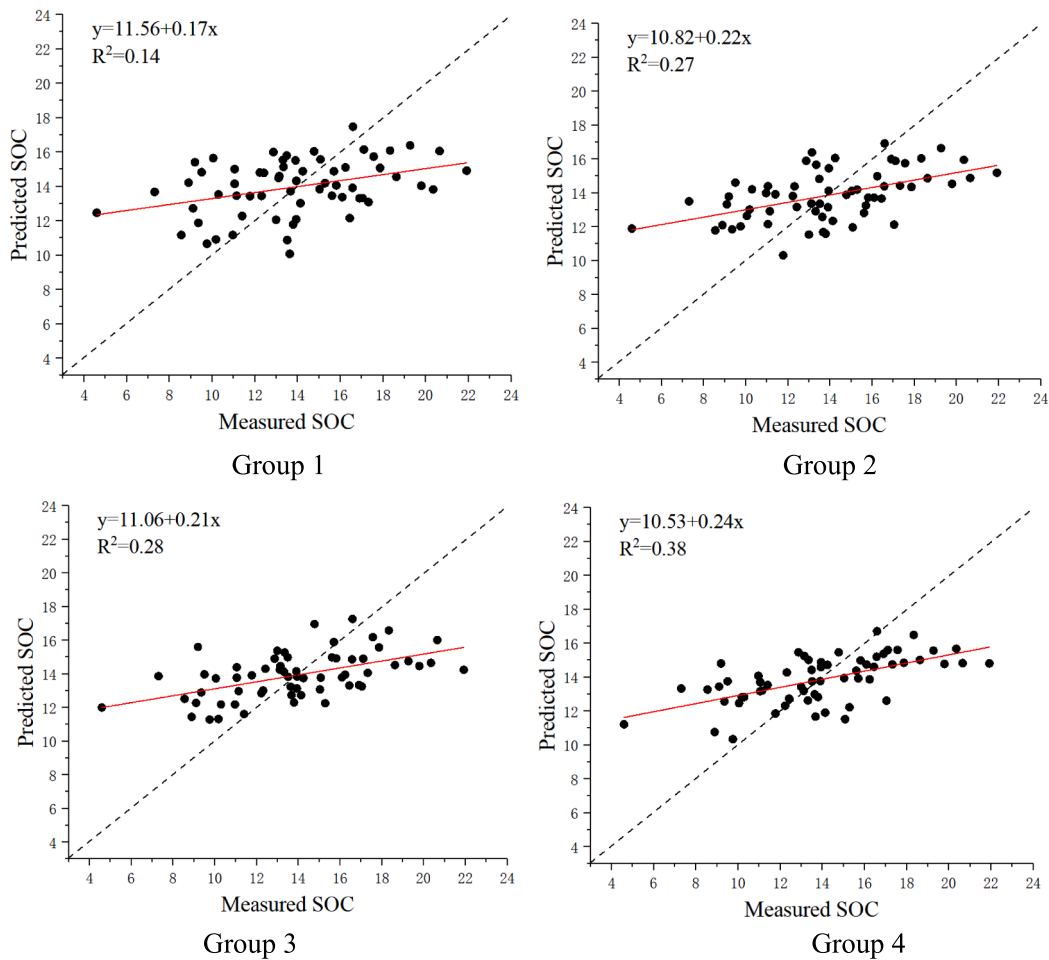
Fig. 6. The scatterplots of the measured against predicted SOC (g/kg) for group 1 (Natural predictors), group 2 (Natural and, remote sensing predictors), group 3 (Natural and phenological predictors), group 4 (all predictors).
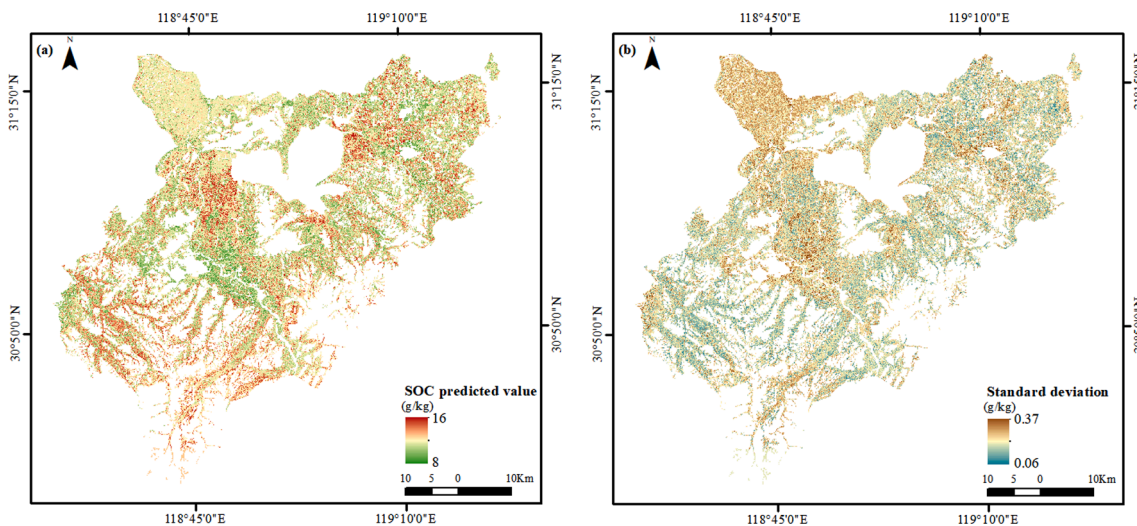

