

TOPICAL REVIEW • OPEN ACCESS

A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect

To cite this article: Haili Huang *et al* 2022 *Environ. Res. Lett.* **17** 123004

View the [article online](#) for updates and enhancements.

You may also like

- [Modelling the impacts of intensifying forest management on carbon budget across a long latitudinal gradient in Europe](#)
Anu Akujärvi, Anatoly Shvidenko and Stephan A Prietsch
- [Increased forest coverage will induce more carbon fixation in vegetation than in soil during 2015–2060 in China based on CMIP6](#)
Mengyu Zhang, Honglin He, Li Zhang et al.
- [Terrestrial carbon sequestration under future climate, nutrient and land use change and management scenarios: a national-scale UK case study](#)
Dmitry Yumashev, Victoria Janes-Bassett, John W Redhead et al.

ENVIRONMENTAL RESEARCH
LETTERS

TOPICAL REVIEW

OPEN ACCESS

RECEIVED
8 May 2022REVISED
23 September 2022ACCEPTED FOR PUBLICATION
18 November 2022PUBLISHED
5 December 2022

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



A review on digital mapping of soil carbon in cropland: progress, challenge, and prospect

Haili Huang¹, Lin Yang^{1,*} , Lei Zhang¹ , Yue Pu¹, Chenconghai Yang¹, Qi Wu¹, Yanyan Cai¹, Feixue Shen¹ and Chenghu Zhou^{1,2}¹ School of Geography and Ocean Science, Nanjing University, Nanjing 210023, People's Republic of China² State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, CAS, Beijing 100101, People's Republic of China

* Author to whom any correspondence should be addressed.

E-mail: yanglin@nju.edu.cn**Keywords:** soil carbon, cropland, digital soil mapping, soil information, environmental covariates, predictive models**Abstract**

Cropland soil carbon not only serves food security but also contributes to the stability of the terrestrial ecosystem carbon pool due to the strong interconnection with atmospheric carbon dioxide. Therefore, the better monitoring of soil carbon in cropland is helpful for carbon sequestration and sustainable soil management. However, severe anthropogenic disturbance in cropland mainly in gentle terrain creates uncertainty in obtaining accurate soil information with limited sample data. Within the past 20 years, digital soil mapping has been recognized as a promising technology in mapping soil carbon. Herein, to advance existing knowledge and highlight new directions, the article reviews the research on mapping soil carbon in cropland from 2005 to 2021. There is a significant shift from linear statistical models to machine learning models because nonlinear models may be more efficient in explaining the complex soil-environment relationship. Climate covariates and parent material play an important role in soil carbon on the regional scale, while on a local scale, the variability of soil carbon often depends on topography, agricultural management, and soil properties. Recently, several kinds of agricultural covariates have been explored in mapping soil carbon based on survey or remote sensing technique, while, obtaining agricultural covariates with high resolution remains a challenge. Based on the review, we concluded several challenges in three categories: sampling, agricultural covariates, and representation of soil processes in models. We thus propose a conceptual framework with four future strategies: representative sampling strategies, establishing standardized monitoring and sharing system to acquire more efficient crop management information, exploring time-series sensing data, as well as integrating pedological knowledge into predictive models. It is intended that this review will support prospective researchers by providing knowledge clusters and gaps concerning the digital mapping of soil carbon in cropland.

1. Introduction

Soil constitutes the largest carbon pool in terrestrial ecosystems (Paustian *et al* 2016, Sun *et al* 2020). Soil carbon is the critical component of soil fertility to sustain plants, animals, and humans (Rumpel *et al* 2018). The spatial distribution of soil carbon not only helps in figuring out soil organic carbon (SOC) content/stock over regions but also contributes to supporting Earth system modeling and sustainable land management (Dignac *et al* 2017, Mandal *et al* 2020). However, precise quantification of global soil carbon

stock remains challenging mainly because of sparse soil observations over the world, especially at deep depths, also leading to the great uncertainty in terrestrial carbon stock estimation (Piao *et al* 2009). Therefore, it is important to develop effective methods to measure and monitor the spatially accurate information of soil carbon, especially based on limited sample data (Tautges *et al* 2019, Varney *et al* 2020).

Cropland covers 12% of the Earth's ice-free land (Foley *et al* 2011). Arable soils under considerable threat in many regions are facing the issue of soil

quality degradation (Haddaway *et al* 2016, Clark and Tilman 2017). Evidence points to the long decline in the soil carbon stock contributing 116 Pg of carbon (as CO₂) to the atmosphere and less organic carbon to hydrographic environmental sediments with the first furrow of human cultivation (Schlesinger and Amundson 2019). Concerns about increased greenhouse gas emissions from degraded cropland have drawn international attention. The international initiative '4 per 1000' attaches great importance to the managed agricultural fields for the potential of increasing SOC stock (Minasny *et al* 2017). Other global initiatives included RECSOIL, launched by FAO, recalled for sequestering more carbon in potential soils (cropland and degraded soils) (FAO 2019). Cropland has great potential in carbon sequestration which can be achieved through effective management practices (Nikolaidis 2011, Gattinger *et al* 2012, Paustian *et al* 2016). To understand the current status and changing of cropland carbon content and stock can shed light on maintaining soil fertility, decision-making for land management and enabling a realistic assessment of carbon sequestration capacity (Lacoste *et al* 2014, Lamichhane *et al* 2019).

Conventional soil carbon investigation methods based on field experiments are time- and labor-consuming (Guo *et al* 2020). In recent decades, digital soil mapping (DSM) has become a vital way to obtain spatial information on soil carbon (McBratney *et al* 2003, Sanchez *et al* 2009, Minasny and McBratney 2016). Following the soil formation theory (Jenny 1941), McBratney *et al* (2003) developed a conceptual DSM paradigm describing the empirical for empirical quantitative relationships between soil and environmental covariates, called 'scorpan' model:

$$S_C = f(s, c, o, r, p, a, n).$$

The environmental covariates, comprised of soil (*s*), climate (*c*), vegetation (*o*), topography (*r*), parent materials (*p*), time (*a*), and space (*n*) are represent the soil-forming environment (Jenny 1941, McBratney *et al* 2003). With DSM techniques, a quantitative relationship (*f*) between observed soil properties and environmental covariates is firstly constructed, and then used to predict soil properties of un-sampled locations based on the relationship (figure 1). With advances in DSM, the estimates of soil carbon have improved over the past decade. Plot-scale, local, regional, and global applications of mapping soil carbon have been conducted (Guevara *et al* 2018, Poggio *et al* 2021).

Despite great advances in DSM in recent years, acquiring satisfying accuracy for mapping cropland's carbon is not a simple exercise because of uncertainties surrounding carbon predictions in plains or small-scale cropland (Guo *et al* 2020, Wadoux *et al* 2021). Most cropland worldwide is

distributed in the area with flat terrain, thus the observed heterogeneities for commonly involved natural factors (such as topography) are limited (Stevens *et al* 2014). Moreover, human activities such as irrigation, crop rotation, and fertilization have a greater impact on soil qualities (Mandal *et al* 2020), while obtaining quantitative agricultural management information is difficult. Using weakly related environmental covariates to represent the characteristics of the spatial distribution of soil carbon in agricultural areas paints only a partial picture (Hamzehpour *et al* 2019). Different strategies could be obtain scientific insights into the soil processes, pedological knowledge is required when selecting environmental covariates (Wadoux *et al* 2021). Research into and exploring new covariates in cropland should be improved to provide better maps and tools for the accurate assessment of soil carbon to promote action and multi-stakeholder participation in crop production.

Recent advances in both sampling, environmental covariates availability, and spatial modeling improved our ability to measure cropland soil carbon and its dynamics (Lacoste *et al* 2014, Lamichhane *et al* 2019). For example, World Soil Information Service provides quality-assured and standardized soil profile data over 5.8 million records to assist large-scale DSM and environmental applications (Batjes *et al* 2020). Additionally, numerous potential covariates are provided by remote sensing techniques such as the Landsat satellite (Taghizadeh-Mehrjardi *et al* 2016), the moderate resolution imaging spectroradiometer (MODIS)-based satellite (Chen *et al* 2019), and Sentinel-2 hyperspectral sensor (Zepp *et al* 2021). Progress in proximal sensing and the development of spectral libraries showed great promise in monitoring soil carbon (Arrouays *et al* 2020). To handle the booming geographic covariates, diverse advanced computational techniques have been explored for their applicability in optimizing the prediction of soil carbon (Wadoux *et al* 2020).

This paper reviews the digital mapping of soil carbon in cropland, including predictive methods and environmental covariates, and highlights some challenges and opportunities to accurately map and monitor soil carbon in cropland for future work. First, this article introduces an overview of general findings in soil carbon mapping within cropland, in terms of the temporal and spatial characteristics of the relevant studies. Second, the article identifies the most regularly used predictive models and environmental covariates in recent years in the context of the particular conditions of cropland. Particularly, we put more focus on agricultural information and discuss the section in two parts, crop information and agricultural management. Then, the review also assesses studies using soil information based on proximal and remote soil sensing. Based on the above summary, we

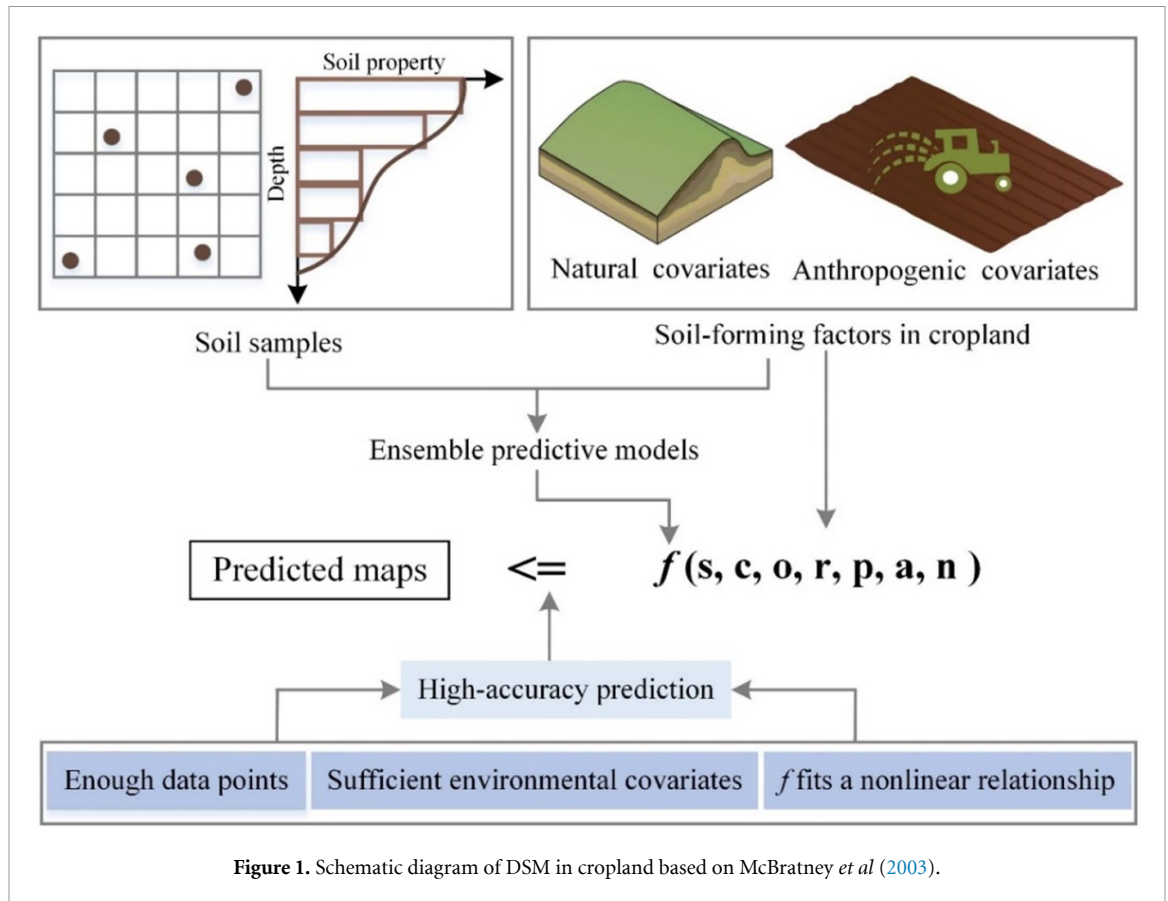


Figure 1. Schematic diagram of DSM in cropland based on McBratney et al (2003).

identify some challenges followed by an analysis on how to get further improvement. Finally, this review proposes future research directions and trends for digital mapping of soil carbon in cropland.