Fig. 7. The predicted SOC and uncertainty maps with the most accurate combination of environmental variables in group 4.

only 62 samples collected at a regional scale (0.024 point/km^2) (Guo et al., 2018). The SOC prediction accuracy with a highest R^2 of 0.38 in our study is similar with the studies of Ahikari et al. (2014), Zhao et al. (2014), and Liang et al. (2019), and higher than other SOC mapping studies in farmland, for example, the R^2 of Wiesmeier et al. (2019) is

0.21, and the R^2 of Funes et al. (2019) is 0.20–0.35. For the SOC prediction in the same area, the maximum R^2 (0.38) of the prediction accuracy in this study is slightly higher than the R^2 (0.37) of Zeng et al. (2016) with a Mixed Geographically Weighted Regression (MGWR) method. Although the prediction accuracy is smaller than Yang et al.,

(2020) in the same study area which used samples collected in 2010 , our study obtained a similar conclusion that adding phenological paraments increased the mapping accuracy. In addition, our study explored which remote sensing predictors work in SOC mapping in farmland.

4.3. Limitations and future perspectives

This study found the two main components that may affect the accuracy of phenological parameters extracted from EVI. Firstly, the time interval of the Sentinel-2 time-series images used in this study was not equidistant, which may cause inconsistency in the extraction accuracy of the phenological parameters (Yang et al., 2020a). The other reason is the effects of cloud pollution in the Sentinel-2 images in May of 2018, which is critical to accurately extract the key growth information about early rice, rape, and wheat. However, there is no alternative to be used during this period. Therefore, it is necessary to integrate multi-sources remote sensing data to resolve these limitations in future works.

In this study, four compound phenological parameters were introduced to reflect the phenological characteristics between two individual seasons. The complement of small integrals (c15) was selected as an important predictor. However, it is necessary to introduce more compound phenological parameters to comprehensively extract the phenological characteristics. For example, the periodic growth law of crops in the time series during a longer time period can be extracted through the model (Zhu et al., 2018; Zhao et al., 2019) to be used as an environmental variable for DSM.

5. Conclusion

This study evaluated the utility of phenological parameters and remote sensing predictors extracted from the Sentinel-2 data for the SOC prediction in farmland. The results showed that the phenological parameters improved the performance in the accuracy of SOC prediction. Among them, this study identified the most important phenological predictors for topsoil SOC content: the largest data value for the fitted function during the season (a6), rate of increase at the beginning of the season (a8) and large seasonal integral (a10). Regarding the remote sensing predictors, near-infrared (Band8), short-wave infrared (Band11), and Soil Adjusted Total Vegetation Index (SATVI) were ranked higher and provide higher added-values in the SOC prediction. This study also intercompared the performance for the four groups according to the combinations of natural, phenological parameters, and remote sensing predictors. Among them, group 4 that incorporates all of the natural, phenological, and remote sensing predictors performed the best, which improved the performance measures, R^2 by 171%, LCCC by 52%, RMSE by 13% and MAE by 15%. Therefore, this study proved that the phenological parameters and remote sensing predictors extracted from the Sentinel-2 images can enhance the ability in SOC mapping in agricultural areas.