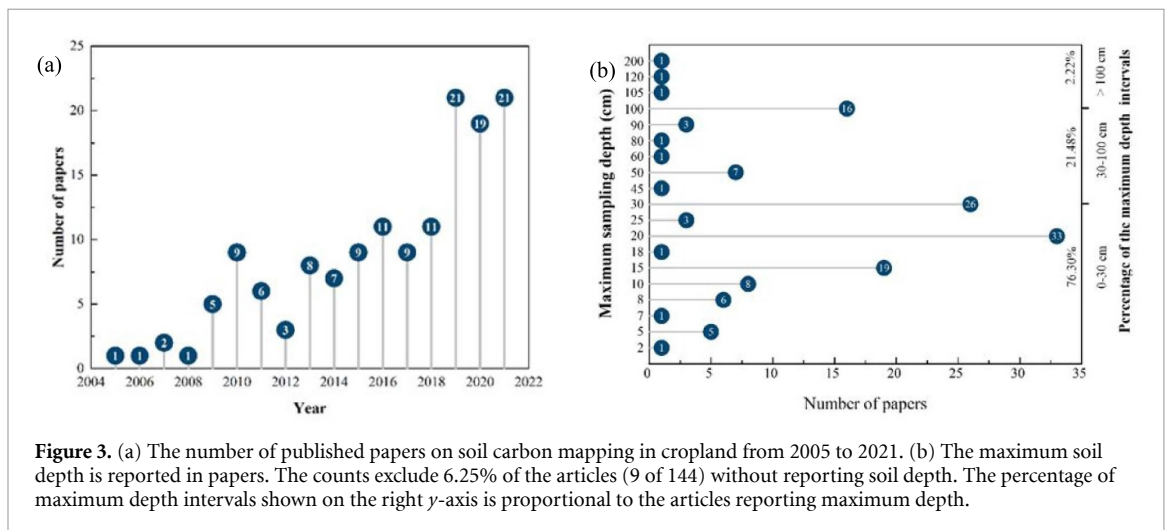
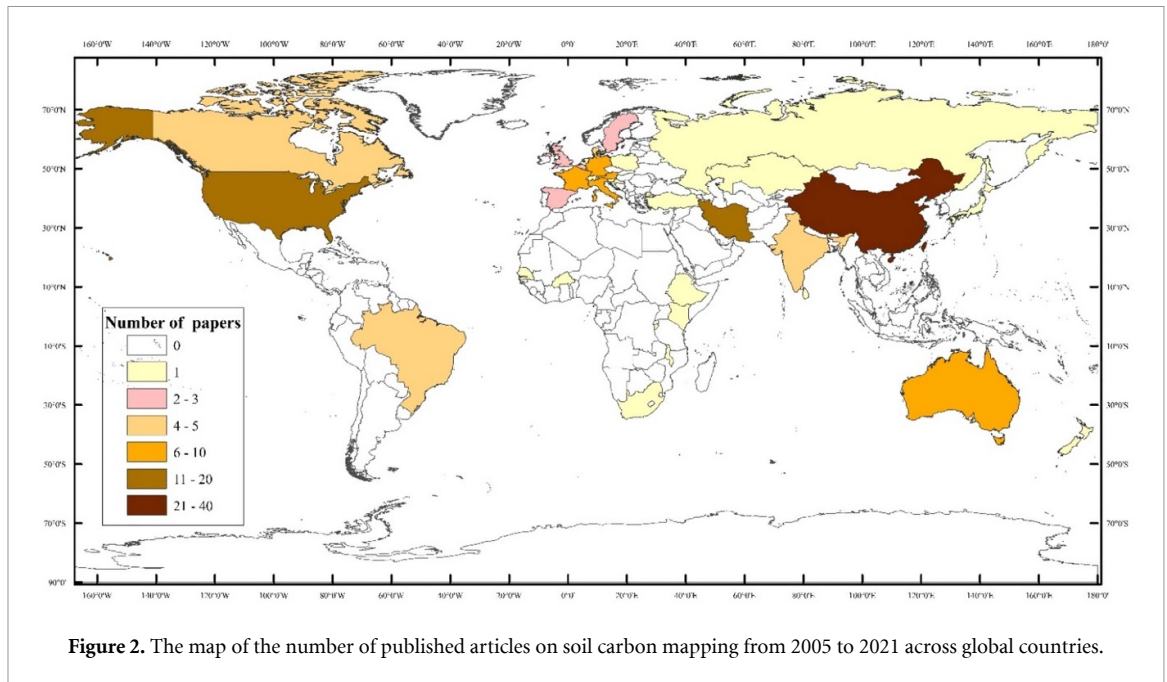
2. A summary of digital mapping of soil carbon in cropland

Focusing on the digital mapping of soil carbon in cropland, we conducted a literature search that was confined to only those employed environmental covariates in digital mapping of soil carbon studies (McBratney et al 2003) and published in English. Web of Science was queried by several expressions:

- crop* AND carbon AND map* AND 'digital mapping'
- crop* AND carbon AND map* AND 'digital soil mapping'
- crop* AND carbon AND map* AND predict*
- agricul* AND carbon AND map* AND 'digital mapping'
- agricul* AND carbon AND map* AND 'digital soil mapping'
- agricul* AND carbon AND map* AND predict*
- agricul* AND carbon AND map* AND 'digital mapping'
- field* AND carbon AND map* AND 'digital mapping'
- field* AND carbon AND map* AND 'digital soil mapping'
- field* AND carbon AND map* AND predict*

- crop* AND 'organic matter' AND map* AND 'digital mapping'
- crop* AND 'organic matter' AND map* AND 'digital soil mapping'
- crop* AND 'organic matter' AND map* AND predict*
- agricul* AND 'organic matter' AND map* AND 'digital mapping'
- agricul* AND 'organic matter' AND map* AND 'digital soil mapping'
- agricul* AND 'organic matter' AND map* AND predict*
- field* AND 'organic matter' AND map* AND 'digital mapping'
- field* AND 'organic matter' AND map* AND 'digital soil mapping'
- field* AND 'organic matter' AND map* AND predict*
- farm* AND 'organic matter' AND map* AND 'digital mapping'
- farm* AND 'organic matter' AND map* AND 'digital soil mapping'
- farm* AND 'organic matter' AND map* AND predict*

Each potential study was assessed suitability by title, then by abstract to estimate the possibility of using environmental covariates, and lastly confirmed according to the main text. From this search, we extracted 144 previous studies through Web of Science databases from 2005 to 2021.



2.1. Geographic distribution and temporal trend of the published studies

Figure 2 depicts the spatial distribution of the number of articles over the study period of this review. These researches were conducted in 34 countries, mainly including China (26.49% of articles), the United States (11.26% of articles), Iran (7.28% of articles), and Australia (6.62% of articles), etc. Basic national conditions (i.e. policy environment, historical foundation, and land resources) influence the distribution of soil map studies.

Figure 3(a) gives details of each year's publications during the period 2005–2021. A growing number of studies have suggested a higher demand for soil carbon information. Simultaneously, greater access to computational tools coupled with the development of prediction methods has promoted more efficient and rapid ways to map soil carbon in different regions

(Miller and Schaetzl 2014, Brevik *et al* 2016). Thus, the 21st century has become a period of booming research on the digital mapping of soil carbon.

2.2. Keywords analysis

Keyword frequency analysis can examine the hot research contexts. Figure 4 shows that key words about various predict models occur frequently, suggesting that the effective of models has become a common research field. Meanwhile, keywords relevant to technique to obtain environmental covariates (e.g. 'imaging spectroscopy', 'landsat', and 'sentinel-2') sustained rapid growth. This indicates that the spectroscopy methods are becoming more popular and accessible to researchers in the DSM. The emergence of 'climate change' indicates the growing concern of soil carbon in the science of climate change.

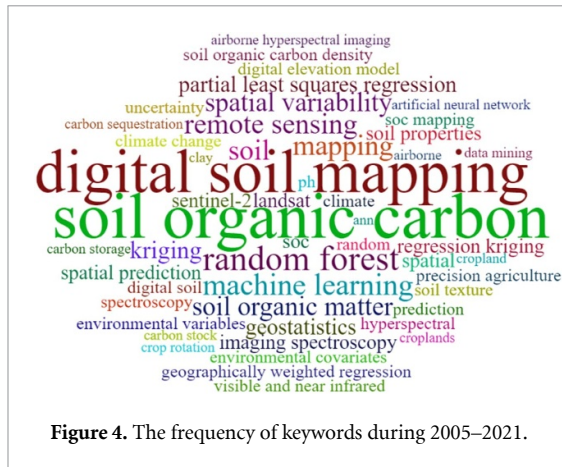


Figure 4. The frequency of keywords during 2005–2021.

2.3. Soil sampling and the maximum depth of interest

The design of sampling points is a vital, first step in DSM (Zhu *et al* 2008, Brus 2019), and affects the mapping result and accuracy of DSM (Piikki and Söderström 2019, Paul *et al* 2020a, Zhang *et al* 2022). The increasing study interest from mapping the spatial variation of SOC to mapping the dynamics of SOC requires an increasing demand for effective soil sampling. Common sampling methods include design-based sampling (e.g. stratified random sampling, systematic sampling) and model-based sampling (e.g. geostatistical sampling and centered grid sampling) (Zhang *et al* 2017a, Brus 2019).

Figure 3(b) presents the maximum sampling depth reported in 143 articles, 6.25% (9) of which did not report the sampling depth. Among the 135 articles reporting a sampling depth, 102 (76.30%) articles focused on soil information in topsoil with a sampling depth less than 30 cm. Studies on deep soil with a sampling depth larger than 30 cm were less represented and only 2.22% of the articles sampled soil larger than 1 m, though the knowledge of soil carbon at the depths plays a vital way in accounting soil carbon stock (Minasny *et al* 2013, Wu *et al* 2017). There are 116 articles focused on mapping soil carbon concentration or content (% or g kg^{-1}), but only 24 articles mapped area-based SOC stock (SOC density).

2.4. Scales, environmental covariates, models

There is a large range of case studies assessing soil carbon from the plot ($<1 \text{ km}^2$) to the regional ($>10^4 \text{ km}^2$) scale. The large variation between scales could be due to various research objectives, e.g. the impact of environmental covariates (Stevens *et al* 2014), the effectiveness of the model (Dharumarajan *et al* 2017), and the evaluation of SOC losses (Paul *et al* 2020b).

Based on the reviewed papers, we summarize the main drivers of soil carbon, environmental covariates, and models used for digital mapping of soil carbon in cropland on different scales (figure 5). Identifying the

main drivers of soil carbon on a special scale can not only improve the accuracy of DSM but also be helpful to understand how environmental covariates influence the spatial distribution of soil carbon. As shown in figure 5, a wealth of environmental covariates for DSM are available, including topography covariates derived from digital elevation models (DEMs), parent material, climate covariates generated by interpolation of observations from meteorological stations, remote sensing images of surface reflectance or transmission, and soil properties acquired by on-the-go sensors. Soil properties, vegetation, and topography tend to be important drivers of soil carbon from the plot ($<1 \text{ km}^2$) to the regional ($>10^4 \text{ km}^2$) scale, while climate, parent material, and anthropogenic influence appear to be vital factors at larger scales.

Model selection may be one of the core objectives of DSM. We divided DSM models of the reviewed 144 articles fall into four generic methods: linear statistical methods, geostatistical methods (mainly cokriging (CoK) in this review), machine learning (ML) methods, and hybrid methods. As we choose the digital mapping of soil carbon studies which used the environmental covariates (McBratney *et al* 2003), kriging methods are only based on sample distances but do not use environmental covariates (e.g. ordinary kriging (OK) and simple kriging (SK)) were not considered in this review. Besides, OK or SK has been reported not as effective as regression kriging (RK) or CoK (Mirzaee *et al* 2016, Wu *et al* 2021). According to figure 5, linear statistical methods and ML methods were the most commonly used algorithms. Linear statistical methods occurred frequently at a local scale and plot scale, while they are used frequently in earlier papers. ML methods are real heated methods that are dominant on all scales. It concerns the effectiveness of tackling the complicated relationships among various environmental covariates and soil carbon. Hybrid methods work in a broader range of scales, compared with linear statistical methods.

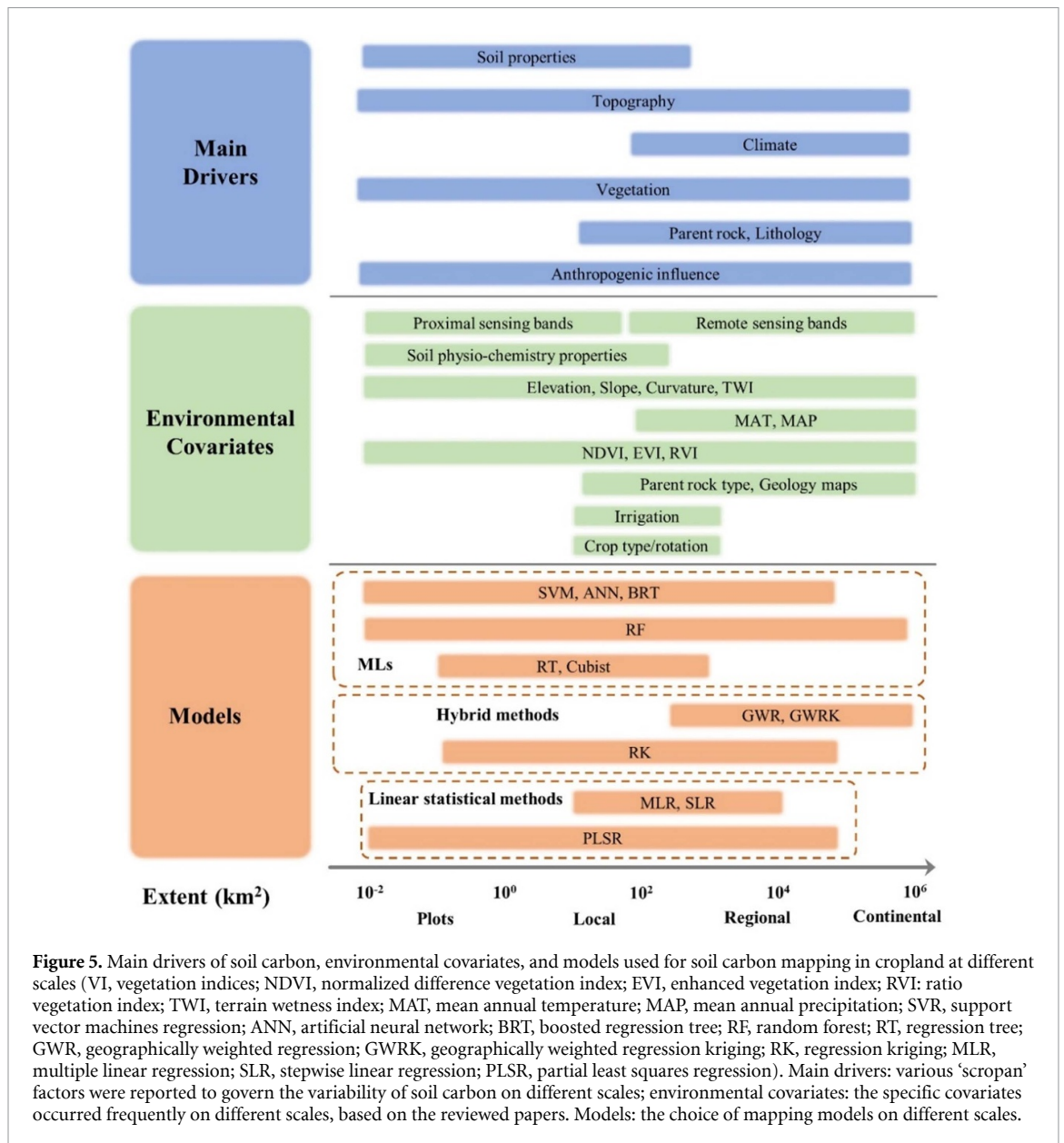
3. Models and environmental covariates of digital mapping of soil carbon in cropland

In this section, we focus on predictive models and environmental covariates in digital mapping of soil carbon in cropland to find advanced existing knowledge, aiming to highlight where further DSM studies in cropland can be carried out.

3.1. Models

3.1.1. Linear statistical methods

Linear statistical methods are commonly adopted in the digital mapping of soil carbon because of their easiness and wider availability. They form a specific model between a response variable (Y) and a set of dependent variables (X) (McBratney *et al* 2003, Summers *et al* 2011). Linear statistical methods included in the section are mainly regression



models using ordinary and generalized least squares. In particular, the two most common models are multiple linear regression (MLR) and partial least squares regression (PLSR).

MLR assumes a linear relationship between soil carbon and its environmental variables (Hounkpatin *et al* 2018). Currently, MLR has come up most frequently in the context of comparing the predictive efficiency of various algorithms for mapping carbon. For example, Khanal *et al* (2018) used five ML algorithms (random forest (RF), neural network (NN), support vector machine (SVM) with radial and linear kernel functions, gradient boosting model, and cubist) and took MLR for comparison for prediction of soil organic matter (SOM) (%). Zeraatpisheh *et al* (2019) gave an example for using non-linear models (RF, cubist, and regression tree (RT)) in mapping SOC content combined with topographic covariates

and remote sensing covariates, while MLR was chosen as the compared method.

Data from several studies suggest that MLR may be decent when the number of multispectral satellite imagery is limited, while PLSR can handle a large number of highly correlated spectral bands (Vaudour *et al* 2013, 2016). PLSR projects predictors (X variables) onto a low-dimensional space and finds a small number of latent variables that explain the covariance between X and Y (Stevens *et al* 2010, Summers *et al* 2011, Kuang *et al* 2015). As a result, PLSR can create the relationships between the input covariates and soil carbon in addition to avoiding the multicollinearity problem among the input covariates (Guo *et al* 2021a). PLSR was chosen to test the suitability of airborne hyperspectral imaging in the detection of SOC content variability at a small scale (Hbirkou *et al* 2012). Based on PLSR,

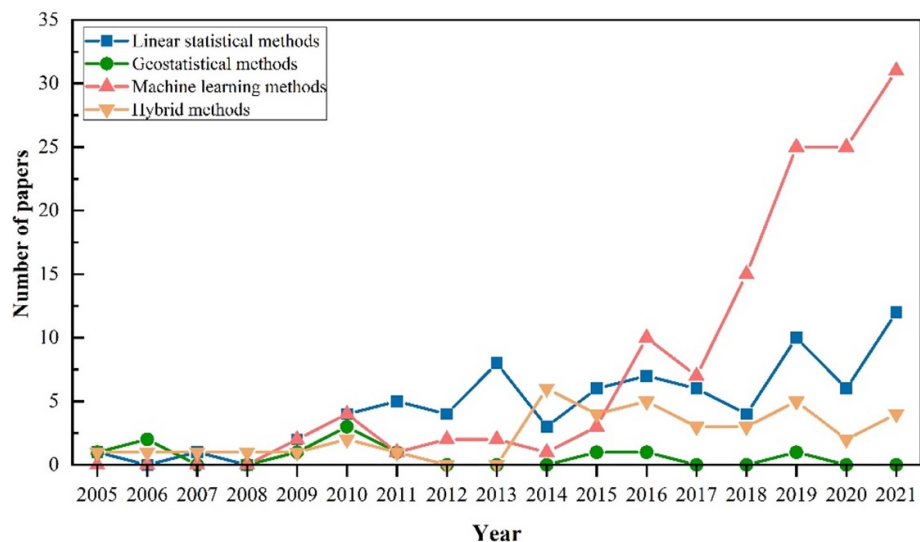


Figure 6. The number of predict models from 2005 to 2021. Note that many articles compared the model performance using several types of models, thus the sum of all the counts is far more than the total number of papers.

Guo *et al* (2019) and Kodaira and Shibusawa (2013) investigated the capability of bare soil spectra to map the topsoil SOM content in cropland. Study such as those conducted by Bartholomeus *et al* (2011) has shown that the PLSR-based prediction of SOC content is sensitive to vegetation influence when exploring to filter away the influence of maize from the mixed spectra.

There has been a significant shift from linear statistical methods to ML methods for mapping soil carbon in the last 20 years (figure 6). One reason is that linear statistical methods can suffer the problem of collinearity among variables. Moreover, the linear statistical methods may not satisfy the complex soil-environment relationships. Linear statistical methods occur frequently when environmental covariates consist of the electromagnetic spectra or near- and mid-infrared bands due to the inherent ability to handle many spectral data with co-linearity (Debaene *et al* 2014, Kuang *et al* 2015, Wetterlind *et al* 2015, Priori *et al* 2016), otherwise, linear statistical methods are frequently used as a benchmark upon which the performance of other models could be assessed (Lamichhane *et al* 2019).

3.1.2. Geostatistical methods

Geostatistical methods are effective in quantifying and model the spatial variation of the variable of interests assuming that samples close together, on average, are valued more similarly than those that are farther apart (Wu *et al* 2021), and have been widely applied in mapping of soil properties (Lopez-Granados *et al* 2005, Minasny and McBratney 2016). Nevertheless, these methods require fairly dense points data to obtain reliable semivariograms, and the relative importance of the different drivers of soil carbon dynamics is hardly reflected in geostatistical

methods (Lacoste *et al* 2014). In the review, we only focus on the geostatistical algorithms that incorporate environmental covariates into the kriging system. CoK is an extension of OK that employs one or more covariables (Z_2) in the estimation at unsampled locations by considering the correlations with the primary variable (Z_1) (Wu *et al* 2009). The semivariogram and cross-semivariogram of all Z_1 and Z_2 can be computed and modeled as a linear model of coregionalization (Simbahan *et al* 2006). Numerous examples of CoK applications can be found in DSM (Simbahan *et al* 2006, Mirzaee *et al* 2016, Dong *et al* 2019). Wu *et al* (2009) found that CoK with remote sensing data showed better performance than kriging when facing limited available soil samples. The results of these studies indicated that geostatistical models coupled with environmental covariates can be helpful in mapping soil carbon.

3.1.3. ML methods

ML methods do not require rigorous statistical assumptions about the distribution of input data, compared with geostatistical models (Lacoste *et al* 2014). They could effectively work with the non-linear and complex relationship between soil carbon and environmental covariates known as ‘SCORPAN’ factors (McBratney *et al* 2003, Lacoste *et al* 2014, Khanal *et al* 2018). The article defines ML methods as a large class of methods using data mining to learn a pattern and then perform regression or classification tasks (Wadoux *et al* 2020). The more frequently used models identified in this review name cubist, RF, SVM, and artificial neural network (ANN).

Cubist is different from the conventional type of RTs in that the prediction values are based on linear regressions instead of discrete values, thus leading to small trees and good prediction accuracy

(Minasny and McBratney 2008, Taghizadeh-Mehrjardi *et al* 2016). The advantage of cubist is that it allows an easy interpretation of the predictive models by producing the learning rules and the importance of the covariates between soil properties and its environment (Lacoste *et al* 2014). Moreover, Lacoste *et al* (2014) applied cubist to output the importance of the covariates on explaining the distribution of SOC content/stock at the landscape scale. Zeraatpisheh *et al* (2021) delivered a great potential in using cubist for mapping SOC content. Zeraatpisheh *et al* (2019) concluded that cubist showed superiority in predicting SOC content among other methods (i.e. RT and MLR) in the semi-arid region.

RF is an ensemble of randomized RTs that are aggregated to give a single prediction (Taghizadeh-Mehrjardi *et al* 2016, Pahlavan-Rad *et al* 2018). RF is robust against overfitting and can work out large datasets and superior predictive performance over many other models (Dharumarajan *et al* 2017, Hounkpatin *et al* 2018). An empirical example is by Hounkpatin *et al* (2018), who find RF performed better than the MLR due to the RF's advantage in handling nonlinear patterns in datasets. Dong *et al* (2019) indicated that the RF model performed better than OK, CoK, and ANN models for predicting topsoil SOC content, especially when sample data are limited. What's more, Guo *et al* (2021a) and Yang *et al* (2019) discussed the application of the RF model for predicting topsoil SOC content in an intensively cultivated region of China. In a watershed, RF shows lower uncertainty compared to the cubist model with not dense sampling density (Fathololoumi *et al* 2020). RF is now widely used in predicting soil attributes and seems to work well with diverse landscape features and limited sample size.

Support vector machines regression (SVR) implements linear/non-linear regression functions in a multi-dimensional feature space, thus it can solve the problem of nonlinear discrimination (Aldana-Jague *et al* 2016, Taghizadeh-Mehrjardi *et al* 2016). One of the great advantages of SVR is that it can deal with noise from high-dimensional data, therefore it is always selected to solve the multivariate calibration problem (Stevens *et al* 2010, 2012). However, selecting the right kernel and SVR parameters can require a lot of computing work or experience. Suleymanov *et al* (2021) applied SVR to map the SOC content in the arable lands of Trans-Ural steppe zone with 17 topographic indices. In the study, SVR is the more accurate method in predicting the spatial variation of SOC content with comparison to MLR. Xu *et al* (2020) concluded that SVR was suitable for the high-resolution mapping of SOC content based on laboratory hyperspectral imaging spectroscopy, which is aligned with the results of Aldana-Jague *et al* (2016).

ANNs are notable for their ability to efficiently manage and modify a large amount of data from a variety of sources with varying levels of noise and

precision, as well as their generalization capability (Sindayihebura *et al* 2017, Kalambukattu *et al* 2018, Nabiollahi *et al* 2018). Kuang *et al* (2015) recommended ANN for spectral calibration and prediction, because it can overcome the non-linear behavior documented for SOC content. Other examples include Kalambukattu *et al* (2018) and Taghizadeh-Mehrjardi *et al* (2016), who used ANN to map SOC content on local and regional scale, respectively. However, it is excessively time consuming, inefficient at handling mixed types of data, and susceptible to missing data and outliers (Lek and Guegan 1999). Dong *et al* (2019) used ANN based on multi-layer perceptron to predict SOC content, while ANN performed worst in four models (the other three are OK, CoK, and RF). This result may be caused by the ANN model's greater propensity for overfitting when sample data are scarce. However, in the context of big data mining, ANNs are still remarkable in DSM.

3.1.4. Hybrid methods

The hybrid methods are defined as the summation of at least two different algorithms, such as RK, random forest plus residuals kriging (RFRK), and RF plus geographically weighted regression. Generally speaking, such algorithms achieve better predictive performance than simple geostatistical algorithms or linear statistical methods (Bilgili *et al* 2011, Mirzaee *et al* 2016).

As one of the common hybrid methods, RK incorporates environmental covariates into the kriging models by combining regression between the primary (target) variable and secondary variable(s) using kriging residuals derived from the regression (Bilgili *et al* 2011, Piccini *et al* 2014). Simbahan *et al* (2006) advocated employing using independently measured, multivariate secondary information in RK approaches for mapping of SOC stock. Minasny *et al* (2009) used RK to demonstrate the application of near-infrared diffuse reflectance spectroscopy measurements in DSM. Wu *et al* (2020) evaluated the performance of RK combined with cropping systems and natural covariates. Even though RK has been confirmed to have the advantages of easy implementation, detailed results, and good prediction accuracy, its validity is strongly reliant on the selection of environmental covariates (Wu *et al* 2019).

Recently, some studies have focused on hybrid geostatistical procedures, mostly in the form of a combination of geostatistics and ML methods. For example, Mirzaee *et al* (2016) conducted ANN-OK and ANN-SK for the spatial variability of SOM content using remote sensing data. The result showed the hybrid geostatistical methods provided more reliable predictions than the simple geostatistical methods including SK, OK, and CoK. Dai *et al* (2014) concluded that ANN-kriging (ANNK) was particularly effective in improving the accuracy of SOM content mapping compared with universal kriging, and

inverse distance weighting at a large scale. This is mainly because the type of methods uses both environmental and spatially auto-correlated information based on residual estimation by kriging (Mirzaee *et al* 2016). Nevertheless, some studies found that kriging achieved a better performance than the hybrid geostatistical methods because the sample density could influence the involvement of environmental covariates (Li 2010, Pouladi *et al* 2019). Thus there is a need to estimate the correlation between environmental covariates and target property when using the interpolation methods by hybrid geostatistical methods such as RK and ANNK (Mirzaee *et al* 2016).

3.1.5. Other methods

Some other innovative methods that do not fall into the above categories for soil mapping have been proposed. Soil type-specific depth functions were introduced by Kempen *et al* (2011). This approach uses general pedological knowledge for defining depth function structures with geostatistical modeling then allows the construction of the depth function of SOM content for each soil type at each prediction site (Kempen *et al* 2011). However, the methodology requires a large number of existing soil datasets, which may be not applicable in areas with limited samples. Zhu *et al* (2015) presented a new method named ‘individual predictive soil mapping’, which can make good use of a limited quantity of soil samples based on their representativeness. Assuming similar soil attribute characteristics under similar environmental conditions, this method calculates similarities between environmental covariates of the unvisited points and the sampling points, then uses a similarity weighted average method to calculate the target attributes of the whole area (Zhu *et al* 2015).

3.1.6. Summary of soil mapping models

As the above study results show, no model is found to have consistently superior performance. Currently, many studies show that ML methods seem to be more efficient in incorporating non-linear correlations between soil carbon and environmental covariates during model construction, thus performing better than the linear statistical methods (Khanal *et al* 2018, Xu *et al* 2020). Bou Kheir *et al* (2010) pointed out that ML methods could satisfy the demands of indicating the relative importance of each environmental covariate and moderate requirements on the number of sampling points. Nevertheless, ML methods are data-driven and the quantity and representativeness of the input training observations determine the prediction results to a great extent. In some cases, ML methods could even give worse results than simple models in some cases. In a spatial downscaling study, Roudier *et al* (2017) pointed out that simpler methods such as linear statistical methods or general additive model outperformed RF and cubist to capture the important variations when facing a limited

number of training samples. What’s more, the performance of the same model is different in different soil depths. In the study of four standardized depths (0–15, 15–30, 30–60, and 60–100 cm), Taghizadeh-Mehrjardi *et al* (2016) indicated that the performance of predictive models were decreased with depth. The best performance of SVR in predicting SOC content was obtained in the 0–15 cm depth of soil profiles, while the lowest accuracy of estimation is for the 60–100 cm depth interval with far greater root mean square errors (RMSEs) than the first two layers (0–30 cm of soil profiles).