Declaration of Competing Interest

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

Acknowledgement

References

Landré, A., Cornu, S., Meunier, J., Guérin, A., Saby, N., 2020. Do climate and land use affect the pool of total silicon concentration? a digital soil mapping approach of french topsoils. Geoderma 364.

Akir, R., 2004. Effect of water stress at different development stages on vegetative and reproductive growth of corn. Field Crops Research 89 (1), 1–16.

Alberton, B., Torres, R., Cancian, L., Borges, B., Morellato, L., 2017. Introducing digital cameras to monitor plant phenology in the tropics. applications for conservation.

Asa, G., Daniel, Z., Mohammadmehdi, S., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sensing of Environment.

Attila, N., János, F., János, T., 2018. Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. Comput. Electron. Agric. 151, 41–49.

Bagheri, N., 2020. Application of aerial remote sensing technology for detection of fire blight infected pear trees. Computers and Electronics in Agriculture 168, 105147.

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using dais-7915 hyperspectral scanner data - a case study over clayey soils in israel. International Journal of Remote Sensing 23 (6), 1043–1062.

Breiman, L., Breiman, L., Cutler, R., 2001. Random forests machine learning. journal of clinical microbiology 2, 199–228.

Camera, C., Zomeni, Z., Noller, J., et al., 2016. A high resolution map of soil types and physical properties for cyprus: a digital soil mapping optimization. Geoderma 285, 35–49.

Campos, V., Gatica, G., Cappa, F., Giannoni, S., Campos, C., 2018. Remote sensing data to assess compositional and structural indicators in dry woodland. Ecological Indicators 88, 63–70.

Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J., Wang, C., Yang, W., 2018. A simple method to improve the quality of ndvi time-series data by integrating spatiotemporal information with the savitzky-golay filter. Remote Sensing of Environment 217, 244–257.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing 147, 267–282.

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote Sensing of Environment 179, 54–65.

Cheabu, S., Moung-Ngam, P., Arikit, S., Vanavichit, A., Malumpong, C., 2018. Effects of Heat Stress at Vegetative and Reproductive Stages on Spikelet Fertility. Rice Science 25, 218–226.

Chen, D., Chang, N., Xiao, J., Zhou, Q., Wu, W., 2019. Mapping dynamics of soil organic matter in croplands with modis data and machine learning algorithms. Science of The Total Environment 669, 844–855.

Chen, Z., Chen, D., Wen, W., Zhuang, Y., Kwan, M., Chen, B., Zhao, B., Yang, L., Gao, B., Li, R., Xu, B., 2019. Evaluating the "2+26" Regional Strategy for Air Quality Improvement During Two Air Pollution Alerts in Beijing: variations of PM2.5 concentrations, source apportionment, and the relative contribution of local emission and regional transport. Atmospheric Chemistry and Physics 19 (10), 6879–6891.

Chen, Z., Chen, D., Zhao, C., Kwan, M., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., Xu, B., 2020. Influence of meteorological conditions on PM2. 5 concentrations across China: A review of methodology and mechanism. Environment International 139, 105558.

Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L., Sanches, I., da Silva, R., Huang, J., Luiz, A., de Oliveira, M., 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. International Journal of Applied Earth Observation and Geoinformation 69, 133–147.

Craig, L., David, M., David, C., et al., 2015. Predictive mapping of soil organic carbon stocks in south australia's agricultural zone. Soil Research 53, 956–973.

Da Silveira, H., Galvão, L., Sanches, I., de Sá, I., Taura, T., 2018. Use of MSI/Sentinel-2 and airborne LiDAR data for mapping vegetation and studying the relationships with soil attributes in the Brazilian semi-arid region. International Journal of Applied Earth Observation and Geoinformation 73, 179–190.

Edmondson, J., Davies, Z., McCormack, S., Gaston, K., Leake, J., 2014. Land-cover effects on soil organic carbon stocks in a European city. Science of The Total Environment 472, 444–453.

Escadafal, R., 1989. Remote sensing of arid soil surface color with landsat thematic mapper. Advances in Space Research 9 (1), 159–163.

Fraga, H., Amraoui, M., Malheiro, A., Moutinho-Pereira, J., Eiras-Dias, J., Silvestre, J., Santos, J., 2014. Examining the relationship between the enhanced vegetation index and grapevine phenology European. J. Rem. Sen. 47 (1).

Fabio, C., Sabine, C., Arwyn, J., Kristin, V., Bart, B., Bas, V., 2018. Soil organic carbon estimation in croplands by hyperspectral remote apex data using the lucas topsoil database. Remote Sensing 10 (2), 153.

Fathololoumi, S., Vaezi, A., Alavipanah, S., Ghorbani, A., Saurette, D., Biswas, A., 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. Science of The Total Environment 721, 137703.

Forkuor, G., Hounkpatin, O., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western burkina faso: a comparison of machine learning and multiple linear regression models. Plos One 12 (1), e0170478.

Gao, X., Li, H., Zhao, X., et al., 2018. Identifying a suitable revegetation technique for soil restoration on water-limited and degraded land: considering both deep soil moisture deficit and soil organic carbon sequestration. Geoderma 319, 61–69.

Gazzola, G., Jeong, M., 2019. Dependence-biased clustering for variable selection with random forests. Pattern Recognition 96, 106980.

Genuer, R., Poggi, J., Tuleau-Malot, M., 2010. Variable selection using random forests. Pattern Recognition Letters 31, 2225–2236.

Grimm, R., Behrens, T., Maerker, 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.

Grinand, C., Maire, G., Vieilledent, G., Razakamanarivo, H., Razafimbelo, T., Bernoux, M., 2017. Estimating temporal changes in soil carbon stocks at ecoregional scale in madagascar using remote-sensing. International Journal of Applied Earth Observation Geoinformation 54, 1–14.

Grunwald, S., Thompson, J., Boettinger, J., 2011. Digital soil mapping and modeling at continental scales: finding solutions for global issues. Soil ence Society of America Journal 75 (4).

Guo, L., Linderman, M., Shi, T., Chen, Y., Duan, L., Zhang, H., 2018. Exploring the Sensitivity of Sampling Density in Digital Mapping of Soil Organic Carbon and Its Application in Soil Sampling. Remote Sensing 10 (6), 888.

Hamzehpour, N., Shafizadeh-Moghadam, H., Valavi, R., 2019. Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture. Catena 182.

Heung, B., Bulmer, C., Schmidt, M., 2014. Predictive soil parent material mapping at a regional-scale: a random forest approach. Geoderma 214–215, 141–154.

Hu, F., Wei, Z., Zhang, W., Dorjee, D., Meng, L., 2020. A spatial downscaling method for smap soil moisture through visible and shortwave-infrared remote sensing data. Journal of Hydrology 125360.

Hutchinson, M., 1998. Interpolation of Rainfall Data with Thin Plate Smoothing Splines: II. Analysis of Topographic Dependence. Journal of Geographic Information and Decision Analysis 2, 168–185.

Jeong, G., Oeverdieck, H., Park, S., Huwe, B., et al., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. Catena 154, 73–84.

Jiang, H., Han, X., Zou, W., Hao, X., Zhang, B., 2018. Seasonal and long-term changes in soil physical properties and organic carbon fractions as affected by manure application rates in the mollisol region of northeast china. Agriculture Ecosystems Environment 268, 133–143.

Jin, X., Song, K., Du, J., Liu, H., Wen, Z., 2017. Comparison of different satellite bands and vegetation indices for estimation of soil organic matter based on simulated spectral configuration. Agricultural and Forest Meteorology 244–245, 57–71.

Jönsson, P., Eklundh, L., 2004. TIMESAT-A program for analyzing time-series of satellite sensor data. Comput Geosci 30, 833–845.

Júnior, R., Siqueira, E., Valera, C., et al., 2019. Diagnosis of degraded pastures using an improved NDVI-based remote sensing approach: An application to the Environmental Protection Area of Uberaba River Basin (Minas Gerais, Brazil). Remote Sensing Applications: Society and Environment 14, 20–33.

Karunaratne, S., Bishop, T., Baldock, J., Odeh, I., 2014. Catchment scale mapping of measureable soil organic carbon fractions. Geoderma 219–220, 14–23.

Kaufman, Y., Tanre, D., 1992. Atmospherically resistant vegetation index (arvi) for eos-modis. IEEE Transactions on Geoscience Remote Sensing 30 (2).

Köchy, M., Don, A., Van, D., Freibauer, A., 2015. Global distribution of soil organic carbon - Part 2: Certainty of changes related to land use and climate. Soil 1, 367–380.

Khaledian, Y., Miller, B., 2020. Selecting appropriate machine learning methods for digital soil mapping. Applied Mathematical Modelling 81, 401–418.