Different predictive models might suit different sets of environmental covariates and target soil properties (Beguin *et al* 2017, Zhou *et al* 2020). According to certain studies, the correlation strength of the target variable with environmental covariates can be a critical measure in determining which models to utilize *a priori* (Hengl *et al* 2004, Taghizadeh-Mehrjardi *et al* 2016, Hounkpatin *et al* 2018). Also, some predictive models are only capable of predicting while others can also provide soil-environment relationships (Roudier *et al* 2017). Some studies demonstrate that RFRK obtains promising performance in modeling complex relationships among variables, while generalized additive mixed model and RT are conducive to interpreting relationships among variables (Hamberg *et al* 2008, Kelly *et al* 2013, Guo *et al* 2015). Therefore, multiple complementary methods used make it possible to get better predictions than using these models individually.

3.2. Environmental covariates

Environmental covariates are used as predictors in the digital mapping of soil carbon, which are supposed to represent the environmental conditions governing the level of soil carbon. Thus, it is important to identify the contribution of these covariates for mapping soil carbon so that the selection of relevant environmental covariates can be guided. The review divides all environmental covariates into two main categories. One is natural environmental covariates, the other is anthropogenic covariates. The most common natural covariates are annual average temperature and precipitation, topography attributes (e.g. slope, aspect, terrain wetness index (TWI)), geological/soil parent materials maps, and previously measured soil attributes or class maps. The most common anthropogenic covariates include crop information (e.g. crop type, vegetation indices (VI) derived from satellite images) and agricultural management (e.g. irrigation, fertilizer). In this review, the time factor is not as a separate section like other common covariates. Soil carbon dynamics is a time-dependent process (Bolinder *et al* 2020), while it is difficult to get appropriate factors to measure the influence of ‘time’. To express the role of ‘time’, some studies used the long-term climatic covariates (Schillaci *et al* 2017) or VI (Yang *et al* 2021), cultivation history

(Wang *et al* 2019), or legacy soil maps (Costa *et al* 2018). It seems the role of 'time' is reflected in these accessible covariates.

3.2.1. Natural environmental covariates

3.2.1.1. Climate covariates

Climate is a pivotal force affecting plant type, crop growth, and soil conditions, thus impacting carbon input and litter decomposition (Wiesmeier *et al* 2019). A higher SOC concentration is frequently seen in places with high rainfall and low temperatures, which encourage the accumulation of carbon and nitrogen in soils (Hounkpatin *et al* 2018). High precipitation is typically correlated with abundant growth and high rates of carbon inputs to soils, whereas low temperatures may noticeably slow down the microbial decomposition of organic matter (Bai *et al* 2019). What's more, the crop species selected for regional agriculture are usually recommended by regional climatic characteristics.

The (long term) annual average temperature and average annual precipitation are the climatic covariates commonly used in mapping soil carbon (Zeng *et al* 2016, Wang *et al* 2017). Li *et al* (2020a) stated the annual average temperature was the most important explanatory variable for predicting SOC content in black soil area of northeastern China. Moreover, other climatic covariates, such as maximum/minimum temperature (Tayebi *et al* 2021), maximum/minimum precipitation (Sreenivas *et al* 2016), and moisture index (Dong *et al* 2019) have also been found to be powerful in explaining the variability of soil carbon. Some studies have shown that climate changes will affect gross productivity, the decomposition of microorganisms (Piao *et al* 2013, Anderson *et al* 2020). In turn, changes in the carbon-cycle can affect climate on a variety of scales (van Meij *et al* 2018, Messori *et al* 2019). The relationship between climate covariates and organic matter becomes highly complex in the context of global warming.

Climatic conditions may be the main drivers of soil carbon on larger scales, such as regional and continental scales. However, at a small scale, sometimes, it is not always an effective covariate. Dong *et al* (2019) found annual average precipitation ranked lowest among the ten environmental covariates in the study of arid agricultural plain (509.17 km²). Paul *et al* (2020b) found climate variables performed less well than other environmental covariates (VI, topographic covariates, and soil covariates) for predicting SOC content in a study area of 120 km². This trend might be explained by the cultivation always in low-elevation plains or hillocks where the land is relatively flat. On small scale, climate effects may be masked in topsoil by land use/management, particularly in cropland soils, where climate effects may be mediated by intensive management (fertilization, irrigation, etc) (Dong *et al* 2019, Paul *et al* 2020b). This shows other environmental covariates than climate covariates may

play a more important role in SOC decomposition and forming of soil carbon at a small scale.

3.2.1.2. Parent materials

Parent material is formed based on bedrock, which is the basis of the physical and chemical properties of the soil (Zhang *et al* 2020). For example, the soils, derived from carbonate rocks in Southwest China, are rich in calcium (Tu *et al* 2018). As the original material sources, parent materials is related to soil properties such as soil texture, pH, and clay mineralogy, subsequently affecting other soil characteristics, including SOC content and the soil's potential to store carbon (Ingram and Fernandes 2001). However, the parent material is not often available in many areas, and is commonly generated from geological maps of different scales, such as 1:500 000 geological maps in China (Zeng *et al* 2016) and 1:25 000 National Geological Map of Denmark (Bou Kheir *et al* 2010).

Studies have employed parent material in predicting SOC content variability from regional to plot scales. Dong *et al* (2019) pointed out that soil parent material was a vital factor influencing SOC content at a regional scale. The finding was in line with the results of Bou Kheir *et al* (2010), who reported parent material as a major factor affecting SOC content in the 5748 km² area mainly farmed with cereals. Kumar (2015) concluded parent material could significantly increase the prediction accuracy of the SOC density in the study of the Midwestern region of the United States. However, the prediction ability of the categorical parent materials variable is limited in a 15 km² study area mainly due to its small variability (Sindayihebura *et al* 2017). Similarly, a study covering an area of 10 km², found that geological data was a less significant factor in predicting SOC stocks (Lacoste *et al* 2014). Overall, selecting parent material as a predictor seems to be a wise choice on a regional scale, whereas is needed to weigh the options at local and plot scales.

3.2.1.3. Topography covariates

Topography condition determines soil particle movement, thus affecting the distributions of soil carbon content. Topographic covariates are generally extracted from existing DEM including slope, elevation, aspect, TWI, plane curvature, and profile curvature (Lacoste *et al* 2014). Specifically, slope affects soil erosion, drainage, and runoff, while TWI reflects the state of soil water, therefore they both played important roles in the spatial patterns of SOC contents in dry farmed cropland (Tu *et al* 2018).

Topographic covariates have been widely adopted in digital mapping of soil carbon and acknowledged as an influential factor, especially at watershed scales. Zhou *et al* (2020) considered topography as one of the main factors affecting soil moisture-temperature conditions, and it proved to be the main explanatory variable for SOC content prediction at

coastal agroecosystems. Bonfatti *et al* (2016) found that for soils below 30 cm depth, aspect was one of the important factors for predicting SOC content. For wet cultivated lands in the study of Bou Kheir *et al* (2010), topographic covariates such as elevation and plan curvature contribute more to the prediction of the spatial distribution of SOC content than other factors such as normalized difference vegetation index (NDVI), soil color, and flow accumulation. Similar results were reported in other studies (Wang *et al* 2018, Hu *et al* 2021, Zeraatpisheh *et al* 2021), which illustrated the important role of topography in the digital mapping of soil carbon. In the alluvial plains of Ningxia province, China, elevation is only one factor that significantly affects the distribution of SOC content among ten topographic covariates, which might cause by the flat land in agriculture area (Dong *et al* 2019). High altitude regions typically have lower temperatures, which may partly explain the accumulation of SOC (Tu *et al* 2018). Nevertheless, the elevation was also a proxy for deposited material and high SOC stocks were located in lower elevation areas with deeper soils (Zeraatpisheh *et al* 2019). The relative importance of elevation varies widely for different agricultural area due to their specific landforms and climate conditions.

These studies demonstrate that topography covariates are strong predictors for predicting soil carbon at different scales, extents, and depths. Yet, the topographic covariates may be not effective in relatively flat areas because they may be too homogenous to reflect soil conditions well in gentle terrain, usually in cropland (Wu *et al* 2020, He *et al* 2021).

3.2.1.4. Soil covariates

Soil properties such as pH, clay content, soil moisture content, and bulk density affect the accumulation and decomposition of soil carbon (Wu *et al* 2020, Naspendra *et al* 2021). Soil pH may influence the SOC turnover rate through its effect on microbial species and diversity, which is a comprehensive indicator of the physico-chemical characteristics of the soil. Fine-textured soils tend to have more organic matter (Bai *et al* 2019). And the reason for this is twofold: clay plays a significant part in stabilizing carbon by adsorption and aggregation; clay-textured soils have greater water retention capacity and affect plant yield (Jackson *et al* 2017, Fan *et al* 2020). Therefore, existing soil property data or maps of an area have been used as predictors for soil mapping (McBratney *et al* 2003). In addition, different soil types have different soil properties, thus existing soil type maps are also widely used in soil mapping.

Soil properties, such as soil bulk density, texture, pH, have been reported to be influencing factors in determining the level of SOC. Deng *et al* (2020) used soil bulk density to map soil carbon/nitrogen (C/N) ratio based on 29 927 topsoils (0–20 cm) samples, and found played a crucial role in soil

mapping. Soil sand content showed a negative correlation with SOC content variation, whereas soil silt content appeared a positive relationship at depth of 0–30 cm (Hamzhepour *et al* 2019). Examples of studies using soil texture for mapping soil carbon are Schillaci *et al* (2017), Meersmans *et al* (2008), and Stevens *et al* (2015). Zhang *et al* (2017b) indicated that the soil pH was the most important variable impacting SOC content, followed by silt contents. However, some studies noted that the explanatory power of soil properties on a local scale was limited, which may be due to the influence of long-term cultivation and relatively rough resolution (Wu *et al* 2020).

In addition to the soil property sample data, a growing number of studies have advocated the effectiveness of existing soil information as covariates in the digital mapping of soil carbon. For example, Paul *et al* (2020b) utilized a 30 m grid to extract soil sand, silt, clay, and cation exchange capacity from the existing detailed Canadian soil survey, and the historic soil property variables were the strongest predictors for predicting SOC content. Dong *et al* (2019) obtained five soil properties (bulk density, clay content, silt content, sand content, and coarse fragments volumetric) from the SoilGrid 250 m dataset (Hengl *et al* 2014) to map SOC content on an alluvial-diluvial plain. Bou Kheir *et al* (2010), Costa *et al* (2018), and Wu *et al* (2019) selected the legacy soil class map as a covariate. These studies indicated that the existing soil layers should be a valuable source of information to build soil mapping models.

3.2.2. Anthropogenic covariates

Covariates indicating cropping system and agricultural management in digital mapping of soil carbon have attracted increasing attention recently (Han *et al* 2018, Zhou *et al* 2019, Zhang *et al* 2020).

Developments in remote sensing, geographic information systems (GIS), and mobile technologies combined with field investigation data are providing available ways to capture information on crop growth and agricultural management practices. There is immense potential in using agricultural information as predictors in the digital mapping of soil carbon, and the application of this kind of variable is gradually emerging. Some studies even point out that the results of the evaluation of soil carbon in crop areas are inaccurate without considering agricultural information (Simbahan *et al* 2006, Bell and Worrall 2009, Nawar *et al* 2015).

3.2.2.1. Crop information

Crop species and cropping rotations influence soil carbon dynamics due to the strong interactions between SOC pool and vegetation (Lal 2011). The use of cropping patterns and various VI is of great concern in the digital mapping of soil carbon (Araya *et al* 2016, Yang *et al* 2019).

3.2.2.1.1. Cropping patterns

Cropping patterns include planting crop species, area, and spatial layout of crops. Crop species, usually a result of farmers' decisions according to climate zone and market or policy, are highly related to soil carbon content (Wu *et al* 2021). Because crop residues of different species influence soil carbon input and crop roots are related to microbial activities. The addition of fresh organic amendment into the soil is known as soil priming (Kuzyakov *et al* 2000). The soil priming usually causes carbon dynamics (Mandal *et al* 2020). Based on the National Cropland Data Layer of the National Agricultural Statistics Service, Flathers and Gessler (2018) extracted five crop species to develop a SOC content map of major cereal crop-growing regions in the northwestern United States. Moreover, a shift of crop species and the corresponding cropping system usually results in changes in soil carbon. For example, the SOC content increased under a change from a dry field to a paddy field (Huang *et al* 2007a).

In addition to crop species, cropping rotation mode, the sequence of crop species planted in a year is another variables influencing the soil carbon dynamics (Wu *et al* 2021). Hu *et al* (2021) revealed that soils with a double-cropping system had the highest of SOM content compared with a single cropping system. Song *et al* (2017) generated a map of crop rotation types based on remote sensing data, and showed that using the crop rotation type variable increased the SOM content prediction accuracy more than NDVI or land use. Yang *et al* (2019, 2020) used the crop rotation information obtained through field investigation in a largely agricultural area, and showed the effectiveness of incorporating such information in SOC content prediction. Ellili *et al* (2019) concluded that crop rotation accounts for 20% stock of SOC change variability in eastern Brittany of France with a 10 km² area. Additionally, Wu *et al* (2019, 2021) employed crop rotation in mapping SOC density in plains.

3.2.2.1.2. VI of crops

VI indicate vegetation growth status, which are obtained based on multispectral or hyperspectral remote sensing bands. VI are easily-obtained, and thus have provided an available way to monitor inter-annual variations and long-term trends in vegetated land surface characteristics. Due to the strong relationships between soil and vegetation, VI have been used as predictors in mapping SOC content or stock for a long time (Simbahan *et al* 2006, Rizzo *et al* 2016, Fatholouloumi *et al* 2020, Li *et al* 2020a).

Several VI have been developed using different algorithms to describe plant characteristics. NDVI, one of the commonest VI, has been widely used. Multi-year average NDVI, as well as remote sensing imagery in several growing seasons, are mostly adopted. Commonly used NDVI products are MODIS NDVI data (Guo *et al* 2015), Landsat

TM data (Guo *et al* 2015), and Landsat ETM+ data (Sindayihebura *et al* 2017). Other VI (table 1), such as enhanced vegetation index (EVI) (Zeng *et al* 2016), soil adjusted vegetation index (Nabiollahi *et al* 2018), and ratio vegetation index (Zeraatpisheh *et al* 2019), have also been widely used as explanatory covariates in the predictive mapping of soil carbon. Kalambukattu *et al* (2018) discovered brightness index, coloration index, hue index, and saturation index were directly proportional to the amount of SOC content in a study area in the northwestern Himalayas and obtained an accuracy of 0.62 (R^2) with the above multiple VI and topographic covariates. Meanwhile, there are also studies using net primary productivity (NPP) to predict SOC stocks distribution (Martin *et al* 2011). This is mainly because NPP has a great relationship with vegetation type, thus it can indicate the amount of litterfall.

Compared with monotemporal VI, it was reported that time-series VI data capturing the spatial-temporal change of crop growth could be better predictors for digital mapping of soil carbon. Time-series VI can be regarded as more effective information to capture the long-term growth status of the crop (Zeng *et al* 2020b), furthermore, the VI vary in different years on account of crop patterns (Guo *et al* 2021b). In low relief cropland regions, Guo *et al* (2020) provide the feasibility of a short period (about five months) of multi-temporal images in mapping SOC stocks. Tayebi *et al* (2021) used 30 year time-series average VI (NDVI, EVI, normalized difference water index) to map SOC stocks and obtained prominent prediction results. These studies give us insightful inspiration on the application of time-series VI in digital mapping of soil carbon, which can be regarded to overcome the limitation of single images in cropland.

Now, innovative environmental covariates extracted based on time-series VI has been recently developed in the digital mapping of soil carbon (Zeng *et al* 2020b). Yang *et al* (2019) stated the time-series NDVI derived through Fourier transform resembled different periodic characteristics of crop rotation, thus the Fourier decomposed variables improved the accuracy of digital mapping of soil carbon at a regional scale. In addition, remote sensing-based phenological parameters, that describe the beginning and end time of the growing season, the length of the growing season, the initial growth rate, the withering rate, and other characteristics (Brown *et al* 2012), have been applied as a new type of environmental covariate in digital mapping of soil carbon. For example, Yang *et al* (2020) used 11 phenological parameters extracted from time-series NDVI in two growing seasons of a year, and suggested that the phenological parameters effectively indicated the spatial variability of SOC content in the region. He *et al* (2021) extracted 17 phenological parameters from 34 Sentinel-2 time-series images from 2018 to

2019 to predict SOC content in Anhui Xuancheng, China. Yang *et al* (2021) obtained ten-year phenology variables for mapping SOC content with a convolutional NN model, and showed that the time-series phenology variables were more effective predictors in improving accuracy than NDVI.

Taken together, these results show great promise of remote sensing techniques for providing valuable information on vegetation types and their growth status. Compared with traditional natural environmental variables, remote-sensing-derived predictors require less cost of acquiring data, thus representing a promising tool in the cropland's soil mapping. However, we need to tackle the influence of weather conditions and the return period of satellites when collecting images. Moreover, topography and climate may influence the efficacy of VI variables because the distribution of vegetation was bound up to topography and climate (Wang *et al* 2018, 2021b).

3.2.2.2. Agricultural management

Agricultural management including irrigation, fertilization, and straw returning has a major effect on the quantity and quality of fresh organic matter inputs into the soil and influences the organic carbon decomposing (Simbahan *et al* 2006, Ellili *et al* 2019, Luo *et al* 2019). Specifically, fertilization directly changes the soil structure and the level of SOC stock (Han *et al* 2018). Meanwhile, it promotes the growth of crops, thereby increasing the input of organic matter and contributing to accumulation in both topsoil and subsoil (Dendoncker *et al* 2004, Huang *et al* 2007a). Straw returning increases soil carbon input (Huang *et al* 2007a), improves soil structure and nutrient availability, and increases soil microbial population size (Li *et al* 2020b). No-tillage slows down soil carbon emissions and increases soil water retention capacity, which in turn increases SOC stock (Abdalla *et al* 2016, Li *et al* 2019, 2020b). The irrigation and drainage system play an important role in the regulation of soil moisture, which will affect the activities of microbial communities and then the decomposition rate of SOM density (Trost *et al* 2013, Sreenivas *et al* 2016).

Much recent literature has attempted to use irrigation information in the digital mapping of soil carbon. Sreenivas *et al* (2016) extracted the irrigation status from the irrigation atlas by Water Resources Information System project. However, most of the study areas do not have available irrigation maps. Cropland generally has well-developed irrigation systems, so the distance from the main channels or water sources is a more available variable representing the irrigation information. For example, Zeng *et al* (2016) used the distance from the nearest drainage system and the height of the closest drainage system as environmental covariates in gentle cropland and proved that these covariates played an important role in the distribution of SOM concentration.

Dong *et al* (2019) calculated the distance to rivers and used it to map SOC content in an alluvial-diluvial plain in China, and an irrigation class map with three classes (well-irrigated, moderately well-irrigated, and poorly irrigated) distinguished by local experts was also employed. The distance to rivers and the irrigation class map were the most important among the ten covariates (Dong *et al* 2019). Cambule *et al* (2013) analyzed the spatial characteristics of NDVI during the dry and wet seasons and found that NDVI along the drainage system lines was higher than in other places in the dry season. According to the above studies, the irrigation information would be a useful exploration especially for regions highly depending on irrigation, such as in arid or semi-arid areas.

Fertilizer information has been used to improve the accuracy of predicting soil carbon. For example, Deng *et al* (2018) obtained the county-level nitrogen-fertilizer application rates, which proved to be able to explain a part of SOC stock variation in Zhejiang province ($\sim 102\,646\text{ km}^2$) in East China. Furthermore, they emphasized the importance of the nitrogen-fertilizer application rate in mapping the C/N ratio (Deng *et al* 2020). Meersmans *et al* (2012) obtained Manure application and animal excrement production statistics (t ha yr^{-1}) at departmental level from French Environment and Energy Management Agency. The relation between SOC content and manure was not clearly identified due to the coarse spatial scale of manure data. Yet, it is ambitious to obtain detailed spatial fertilizer information because fertilization mode appears personal. The county-level fertilization information obtained through the statistical yearbook can be a reasonable variable for regional studies but may be not informative at local scales.

The historical agricultural management information representing the impact of human activities on agricultural ecosystems on a long-time scale attracts attention. A study on an estate in the United Kingdom suggested that farm tenancy was a better predictor of mapping SOC concentrations than soil series or land use (Bell and Worrall 2009). The length of cultivation for the past 300 years in Northeast China has been proved to improve the prediction accuracy of SOC content (Wang *et al* 2019). In Kravchenko and Robertson (2007), the dense long-term yield data obtained via yield monitoring from plots improved the total carbon mapping accuracy. Those studies show that it is an attractive attempt to introduce historical information (such as planting length, planting policy, yield information) into the digital mapping of soil carbon.

Although recent studies have proved the effectiveness of agricultural management information, it remains difficult to obtain detailed information on the spatial distribution of agricultural practices. The subjectivity of determination of agricultural practices by farmers makes it hard to obtain the agricultural management information with a detailed spatial

resolution (Han *et al* 2018, Luo *et al* 2019). This unfortunate reality can be a barrier to involving agricultural management in the digital mapping of soil carbon, which thus needs a coordinated effort to expand these capabilities.

3.2.3. Soil information based on proximal and remote soil sensing

Spectral signatures of soil are defined by their reflectance or absorbance, as a function of wavelength in the electromagnetic spectrum (Brown *et al* 2006, Ladoni *et al* 2010). The visible and near-infrared (VNIR: 400–1100 nm) and shortwave infrared (1100–2500 nm) bands of the electromagnetic spectrum have proven to be highly related to various components of organic materials when the area is bare, dry, and flat (Simbahan *et al* 2006, Hbirkou *et al* 2012, Debaene *et al* 2014). Soil spectra usually are measured in laboratory conditions, *in-situ* conditions by portable spectrometers (e.g. a sensor is mounted on a tractor and its position recorded by a global position system receiver) (Chen *et al* 2000, Fox and Sabbagh 2002, Martin *et al* 2002, Lopez-Granados *et al* 2005, Bricklemeyer and Brown 2010), or remotely-sensed aerial photographs (e.g. the sensor records the surface aboard a plane or satellite) (Stevens *et al* 2010, Bilgili *et al* 2011, Stevens *et al* 2012, Kodaira and Shibusawa 2013). Proximal and remote soil sensing are believed to be a cheaper way to obtain the spatial distribution characteristics of soil compared with conventional laboratory analysis (Minasny *et al* 2009, Miklos *et al* 2010). Moreover, an intriguing fact is that, by adopting this method, there can be enough soil data to allow meaningful (geo)statistical analyses to obtain a reliable assessment of soil carbon on the field.

Proximal soil sensing is defined as *in-situ*, mobile (on-the-go), field based-sensors for sensing soil information promptly (Minasny and McBratney 2016). Proximal soil sensing has been generally applied over different regions and yields comparable results to traditional analytical techniques (Stevens *et al* 2010, Hbirkou *et al* 2012). Huang *et al* (2007b) used on-the-go near-infrared spectroscopy measurements in a 50 ha agricultural field and concluded that the measurements are valuable tools for accurate total carbon content mapping. Recently, gamma-ray spectrometry and radar sensors have also been used (Flynn *et al* 2019, Piikki and Söderström 2019, Zhang *et al* 2020).

Remote soil sensing provided an important data source for the spectra of soil carbon, especially for bare soil. The different sensor types can usually be mounted on either airborne systems (Laamrani *et al* 2019, Guo *et al* 2021b) or as well as satellite platforms (Safanelli *et al* 2021). Vaudour *et al* (2021) used Sentinel-2 time series over several years to obtain soil composites in a heterogeneous, temperate environment, and concluded that Sentinel-2 was very encouraging for further studies about maximizing bare soil

areas in cropland. On a field scale, Hong *et al* (2020) demonstrated that the airborne hyperspectral image could be served as a valuable predictor to map agricultural topsoil SOC content.

Although we can get more soil spectral data from proximal or remote sensing, soil spectroscopy is limited in areas with large vegetation cover. To address this issue, various spectral unmixing techniques have been proposed for segregating bare soils from vegetation cover. Bartholomeus *et al* (2011) applied a method named Residual Spectral Unmixing within fields were partially covered with maize. Recently, a myriad of studies used multi-temporal images to maximize the bare soil coverage thus providing the ability to more accurately map soil spectra (Luo *et al* 2022). Diek *et al* (2017) used a dense time series of Landsat images to detect the least-vegetated area over the Swiss Plateau. Roberts *et al* (2019) provided the first continental-scale mosaic of Australia at its barest state based on the petabyte-scale of Landsat datasets. Vaudour *et al* (2021) highlighted the capability of multi-date Sentinel-2 images to composite bare soil images for predicting SOC content over cropland.

Additionally, soil spectroscopy has remained a unique concern due to platform reliability (i.e. the mounted sensor's spectral range and the limited flight duration), spectral instability effects, and issues regarding image processing (Angelopoulou *et al* 2019, Wang *et al* 2021a). The condition of the soil surface, such as soil moisture and surface roughness, potentially impacts the infrared absorbance, then may influence the prediction accuracy of hyperspectral images (Minasny *et al* 2009, Lu *et al* 2013). For example, Bricklemeyer and Brown (2010) pointed out that the sensor passing soils being collected at different physical locations could result in various wavelengths.

Furthermore, soil spectroscopy could be conducted in typical cropland which has certain representativeness for specific landscapes. Till now, The Soil Spectral Library has been proposed and built (Brown *et al* 2006, Viscarra Rossel *et al* 2016), which takes varying soil surface conditions into account and seems to be a step in the right direction (Hbirkou *et al* 2012). Scientists from eight countries (representing Kenya, Israel, China, Sweden, Germany, United States, Australia, and Brazil) are engaging in the global soil VNIR spectral library, to establish the correlations between the soil attributes (e.g. SOC, CaCO₃, and pH_{Water}) (Viscarra Rossel *et al* 2016). The local soil spectral library enables less effort and cost compared to analytical wet chemistry methods to estimate soil carbon in cropland (Bellinaso *et al* 2010, Rizzo *et al* 2016, Viscarra Rossel *et al* 2016).

3.2.4. Environmental covariates selection

Different combinations of environmental covariates would influence the prediction accuracy, thus selection of the 'optimal' combination has been

increasingly applied (Zeraatpisheh *et al* 2021). A growing number of studies have advocated diverse tailored approaches to select environmental covariates. Ideally, the most effective covariates not only require a small subset of the relevant features with low multicollinearity, but also could interpret the driving features of SOC formation.

One common method to select appropriate covariates for mapping is based on the expert knowledge of soil-environment processes (Zhu *et al* 2010). However, when faced with a large number of covariates, the expert knowledge may be not comprehensive, thus the so-selected covariates could influence prediction accuracy (Shi *et al* 2018, Hu *et al* 2021).

Overall, most papers tend to use statistical methods or ML methods. Commonly, in MLR models, the stepwise technique was used to select the covariates based on the Akaike information criterion (Costa *et al* 2018). Correlation analysis (e.g. Pearson) can be used to avoid multicollinearity effects. Environmental covariates are selected when the coefficient between environmental covariates and soil carbon is greater than a specific determined (Tayebi *et al* 2021). Moreover, pedometricians also joined forces to use ML methods (e.g. recursive feature elimination) (Gregorutti *et al* 2017) on selecting covariates (Shi *et al* 2018). Studies demonstrate that the variable importance index (e.g. mean decrease in accuracy) calculated by RF can facilitate success in evaluating the predicting power of environmental variables (Grimm *et al* 2008, Behrens *et al* 2014, Yang *et al* 2019, Paul *et al* 2020b). These studies remain innovative and significant in targeting the effectiveness of environmental covariates selection. At the meantime, ML-based approaches rely on large training sample data. To combine the data-driven selection results and expert knowledge on soil carbon dynamics could be a more reasonable way (Wadoux *et al* 2021).

3.2.5. Summary of environmental covariates

Various covariates for mapping soil carbon have been proposed and implemented in cropland. Different influential covariates are found among different scales. At the regional scale, SOC content is mainly affected by climate covariates, soil parent material, and soil type, while on a local scale, the differences in SOC content often depend on topography, agricultural management styles, and soil properties (Dong *et al* 2019).

Spatial variability of soil carbon is frequently too fine to be recorded by a coarse sample grid on a field scale. At a local/regional level, the issue becomes even more pressing. Soil carbon variability caused by local transfer processes (e.g. erosion) combines with variability caused by wider pedogenetic factors (climate, soil) at these scales, resulting in sometimes substantial uncertainty in regional soil carbon estimates (Stevens *et al* 2010). The importance of environmental covariates is largely dependent on the study site's specific

environmental covariates, notably topography, and there is no consensus on the best covariates. To highlight the agricultural management information of specific plots, Dong *et al* (2019) have proposed the idea of 'land-parcel'. Environmental covariates were adjusted according to 'land-parcel' and mapping with 'land-parcel' reduced the noise of unrelated environmental covariates and highlighted the boundary that was consistent with the actual field.