Lal, R., 2004. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. Science 304, 1623–1627.

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.

Li, F., Song, G., Liujun, Z., Yanan, Z., Di, L., 2017a. Urban vegetation phenology analysis using high spatio-temporal ndvi time series. Urban Forestry Urban Greening 25, 43–57.

Li, Z., Li, X., Wei, D., Xu, X., Wang, H., 2010. An assessment of correlation on modis-ndvi and evi with natural vegetation coverage in northern hebei province, china. procedia environmental sciences 2.

Li, Z., Liu, S., Tan, Z., Sohl, T.L., Wu, Y., 2017b. Simulating the effects of management practices on cropland soil organic carbon changes in the temperate prairies ecoregion of the united states from 1980 to 2012. Ecological Modelling 365, 68–79.

Liu, F., Geng, X., Zhu, A., Fraser, W., Waddell, A., 2012. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from modis. Geoderma 171, 44–52.

Liu, S., Liu, X., Liu, M., Wu, L., Ding, C., Huang, Z., 2017. Extraction of rice phenological differences under heavy metal stress using evi time-series from hj-1a/b data. Sensors 17 (6), 1243-.

Mahmood, K., Batool, S., Chaudhry, M., 2016. Studying bio-thermal effects at and around msw dumps using satellite remote sensing and gis. Waste Management 55, 118–128.

Marcela, R., Antonio, M., Cortizas, L., 2016. Mapping soil organic carbon content using spectroscopic and environmental data: A case study in acidic soils from NW Spain. Science of The Total Environment 539, 26–35.

Marques, M., Alvarez, A., Carral, P., Sastre, B., Bienes, R., 2020. The use of remote sensing to detect the consequences of erosion in gypsiferous soils. International Soil and Water Conservation Research 8 (4).

Mathieu, R., Pouget, M., Cervelle, B., Escadafal, R., 1998. Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. Remote Sensing of Environment 66 (1), 17–28.

Mcbratney, A., Field, D., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213.

Mcbratney, A., Santos, M., 2003. On digital soil mapping. Geoderma 117, 3–52.

Mercier, A., Betbeder, J., Baudry, J., Le Roux, V., Spicher, F., Lacoux, J., Roger, D., Hubert-Moy, L., 2020. Evaluation of Sentinel-1 2 time series for predicting wheat and rapeseed phenological stages. ISPRS Journal of Photogrammetry and Remote Sensing 163, 231–256.

Minasny, B., Mcbratney, A., Malone, B., Wheeler, I., 2013. Digital mapping of soil carbon. Advances in Agronomy 118 (118), 1–47.

Mohammadi, J., Mirzaee, S., Asadi, M., 2016. Spatial variability of soil organic matter using remote sensing data. Catena.

Möller, M., Gerstmann, H., Gao, F., Dahms, T., Förster, M., 2017. Coupling of phenological information and simulated vegetation index time series: Limitations and potentials for the assessment and monitoring of soil erosion risk. Catena 150, 192–205.

Nandan, R., Singh, V., Singh, S., Kumar, V., Hazra, K., Nath, C., et al., 2019. Impact of conservation tillage in rice–rapeseed cropping systems on soil aggregation, carbon pools and nutrients. Elsevier Sponsored. Documents 340.

Ottoy, S., De Vos, B., Sindayihebura, A., Hermy, M., Van Orshoven, J., 2017. Assessing soil organic carbon stocks under current and potential forest cover using digital soil mapping and spatial generalisation. Ecological indicators 77, 139–150.

Padilla, F., Peña-Fleitas, M., Fernández, M., Del Moral, F., Thompson, R., Gallardo, M., 2017. Responses of soil properties, crop yield and root growth to improved irrigation and n fertilization, soil tillage and compost addition in a pepper crop. Scientia Horticulturae 225, 422–430.

Peng, D., Wu, C., Li, C., Zhang, X., Liu, Z., Ye, H., 2017. Spring green-up phenology products derived from modis ndvi and evi: intercomparison, interpretation and validation using national phenology network and ameriflux observations. Ecological indicators 77, 323–336.

Pouget, M., Madeira, N., José, L., Kamal, S., 1991. Caractéristiques spectrales des surfaces sableuses de la région côtière nord-ouest de l'egypte : application aux données satellitaires spot. j am med assoc 164 (10), 1087–1088.