3.3. Validation and the mapping of uncertainty

Validation and quantifying the prediction uncertainty are important to provide the quality information of the mapping results. Data-splitting technique and cross-validation were common methods for evaluating the predict results. A recent example on validation of a soil map by data-splitting is given by Hu *et al* (2021). One issue with data splitting is that it is not clear how to divide a data set so that accurate estimations of map accuracy may be achieved. Variation in environmental conditions between the training and validation data could result in influencing accuracy of models (Bonfatti *et al* 2016). Increasingly, researchers use several repeats of data-splitting to reduce randomness. In case of relatively small sample size, cross validation (e.g. k-fold cross-validation, leave-one-out cross-validation, repeated k-fold cross-validation) is selected to validate the models (Yang *et al* 2021, Zeraatpisheh *et al* 2021, Zhang *et al* 2021). This differs from data-splitting in that cross-validation repeats the splitting, making it more effective and reliable than data-splitting (Brus *et al* 2011). The RMSE, Lin's concordance correlation coefficient, mean absolute error, the ratio of performance to deviation, and coefficient of determination (R^2) values are always used for determining final prediction accuracy (Kumar 2015, Hong *et al* 2020, Wu *et al* 2021).

DSM predictions are coupled with uncertainty (McBratney *et al* 2003). Model parameter, model input, and model structure could result in model uncertainty (Minasny and McBratney 2002, Liddicoat *et al* 2015). It is crucial to quantify the predictability uncertainty of the maps to confirm their suitability for managerial decision-making processes (Brus *et al* 2011). There are several approaches can be adopted for quantifying the uncertainty of prediction. When using the geostatistical model (e.g. universal kriging) for mapping, the prediction uncertainty can be calculated directly via the kriging variance (Li 2010). Another way is to use the bootstrapping approach, to obtain probability distributions of the outcomes. The 90% or 95% confidence intervals, which is usually used to quantify the uncertainty of soil carbon predictions (Adhikari and Hartemink 2017, Ellili *et al* 2019, Fatholouloumi *et al* 2020). In data partitioning (Malone *et al* 2014) and fuzzy clustering approach (Shrestha and Solomatine 2006), the uncertainty is expressed in the form of quantiles of the underlying distribution of model error. By assuming that

locations with similar environmental conditions have similar soil properties (Hudson 1992), the uncertainty of prediction at each location can also be determined inversely related to its environmental similarity to the soil sample set (Zhu *et al* 2015, 2018).

4. Challenges

Great progress has been achieved in the digital mapping of soil carbon over the last decades. However, mapping soil carbon accurately still poses additional challenges on soil observation sampling, effective anthropogenic covariates generation, predictive models development and so on. In addition, attentions have been increasingly paid on the temporal dimension of soil carbon mapping due to obvious climate change, soil health concern and land system modeling requirement. Pedometricians face an increasing demand in modeling the spatio-temporal soil carbon as well as revealing the underlying soil processes.

4.1. Challenges in collecting soil carbon observations in depths

Sampling design based on environmental covariates feature space has been developed and applied widely. For example, Minasny and McBratney (2006) developed conditional Latin hypercube sampling (cLHS) for soil mapping, which selected sample points by taking a full coverage of each covariate. These methods seemed to obtain more accurate soil maps than the probability sampling methods (Wadoux *et al* 2020). However, when considering a temporal variation of soil carbon, novel and efficient spatial-temporal sampling methods are needed (Wadoux *et al* 2021).

Another important aspect of soil sampling is to consider the characteristics of deeper soil layers. Until recently, studies of soil carbon mapping have paid more attention to surface layers (0–30 cm), while less to deeper soil layers (30–100 cm) (figure 3(b)). Soil processes vary at different depths due to the different influential covariates. For example, soil properties at surface soil (0–30 cm) tend to be influenced by management variabilities such as fertilizer and tillage, while deep soil (30–90 cm) tends to reflect the natural patterns of soil variability (Knadel *et al* 2011). In consequence, sampling design for several depths needs to synthesize different environmental covariates. Simply put, we need to develop new strategies on how to obtain sampling points more efficiently in cropland.

4.2. Challenges in extracting crop information

Various VI are widely used in the studies of agricultural areas. However, most studies did not analyze the VI in combination with the characteristics of soil carbon in cropland, nor did they discuss the spatio-temporal characteristics of VI to a large spatial extent. Particularly noteworthy is the fact that

some areas carry out household-based and scattered farming systems with small fields, for example, in China and Africa, which makes VI in those areas not very ‘informative’ and lack sufficiently fine spatial resolution (Dong *et al* 2019, Weiss *et al* 2020). At present, researches on the extraction of crop rotation or planting area based on remote sensing are well developed (Begue *et al* 2018, Bendini *et al* 2019), while the application of such information is suboptimal. Another potential limitation is that we may not have included VI capable of depicting the pedogenesis and/or temporal fluxes of soil carbon over short-term (seasonal) and long-term (climatic) trends, as well as management impacts.

4.3. Challenges in obtaining agricultural management information

Agricultural activities are increasing as major contributors to the dynamic of soil carbon in the cropland (Foley *et al* 2011). More studies begin to use agricultural management information as environmental covariates in mapping soil carbon rather than just as the background of the study area. Currently, the acquisition of agricultural information is mostly based on some indirect relationships (table 1), such as the distance from the irrigation system or the main rivers in the area (Dong *et al* 2019), or data spatialized from the statistical yearbook (Deng *et al* 2018). The coarse spatial scale of agricultural management could possibly hinder the accuracy of prediction (Martin *et al* 2011). Undoubtedly, agricultural management information such as the amount of fertilization and the amount of straw returning to the field is important but difficult to quantify for each pixel. It is critical to be aware that the acquisition of such quantitative information still requires a large number of field observations (Harms *et al* 2015). In summary, two things conspired towards the present dilemma: (a) the high cost of the fixed-point in large-scale cropland; (b) the difficulty to obtain accurate agricultural management information under the empirical planting pattern.

Therefore, major investment in new measurement will be required in countries with scant data from credible, georeferenced field experiments. The agricultural information would need to keep pace with the effect of spatial differences on soil properties.

4.4. Challenges in the development of predictive models

Research for predicting soil carbon has witnessed a considerable shift from using statistical models to ML models, because ML methods can handle well the non-linear relationship between soil carbon and environmental covariates. While, ML methods have great capacity in predicting rather than explaining. Typically, it is unknown how the covariates relate to one another or whether the input interactions exist (Wadoux *et al* 2020). This has resulted in criticism that the ML methods are ‘black box’ prediction

Table 1. The summary of anthropogenic covariates used in digital mapping of soil carbon in cropland (NDVI, normalized difference vegetation index; GNDVI, green normalized difference vegetation index; EVI, enhanced vegetation index; RVI, ratio vegetation index; GRVI, green-red vegetation index; SAVI, soil adjusted total vegetation index; SATVI, soil adjusted total vegetation index; MSAVI2: second modified soil adjusted vegetation index).

Agricultural information	Variables or indicators		Characteristic	Data source	Articles
Crop information	Cropping patterns	Crop rotation types		Field investigation; HJ-1A/1B CCD	Ellili et al (2019), Wu et al (2019), Yang et al (2019, 2020), Hu et al (2021), Wu et al (2021)
	Cropping systems	Crop types	Agricultural cover types; previous crop types	NDVI time series, SPOT-VGT, and more	Flathers and Gessler (2018), Deng et al (2020)
		Crop yield	Average crop yield	Field investigation; USDA-NASS	Kravchenko and Robertson (2007)
	Vegetation indices	NDVI	$\frac{NIR-Red}{NIR+Red}$	MODIS, HJ-1A/B CCD, Landsat 5 TM/8, Sentinel 2A, and more	Mora-Vallejo et al (2008), Bou Kheir et al (2010), Piccini et al (2014), Mosleh et al (2016), Khanal et al (2018), Yang et al (2020)
		GNDVI	$\frac{NIR-Green}{NIR+Green}$	Sentinel 2A, Landsat 5 TM/7 ETM+/8	Guo et al (2021b), He et al (2021), Zepp et al (2021)
		EVI	$\frac{\rho_{NIR}-\rho_{Red}}{\rho_{NIR}+CI \times \rho_{Red}-C2 \times \rho_{Blue}+L}$	Developed to improve NDVI by optimizing the vegetation signal in areas with a high leaf area index (LAI)	Takata et al (2007), Li (2010), Dharumajan et al (2017), Chen et al (2019), Hamzehpour et al (2019), Zhou et al (2020), He et al (2021), Zepp et al (2021)
		RVI	$\frac{NIR}{Red}$	MODIS, Landsat 5 TM/7 ETM+, Sentinel 2A, and more	Taghizadeh-Mehrdadi et al (2016), Pouladi et al (2019), Nguyen et al (2022)
		GRVI	$\frac{Green-Red}{Green+Red}$	Used for estimating and monitoring green biomass	Mponela et al (2020), He et al (2021), Zepp et al (2021)
			Sensitive to land cover types	Sentinel 2	

(Continued.)

Table 1. (Continued.)

Agricultural information	Variables or indicators	Characteristic	Data source	Articles
SAVI	$\frac{(\rho_{\text{NIR}} - \rho_{\text{Red}})(1 + N)}{\rho_{\text{NIR}} - \rho_{\text{Red}} + N}$	Designed to reduce the impact of soil brightness.	Sentinel 2, Landsat 5 ETM/7 ETM+/8, RapidEye Sensor	Taghizadeh-Mehrdadi <i>et al</i> (2016), Costa <i>et al</i> (2018), Nabiollahi <i>et al</i> (2018), Flynn <i>et al</i> (2019), Paul <i>et al</i> (2020b), Nguyen <i>et al</i> (2022), Zepp <i>et al</i> (2021)
SATVI	$\frac{\rho_{\text{SWIR1}} - \rho_{\text{Red}}}{\rho_{\text{SWIR1}} + \rho_{\text{Red}} + 1} \times 2 - \frac{\rho_{\text{SWIR2}}}{2}$	Sensitive to both green and senescent vegetation	Sentinel 2A, Landsat 5 ETM/7 ETM+	Zhou <i>et al</i> (2020), Wang <i>et al</i> (2021a), Zepp <i>et al</i> (2021), Zeraatpisheh <i>et al</i> (2021)
MSAVI2	$\frac{2 \times \rho_{\text{NIR}} + 1 - \sqrt{(2 \times \rho_{\text{NIR}} + 1)^2 - 8 \times (\rho_{\text{NIR}} - \rho_{\text{Red}})}}{2}$	Designed to be used in the early stages of crop growth to observe the first seedlings emerging; minimize the influence of bare soil	Sentinel 2, Landsat 5 ETM/7 ETM+	Mponela <i>et al</i> (2020), He <i>et al</i> (2021), Zepp <i>et al</i> (2021)
TVI	$\left(\frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} + 0.5 \right)^{0.5}$	Sensitive to living plants	Sentinel 2, Landsat 5 ETM/7 ETM+	He <i>et al</i> (2021), Wang <i>et al</i> (2021a), Zepp <i>et al</i> (2021)
Agricultural management	Averaged nitrogen-fertilizer application rate Irrigation command area; drainage class; irrigation index Lengths of cultivation Soil plow depth	Average fertilization level Irrigation status	Statistical yearbook Designed irrigation atlas or soil map Historical data Field investigation	Deng <i>et al</i> (2018, 2020) Sreenivas <i>et al</i> (2016), Sindayihubura <i>et al</i> (2017), Dong <i>et al</i> (2019) Wang <i>et al</i> (2019) Hu <i>et al</i> (2021)

approaches without an explanatory basis (Benke *et al* 2020). The interpretation of ML methods to reveal the complex system functioning and expose the importance of each driver of soil carbon variation is helpful but relies on the input training sample data. Furthermore, incorporating soil-forming processes in the digital mapping of soil carbon with statistical models or ML is a challenging exercise (Wadoux *et al* 2020, 2021).

There are both static variables (such as topographic variables) and dynamic/temporal variables (such as annual mean temperature, and crop growth variables) in soil mapping. Furthermore, the structure of temporal covariates may contain spatial-temporal characteristics. This requires novel models able to process the spatio-temporal information of different data. However, the current methods commonly used in digital mapping of soil carbon, such as MLR, RF, and RK, only employ the spatial characteristics of the environmental covariates but are not capable to deal with the spatio-temporal characteristics of these data (Kumar 2015). In the future, when using time-series remote sensing data or long-term agricultural information in mapping soil carbon, it is very urgent to explore methods that can deal with both spatial and temporal features in using natural and anthropogenic variables.

Soil carbon dynamics have recently gotten increasing scientific interest (Li *et al* 2016, Mishra *et al* 2019, Xie *et al* 2021). However, most of the current studies only focus on static soil carbon mapping. Therefore, we need to characterize the variation of soil carbon in space and time to improve the mechanism and drivers of soil carbon dynamics.

5. Future horizons

Based on the review, we identify some challenges in the current use of agricultural information and models for digital mapping of soil carbon, then we want to outline several promising solutions to our interwoven request and application challenges (figure 7). Moving forward, we should not be locked into a single approach *a priori*, whether it be efficient sampling strategies, model modification, or suitable agricultural information. A holistic conceptual framework is proposed for better monitoring spatial and temporal variation of soil carbon to support food production, soil protection, carbon sequestration, and emissions mitigation (figure 7).

5.1. Efficient sampling strategies for obtaining representative samples

Encouragingly, pedometricians have paid attention to efficient sampling with lower cost but higher mapping accuracy (Wadoux *et al* 2021). For example, Ma *et al* (2020) proposed that the *in-situ* field sampling based on environmental similarity in case of inaccessibility of pre-designed sample points was more

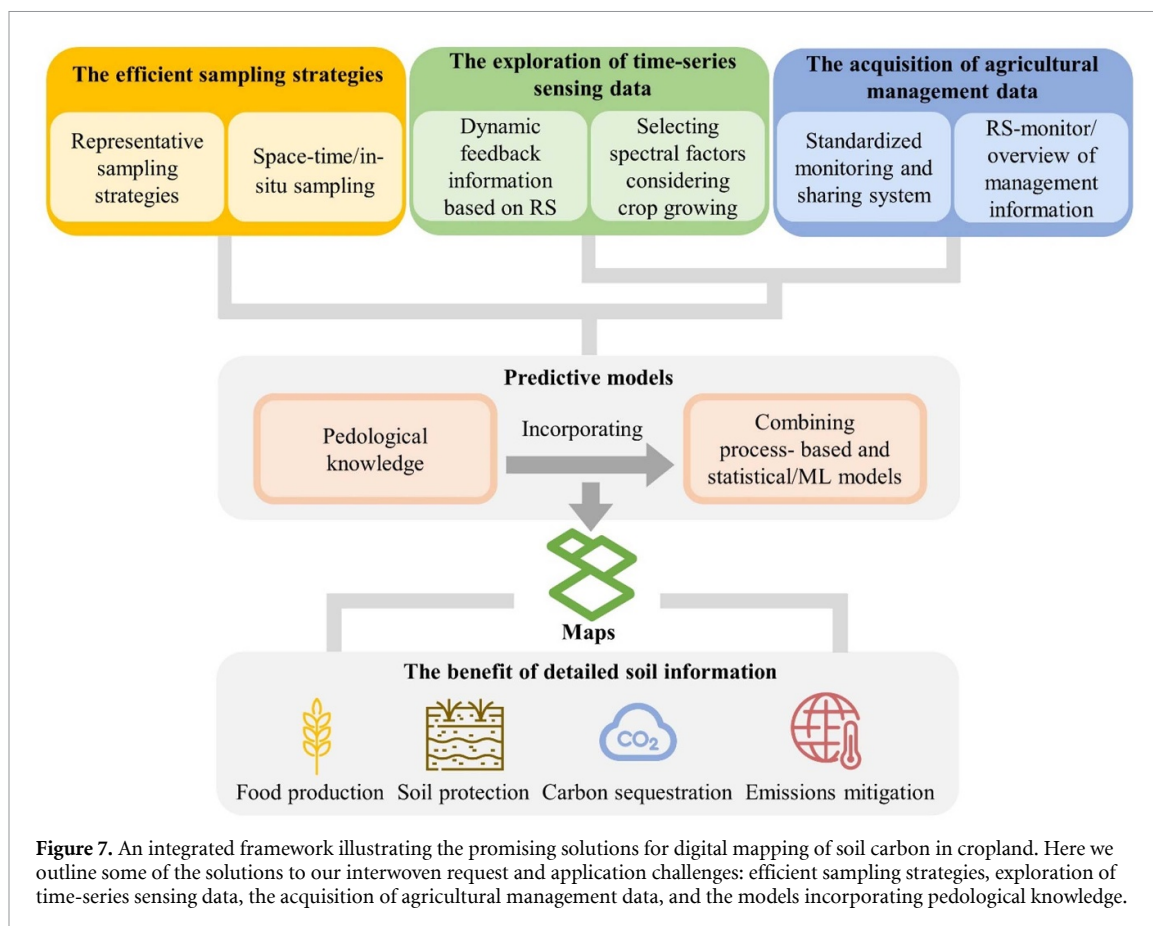
accurate than cLHS in soil sand mapping. Wadoux *et al* (2019) also pointed out that feature space coverage sampling was more efficient than cLHS. Feature space coverage sampling is accomplished by minimizing a feature space distance requirement between sampling and prediction locations by the k-means method, aiming for an even sampling density in the multivariate feature space (Wadoux *et al* 2019). Yang *et al* (2013) and Zhang *et al* (2022) proposed representative sampling strategies to design typical samples based on clustering of environmental covariates. The representative sampling strategy proved determining sampling points representing typical soil variations is a potential efficient sampling strategy. For sampling in cropland, the approach would deserve more attention in considering specific agricultural management of cropland in the future.

When considering the spatio-temporal SOC content changes, efficient space-time samplings are needed to apply. Brus (2014) stated the idea of space-time sampling, which selected both sampling locations and times by sampling probability. Combining the characteristics of long-term agricultural activities to designing a reasonable space-time sample point distribution would tap into the effectiveness of agricultural information, improve the accuracy of dynamic SOC content mapping, and maximize the value of sampling data.

5.2. Exploring time-series sensing data indicating crop growth characteristics

Remote sensing data with good spatio-temporal resolutions provide opportunities for the acquisition of time-series data indicating land surface conditions, such as soil water conditions, soil mechanical composition, and crop characteristics (Zhao *et al* 2014, Fatholouloumi *et al* 2020, Zeng *et al* 2020a). Advanced RS platforms make crop monitor cheaper, faster and more accurate (Weiss *et al* 2020). Time-series data could capture greater variability and thus indirectly indicate agricultural management styles and levels. Therefore, scholars should pay more attention to the application of time-series environmental covariates in the digital mapping of soil carbon and continue to explore various types of dynamic feedback information based on remote sensing. Recent studies have demonstrated the effectiveness of a new covariate known as land surface dynamic feedback (LSDF) which captures temporal spectral of the land surface following a rainfall event (Liu *et al* 2012, Guo *et al* 2016, Zeng *et al* 2019).

The selection of spectral factors in agricultural areas should be combined with the characteristics of the crop growing season. Fatholouloumi *et al* (2020) indicated that the satellite-based factors (albedo, emissivity, land surface temperature, incidence) extracted in June and July were more important than those in August and September due to the less influence of vegetation. And the difference was



more definite under dense cultivation. Remote sensing images tend to reflect the spectral reflectance of the vegetation growth condition in the summer or autumn, while the proximal geophysical covariates can reflect the spectral reflectance of the soil condition in the spring or winter (Guo *et al* 2021b, Wang *et al* 2021a). Thus, we can consider combining the soil spectral bands and VI which can integrate the advantages of remote sensing images on different dates.

5.3. Acquiring agricultural management information

The important sources of agricultural information include statistical yearbooks, farmer surveys data, agricultural production research, management data under precision agriculture, and some fixed-point data serving agricultural policies. There are common methods to get an overview of management information, such as interpolation of statistical data at a county/village level. However, to obtain detailed information on agricultural management, we must invest in explicit consideration when establishing standardized data collecting and sharing system in the future. Government departments and research institutions should also join forces to overcome some data protection challenges when mapping large-scale soil mapping, by giving incentives to local farmers and stakeholders who are often well informed about agricultural management information (Amelung *et al*

2020). For example, an earthworm population survey across 1300 ha in the United Kingdom was conducted by farmers on their land which was devised and lead by Rothamsted Research (Rumpel *et al* 2018). Della Chiesa *et al* (2019) proposed a viable cooperative structure for obtaining soil and agricultural data that connects individual farmers with a variety of stakeholders. The management of agricultural information could be achieved within a GIS framework. Such open platforms provide tools for sharing spatial and temporal agricultural information dynamics.

Field research results and remote sensing can be combined to confirm each other so that crop management information can be better monitored and characterized to a certain extent. Some remote sensing variables can monitor crop status throughout the season, such as crop yield and crop nitrogen content (Weiss *et al* 2020). For instance, the survey and statistical data, before applying, should be tested across a broad range of dynamics as simulated by remote sensing monitoring.

5.4. Incorporating pedological knowledge into statistical/ML models

In addition to increasing the accuracy of prediction models to improve soil carbon content/stock predictions, we should increase our understanding of the mechanism of forming and decomposition of soil carbon under the impact of environmental covariates.

Incorporating pedological knowledge into the statistical or ML models enhances our ability to understand variations among soil carbon and to precisely and accurately estimate and predict changes in soil carbon according to its accumulation and decomposition mechanism. The most basic incorporation is using pedological knowledge in the selection of environmental covariates. For example, Hendriks *et al* (2021) introduced a mechanistic model to predict SOM stocks, which started with identifying major processes that influenced SOM stocks, then collecting the input data for the model. The other alternative method is incorporating pedological knowledge into the conceptual model, such as structural equation modeling (Angelini *et al* 2016).

Statistical models or ML models combined with process-based models, e.g. century, Rothamsted Carbon model (RothC), and DeNitrification and DeComposition, may be a potential solution to model soil carbon in space and time. For example, Xie *et al* (2021) proposed the idea of incorporating the process-based RothC model into the geographically weighted regression kriging to better the space-time modeling of SOC contents in the heavily human-impacted area and the attempt resulted in improved model performance. Process-based models can benefit from simulating point results, while ML models show opportunities to extrapolate relationships by using a large number of covariates, so a method that combines the advantages of the two can be designed for optimal direction.

In the past few years, advanced data mining methods, such as deep learning, have been explored to improve spatio-temporal soil carbon modeling (Wadoux 2019, Yang *et al* 2021). A potential advantage of deep learning is its ability to deal with multiple source data (both spatial and spatio-temporal data) (Wang *et al* 2022). Due to the 'black-box' characteristics of deep learning or ML methods, the incorporation of pedological knowledge would improve the model performance in mapping soil carbon distribution conforming to the law of soil genesis.

6. Conclusion

Soil science is at the interface of several questions relevant to the global agenda of ecosystems, climate change, and agriculture. However, due to the rapid development of contemporary facility agriculture, severe anthropogenic disturbance creates a high level of uncertainty in accurate soil mapping, particularly in cropland. A holistic understanding of agricultural activity on soil carbon spatial distribution patterns and its changing processes is critical. Focusing on cropland, our review of the literature suggests some important points that are likely to function better in the digital mapping of soil carbon and in tackling the common goal of carbon emission-agriculture development on a scale and at a level that

is appropriate for the complex challenge of carbon neutrality.

- (a) The growing publications on digital mapping of soil carbon in cropland state that the attention to cropland's soil carbon is increasing. Studies are clustered in several countries, namely China, the United States, Iran, and Australia.
- (b) Most studies focused on mapping topsoil SOC/SOM content, and fewer studies on SOC stocks. Further efforts are necessary to gain insights into deeper soil which is essential for a better understanding of soil function.
- (c) In many cases, climate covariates, parent material, and soil type play an important role in soil carbon on the regional scale, while on a local scale, the variability of soil carbon often depends on terrain, agricultural management styles, and soil properties. Many researchers have realized the significance of agricultural information as environmental covariates for the digital mapping of soil carbon. Agricultural management information is usually obtained through survey data, remote sensing technology, or expert knowledge. The acquisition of detailed spatial agricultural management information should be improved by establishing standardized monitoring and sharing system.
- (d) There has been a large number of studies using environmental covariates based on proximal/remote sensing. Although the steps towards the time-series sensing data are diverse, imperfect, incremental, and take time, they provide a promising opportunity to capture the spatial variability of soil carbon.
- (e) Among the predictive models, there were a significant number of investigations concerning some promising algorithms, such as PLSR, RK, RF, SVM, and ANN. No single model was found to be the optimum method in all study areas, although ML and hybrid models are increasingly applied. In addition, to develop more advanced prediction methods such as deep learning, more attention should be given to incorporating pedological knowledge in the modeling process, where major improvements in combining process-based and statistical/ML models may reveal new insights into the interpretability of predictive results.

Overall, DSM is a promising tool that helps to assess and monitor soil carbon conditions with enough accuracy for cropland. An integrated framework including efficient sampling strategies, time-series sensing data, agricultural management data, and pedological knowledge-based models will work better in cropland and at different scales. There is a need to design more sophisticated technologies (e.g. sufficient samples/covariates sharing platforms

and dynamic model structures) for modeling soil carbon and processes at fine resolution and with high accuracy.

Data availability statement

No new data were created or analyzed in this study.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (Grant No. 41971054), the Leading Funds for the First Class Universities (020914912203 and 020914902302), and the Research Funds for the Frontiers Science Center for Critical Earth Material Cycling, Nanjing University.

Conflict of interest

All the authors have read and approved the manuscript. There are no conflicts of interest to declare.