Qin, C., Zhu, A., Pei, T., Li, B., Zhou, C., Yang, L., 2007. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. International Journal of Geographical Information Science 21 (4), 443–458.

Ratnayake, R., Karunaratne, S., Lessels, J., Yogenthiran, N., Rajapaksha, P., Gnanavelrajah, N., 2016. Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka. Geoderma 7.

Ren, H., Zhou, G., 2019. Estimating green biomass ratio with remote sensing in arid grasslands. Ecological Indicators 98, 568–574.

Rezzouk, F., Gracia-Romero, A., Kefauver, S., Gutiérrez, N., Aranjuelo, I., Serret, M., Araus, J., 2019. Remote sensing techniques and stable isotopes as phenotyping tools to assess wheat yield performance: Effects of growing temperature and vernalization. Plant Sci. 10.

Rock, B., Vogelmann, J., Williams, D., 1985. Field and airborne spectral characterization of suspected damage in red spruce (picea rubens) from Vermont. International Symposium - Machine Processing of Remotely Sensed Data. NTRS.

Rudnick, D., Irmak, S., Djaman, K., Sharma, V., 2017. Impact of irrigation and nitrogen fertilizer rate on soil water trends and maize evapotranspiration during the vegetative and reproductive periods. Agricultural Water Management 191, 77–84.

Saikia, R., Sharma, S., Thind, H., Sidhu, H., 2019. Temporal changes in biochemical indicators of soil quality in response to tillage, crop residue and green manure management in a rice-wheat system. Ecological Indicators 103, 383–394.

Seidel, F., Natasha, S., Cable, M., Robert, G., Anthony, F., 2018. Imaging spectrometer emulates landsat: a case study with airborne visible infrared imaging spectrometer (aviris) and operational land imager (oli) data. Remote Sensing of Environment 215, 157–169.

Semeraro, T., Mastroleo, G., Pomes, A., Luvisi, A., Gissi, E., Aretano, R., 2019. Modelling fuzzy combination of remote sensing vegetation index for durum wheat crop analysis. Computers Electronics in Agriculture 156, 684–692.

Shi, J., Yang, L., Zhu, A., Qin, C., Liu, p., Zeng, C., 2018. Machine-Learning Variables at Different Scales vs. Knowledge-based Variables for Mapping Multiple Soil Properties. Soil Science Society of America Journal 82 (3), 645–653.

Si, Z., Zain, M., Mehmood, F., Wang, G., Gao, Y., Duan, A., 2020. Effects of nitrogen application rate and irrigation regime on growth, yield, and water-nitrogen use efficiency of drip-irrigated winter wheat in the North China Plain. Agric, Water Manage, p. 231.

Sushil, L., Lalit, K., Brian, W., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma 352.

Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone, B., 2014. Digital mapping of soil salinity in ardakan region, central iran. Geoderma 213, 15–28.

Testa, S., Soudani, K., Boschetti, L., et al., 2018. MODIS-derived EVI, NDVI and WDRVI time series in French deciduous forests. International Journal of Applied Earth Observation and Geoinformation 64, 132–144.

Tong, X., Xu, M., Wang, X., Bhattacharyya, R., Zhang, W., Cong, R., 2014. Long-term fertilization effects on organic carbon fractions in a red soil of China. Catena 113, 251–259.

Tucker, C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing and Environment 8 (2), 127–150.

Van, F., Ruitenbeek, H., Van, D., 2014. Potential of esa's Sentinel-2 for geological applications. Remote Sensing of Environment.

Van, W., Bas, V., Oost, K., 2016. Uas-based soil carbon mapping using vis-nir (480–1000 nm) multi-spectral imaging: potential and limitations. Geoderma An International. Journal of Soil Science.

Vaudour, E., Bel, L., Gilliot, J., Coquet, Y., Hadjar, D., Cambier, P., 2013. Potential of spot multispectral satellite images for mapping topsoil organic carbon content over peri-urban croplands. Soil Science Society of America Journal 77 (6), 2122.

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and mediterranean agroecosystems. Remote Sensing of Environment 223, 21–33.

Venancio, L., Mantovani, E., Amaral, C., Neale, C., Campos, I., 2019. Forecasting corn yield at the farm level in brazil based on the fao-66 approach and soil-adjusted vegetation index (savi). Agricultural Water Management 225, 105779.

Villarreal, M., Norman, L., Buckley, S., 2016. Multi-index time series monitoring of drought and fire effects on desert grasslands. Remote Sensing of Environment 183, 186–197.

Wang, B., Waters, C., Orgill, S., 2018. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern australia. Science of the Total Environment.

Vijith, H., Dodge-Wan, D., 2020. Applicability of MODIS land cover and Enhanced Vegetation Index (EVI) for the assessment of spatial and temporal changes in strength of vegetation in tropical rainforest region of Borneo. Remote Sensing Applications: Society and Environment 18, 100311.

Wang, Z., Liu, C., Huete, A., 2003. From avhrr-ndvi to modis-evi: advances in vegetation index research. Acta Ecologica Sinica 23 (5), 979–987.

Wu, Z.H., Liu, Y.L., Han, YR., Zhou, J.N., Liu, J.M., 2021. Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data. Science of The Total Environment 754, 142120.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., Von, L., Marin-Spiotta, E., Van, W., Rabot, E., Liess, M., Garcia-Franco, N., Wollschlaeger, U., Vogel, H., Koegel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils-A review of drivers and indicators at various scales. Geoderma 333.

Xiangtian, M., Yilin, B., Jiangui, L., Huanjun, L., Xinle, Z., Yu, Z., Peng, W., Haitao, T., 2020. Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data. International Journal of Applied Earth Observations and Geoinformation 89.

Xu, J., Han, H., Ning, T., Li, Z., Lal, R., 2019. Long-term effects of tillage and straw management on soil organic carbon, crop yield, and yield stability in a wheat-maize system. Field Crops Research 233, 33–40.

Xue, B., Huang, L., Huang, Y., Kubar, K., Lu, J., 2019. Straw management influences the stabilization of organic carbon by fe (oxyhydr)oxides in soil aggregates. Geoderma 358.

Yang, L., He, X.L., Shen, F.X., Zhou, C.H., Zhu, A.X., Gao, B., Chen, Z.Y., Li, M., 2020a. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. Soil and Tillage Research 196.

Yang, Q., Shi, L.S., Han, J.Y., Huang, K., 2020b. A near real-time deep learning approach for detecting rice phenology based on UVA images. Agricultural and Forest Meteorology 287, 15.

Yang, R., Zhang, G., Liu, F., Lu, Y., Yang, F., Yang, F., Yang, M., Zhao, Y., Li, D., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecol. Indicators 60, 870–878.

Yang, L., Zhu, A., Zhao, Y.., Li, D., Zhang, G., Zhang, S., Band, L., 2017. Regional soil mapping using multi-grade representative sampling and a fuzzy membership-based mapping approach. Pedosphere 027, 344–357.

Zeng, L., Wardlow, B., Xiang, D., Hu, S., Li, D., 2019. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sensing of Environment.

Zeng, C., Zhu, A., Liu, F., Yang, L., Rossiter, D., Liu, J., et al., 2017. The impact of rainfall magnitude on the performance of digital soil mapping over low-relief areas using a land surface dynamic feedback method. Ecological Indicators 72, 297–309.

Zhang, Y., Guo, L., Chen, Y., Shi, T., Wang, S., 2019. Prediction of soil organic carbon based on landsat 8 monthly ndvi data for the jianghan plain in hubei province, china. Remote Sensing 11 (14), 1683.

Zhang, J., Zhuang, M., Shan, N., Zhao, Q., et al., 2019. Substituting organic manure for compound fertilizer increases yield and decreases NH_3 and N_2O emissions in an intensive vegetable production systems. Science of The Total Environment 670, 1184–1189.

Zhao, K., Wulder, M., Hu, T., Bright, R., Brown, M., 2019. Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: a bayesian ensemble algorithm. Remote Sensing of Environment 232, 111181.

Zhou, J., Khot, L., Bahlol, H., Boydston, R., Miklas, P., 2016. Evaluation of ground, proximal and aerial remote sensing technologies for crop stress monitoring. Ifac Papersonline 49 (16), 22–26.

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A., 2020. 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. Science of The Total Environment 755.

Zhou, Y., Hartemink, A., Shi, Z., Liang, Z., Lu, Y., 2019. Land use and climate change effects on soil organic carbon in north and northeast china. Science of the Total Environment 647, 1230–1238.

Zhu, A.X, Lu, G.N., Qin, C.Z., Zhou, C.H., 2018. Spatial prediction based on Third Law of Geography. Annals of GIS 24, 225–240.