ORCID iDs

Lin Yang  <https://orcid.org/0000-0002-7576-4481>

Lei Zhang  <https://orcid.org/0000-0002-1090-6338>

References

- Abdalla K, Chivenge P, Ciais P and Chaplot V 2016 No-tillage lessens soil CO₂ emissions the most under arid and sandy soil conditions: results from a meta-analysis *Biogeosciences* **13** 3619–33
- Adhikari K and Hartemink A E 2017 Soil organic carbon increases under intensive agriculture in the Central Sands, Wisconsin, USA *Geoderma Reg.* **10** 115–25
- Aldana-Jague E, Heckrath G, Macdonald A, van Wesemael B and Van Oost K 2016 UAS-based soil carbon mapping using VIS-NIR (480–1000 nm) multi-spectral imaging: potential and limitations *Geoderma* **275** 55–66
- Amelung W *et al* 2020 Towards a global-scale soil climate mitigation strategy *Nat. Commun.* **11** 5427
- Anderson R, Bayer P E and Edwards D 2020 Climate change and the need for agricultural adaptation *Curr. Opin. Plant Biol.* **56** 197–202
- Angelini M E, Heuvelink G B M, Kempen B and Morras H J M 2016 Mapping the soils of an Argentine Pampas region using structural equation modelling *Geoderma* **281** 102–18
- Angelopoulou T, Tziolas N, Balafoutis A, Zalidis G and Bochtis D 2019 Remote sensing techniques for soil organic carbon estimation: a review *Remote Sens.* **11** 676
- Araya S, Lyle G, Lewis M and Ostendorf B 2016 Phenologic metrics derived from MODIS NDVI as indicators for plant available water-holding capacity *Ecol. Indic.* **60** 1263–72
- Arrouays D, Poggio L, Guerrero O A S and Mulder V L 2020 Digital soil mapping and GlobalSoilMap main advances and ways forward *Geoderma Reg.* **21** e00265
- Bai X X *et al* 2019 Responses of soil carbon sequestration to climate-smart agriculture practices: a meta-analysis *Glob. Change Biol.* **25** 2591–606
- Bartholomeus H, Kooistra L, Stevens A, van Leeuwen M, van Wesemael B, Ben-Dor E and Tychon B 2011 Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy *Int. J. Appl. Earth Obs. Geoinf.* **13** 81–88
- Batjes N H, Ribeiro E and van Oostrum A 2020 Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019) *Earth Syst. Sci. Data* **12** 299–320
- Begue A, Arvor D, Bellon B, Betbeder J, de Abelleira R, Ferraz R P D, Lebourgeois V, Lelong C, Simões M and R. Verón S 2018 Remote sensing and cropping practices: a review *Remote Sens.* **10** 99
- Beguín J, Fuglstad G-A, Mansuy N and Paré D 2017 Predicting soil properties in the Canadian boreal forest with limited data: comparison of spatial and non-spatial statistical approaches *Geoderma* **306** 195–205
- Behrens T, Schmidt K, Ramirez-Lopez L, Gallant J, Zhu A X and Scholten T 2014 Hyper-scale digital soil mapping and soil formation analysis *Geoderma* **213** 578–88
- Bell M J and Worrall F 2009 Estimating a region's soil organic carbon baseline: the undervalued role of land-management *Geoderma* **152** 74–84
- Bellinaso H, Dematte J A M and Romeiro S A 2010 Soil spectral library and its use in soil classification *Rev. Bras. Cienc. Solo* **34** 861–70
- Bendini H D *et al* 2019 Detailed agricultural land classification in the Brazilian cerrado based on phenological information from dense satellite image time series *Int. J. Appl. Earth Obs. Geoinf.* **82** 101872
- Benke K K, Norng S, Robinson N J, Chia K, Rees D B and Hopley J 2020 Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content *Geoderma* **366** 114210
- Bilgili A V, Akbas F and van Es H M 2011 Combined use of hyperspectral VNIR reflectance spectroscopy and kriging to predict soil variables spatially *Precis. Agric.* **12** 395–420
- Bolinder M A, Crotty F, Elsen A, Frac M, Kismányoky T, Lipiec J, Tits M, Tóth Z and Kätterer T 2020 The effect of crop residues, cover crops, manures and nitrogen fertilization on soil organic carbon changes in agroecosystems: a synthesis of reviews *Mitig. Adapt. Strateg. Glob. Change* **25** 929–52
- Bonfatti B R, Hartemink A E, Giasson E, Tornquist C G and Adhikari K 2016 Digital mapping of soil carbon in a viticultural region of Southern Brazil *Geoderma* **261** 204–21
- Bou Kheir R, Greve M H, Bocher P K, Greve M B, Larsen R and McCloy K 2010 Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: the case study of Denmark *J. Environ. Manage.* **91** 1150–60
- Brevik E C, Calzolari C, Miller B A, Pereira P, Kabala C, Baumgarten A and Jordán A 2016 Soil mapping, classification, and pedologic modeling: history and future directions *Geoderma* **264** 256–74
- Bricklemeyer R S and Brown D J 2010 On-the-go VisNIR: potential and limitations for mapping soil clay and organic carbon *Comput. Electron. Agric.* **70** 209–16
- Brown D J, Shepherd K D, Walsh M G, Dwayne Mays M and Reinsch T G 2006 Global soil characterization with VNIR diffuse reflectance spectroscopy *Geoderma* **132** 273–90
- Brown M E, de Beurs K M and Marshall M 2012 Global phenological response to climate change in crop areas using satellite remote sensing of vegetation, humidity and temperature over 26 years *Remote Sens. Environ.* **126** 174–83
- Brus D J 2014 Statistical sampling approaches for soil monitoring *Eur. J. Soil Sci.* **65** 779–91
- Brus D J 2019 Sampling for digital soil mapping: a tutorial supported by R scripts *Geoderma* **338** 464–80
- Brus D J, Kempen B and Heuvelink G B M 2011 Sampling for validation of digital soil maps *Eur. J. Soil Sci.* **62** 394–407
- Cambule A H, Rossiter D G and Stoorvogel J J 2013 A methodology for digital soil mapping in poorly-accessible areas *Geoderma* **192** 341–53
- Chen D, Chang N, Xiao J, Zhou Q and Wu W 2019 Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms *Sci. Total Environ.* **669** 844–55
- Chen F, Kissel D E, West L T and Adkins W 2000 Field-scale mapping of surface soil organic carbon using remotely sensed imagery *Soil Sci. Soc. Am. J.* **64** 746–53

- Clark M and Tilman D 2017 Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice *Environ. Res. Lett.* **12** 064016
- Costa E M, Tassinari W D S, Koenow Pinheiro H S, Beutler S J and Cunha Dos Anjos L H 2018 Mapping soil organic carbon and organic matter fractions by geographically weighted regression *J. Environ. Qual.* **47** 718–25
- Dai F, Zhou Q, Lv Z, Wang X and Liu G 2014 Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau *Ecol. Indic.* **45** 184–94
- Debaene G, Niedzwiecki J, Pecio A and Zurek A 2014 Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale *Geoderma* **214** 114–25
- Della Chiesa S, la Cecilia D, Genova G, Balotti A, Thalheimer M, Tappeiner U and Niedrist G 2019 Farmers as data sources: cooperative framework for mapping soil properties for permanent crops in South Tyrol (Northern Italy) *Geoderma* **342** 93–105
- Dendoncker N, Van Wesemael B, Rounsevell M D A, Roelandt C and Lettens S 2004 Belgium's CO₂ mitigation potential under improved cropland management *Agric. Ecosyst. Environ.* **103** 101–16
- Deng X, Chen X, Ma W, Ren Z, Zhang M, Grieneisen M L, Long W, Ni Z, Zhan Y and Lv X 2018 Baseline map of organic carbon stock in farmland topsoil in East China *Agric. Ecosyst. Environ.* **254** 213–23
- Deng X, Ma W, Ren Z, Zhang M, Grieneisen M L, Chen X, Fei X, Qin F, Zhan Y and Lv X 2020 Spatial and temporal trends of soil total nitrogen and C/N ratio for croplands of East China *Geoderma* **361** 114035
- Dharumarajan S, Hegde R and Singh S K 2017 Spatial prediction of major soil properties using Random Forest techniques—a case study in semi-arid tropics of South India *Geoderma Reg.* **10** 154–62
- Diek S, Fornallaz F and Schaepman M E 2017 Barest pixel composite for agricultural areas using Landsat time series *Remote Sens.* **9** 1245
- Dignac M F et al 2017 Increasing soil carbon storage: mechanisms, effects of agricultural practices and proxies. A review *Agron. Sustain. Dev.* **37** 14
- Dong W, Wu T, Luo J, Sun Y and Xia L 2019 Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China *Geoderma* **340** 234–48
- Ellili Y, Walter C, Michot D, Pichelin P and Lemercier B 2019 Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale *Geoderma* **351** 1–8
- Fan M M, Lal R, Zhang H, Margenot A J, Wu J, Wu P, Zhang L, Yao J, Chen F and Gao C 2020 Variability and determinants of soil organic matter under different land uses and soil types in eastern China *Soil Tillage Res.* **198** 104544
- Fatholouloumi S, Vaezi A R, Alavipanah S K, Ghorbani A, Saurette D and Biswas A 2020 Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: a case study in Iran *Sci. Total Environ.* **721** 137703
- Flathers E and Gessler P E 2018 Building an open science framework to model soil organic carbon *J. Environ. Qual.* **47** 726–34
- Flynn T, de Clercq W, Rozanov A and Clarke C 2019 High-resolution digital soil mapping of multiple soil properties: an alternative to the traditional field survey? *South Afr. J. Plant Soil* **36** 237–47
- Foley J A et al 2011 Solutions for a cultivated planet *Nature* **478** 337–42
- Food and Agriculture Organisation of the United Nations (FAO) 2019 Recarbonization of global soils: a dynamic response to offset global emissions (available at: www.fao.org/3/i7235en/i7235EN.pdf)
- Fox G A and Sabbagh G J 2002 Estimation of soil organic matter from red and near-infrared remotely sensed data using a soil line Euclidean distance technique *Soil Sci. Soc. Am. J.* **66** 1922–9
- Gattinger A et al 2012 Enhanced top soil carbon stocks under organic farming *Proc. Natl Acad. Sci. USA* **109** 18226–31
- Gregorutti B, Michel B and Saint-Pierre P 2017 Correlation and variable importance in random forests *Stat. Comput.* **27** 659–78
- Grimm R, Behrens T, Marker M and Elsenbeer H 2008 Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis *Geoderma* **146** 102–13
- Guevara M et al 2018 No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America *Soil* **4** 173–93
- Guo L et al 2021b Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas *Geoderma* **398** 115118
- Guo L, Fu P, Shi T, Chen Y, Zeng C, Zhang H and Wang S 2021a Exploring influence factors in mapping soil organic carbon on low-relief agricultural lands using time series of remote sensing data *Soil Tillage Res.* **210** 104982
- Guo L, Fu P, Shi T, Chen Y, Zhang H, Meng R and Wang S 2020 Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images *Soil Tillage Res.* **196** 104477
- Guo L, Zhang H, Shi T, Chen Y, Jiang Q and Linderman M 2019 Prediction of soil organic carbon stock by laboratory spectral data and airborne hyperspectral images *Geoderma* **337** 32–41
- Guo P-T, Li M-F, Luo W, Tang Q-F, Liu Z-W and Lin Z-M 2015 Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach *Geoderma* **237-238** 49–59
- Guo S X, Zhu A-X, Meng L, Burt J E, Du F, Liu J and Zhang G 2016 Unification of soil feedback patterns under different evaporation conditions to improve soil differentiation over flat area *Int. J. Appl. Earth Obs. Geoinf.* **49** 126–37
- Haddaway N R, Hedlund K, Jackson L E, Kätterer T, Lugato E, Thomsen I K, Jørgensen H B and Isberg P-E 2016 How does tillage intensity affect soil organic carbon? A systematic review protocol *Environ. Evid.* **5** 1
- Hamberg L, Malmivaara-Lämsä M, Lehvävirta S and Kotze D J 2008 The effects of soil fertility on the abundance of rowan (*Sorbus aucuparia* L.) in urban forests *Plant Ecol.* **204** 21–32
- Hamzehpour N, Shafizadeh-Moghadam H and Valavi R 2019 Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture *Catena* **182** 104141
- Han D, Wiesmeier M, Conant R T, Kühnel A, Sun Z, Kögel-Knabner I, Hou R, Cong P, Liang R and Ouyang Z 2018 Large soil organic carbon increase due to improved agronomic management in the North China Plain from 1980s to 2010s *Glob. Change Biol.* **24** 987–1000
- Harms B, Brough D, Philip S, Bartley R, Clifford D, Thomas M, Willis R and Gregory L 2015 Digital soil assessment for regional agricultural land evaluation *Glob. Food Sec.* **5** 25–36
- Hbirkou C, Pätzold S, Mahlein A-K and Welp G 2012 Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale *Geoderma* **175** 21–28
- He X L, Yang L, Li A, Zhang L, Shen F, Cai Y and Zhou C 2021 Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images *Catena* **205** 105442
- Hendriks C M J, Stoorvogel J J, Alvarez-Martinez J M, Claessens L, Perez-Silos I and Barquin J 2021 Introducing a mechanistic model in digital soil mapping to predict soil organic matter stocks in the Cantabrian region (Spain) *Eur. J. Soil Sci.* **72** 704–19
- Hengl T et al 2014 SoilGrids1km—global soil information based on automated mapping *PLoS One* **9** e105992
- Hengl T, Heuvelink G B M and Stein A 2004 A generic framework for spatial prediction of soil variables based on regression-kriging *Geoderma* **120** 75–93

- Hong Y *et al* 2020 Exploring the potential of airborne hyperspectral image for estimating topsoil organic carbon: effects of fractional-order derivative and optimal band combination algorithm *Geoderma* **365** 114228
- Houkpatin O K L, de Hipt F O, Bossa A Y, Welp G and Amelung W 2018 Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso) *Catena* **166** 298–309
- Hu B F *et al* 2021 Spatial variability and potential controls of soil organic matter in the Eastern Dongting Lake Plain in southern China *J. Soils Sediments* **21** 2791–804
- Huang B, Sun W, Zhao Y, Zhu J, Yang R, Zou Z, Ding F and Su J 2007a Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices *Geoderma* **139** 336–45
- Huang X, Senthilkumar S, Kravchenko A, Thelen K and Qi J 2007b Total carbon mapping in glacial till soils using near-infrared spectroscopy, Landsat imagery and topographical information *Geoderma* **141** 34–42
- Hudson B D 1992 The soil survey as paradigm-based science *Soil Sci. Soc. Am. J.* **56** 836–41
- Ingram J S I and Fernandes E C M 2001 Managing carbon sequestration in soils: concepts and terminology *Agric. Ecosyst. Environ.* **87** 111–7
- Jackson R B, Lajtha K, Crow S E, Hugelius G, Kramer M G and Pineiro G 2017 The ecology of soil carbon: pools, vulnerabilities, and biotic and abiotic controls *Annu. Rev. Ecol. Syst.* **48** 419–45
- Jenny H 1941 *Factors of Soil Formation: A System of Quantitative Pedology* (New York: McGraw-Hill)
- Kalambukattu J G, Kumar S and Arya Raj R 2018 Digital soil mapping in a Himalayan watershed using remote sensing and terrain parameters employing artificial neural network model *Environ. Earth Sci.* **77** 203
- Kelly L T, Dayman R, Nimmo D G, Clarke M F and Bennett A F 2013 Spatial and temporal drivers of small mammal distributions in a semi-arid environment: the role of rainfall, vegetation and life-history *Aust. Ecol.* **38** 786–97
- Kempen B, Brus D J and Stoorvogel J J 2011 Three-dimensional mapping of soil organic matter content using soil type-specific depth functions *Geoderma* **162** 107–23
- Khanal S, Fulton J, Klopfenstein A, Douridas N and Shearer S 2018 Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield *Comput. Electron. Agric.* **153** 213–25
- Knadel M, Thomsen A and Greve M H 2011 Multisensor on-the-go mapping of soil organic carbon content *Soil Sci. Soc. Am. J.* **75** 1799–806
- Kodaira M and Shibusawa S 2013 Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping *Geoderma* **199** 64–79
- Kravchenko A N and Robertson G P 2007 Can topographical and yield data substantially improve total soil carbon mapping by regression kriging? *Agron. J.* **99** 12–17
- Kuang B, Tekin Y and Mouazen A M 2015 Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content *Soil Tillage Res.* **146** 243–52
- Kumar S 2015 Estimating spatial distribution of soil organic carbon for the Midwestern United States using historical database *Chemosphere* **127** 49–57
- Kuzyakov Y, Friedel J K and Stahr K 2000 Review of mechanisms and quantification of priming effects *Soil Biol. Biochem.* **32** 1485–98
- Laamrani A, Berg A A, Voroney P, Feilhauer H, Blackburn L, March M, Dao P D, He Y and Martin R C 2019 Ensemble identification of spectral bands related to soil organic carbon levels over an agricultural field in Southern Ontario, Canada *Remote Sens.* **11** 1298
- Lacoste M, Minasny B, McBratney A, Michot D, Viaud V and Walter C 2014 High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape *Geoderma* **213** 296–311
- Ladoni M, Bahrami H A, Alavipanah S K and Norouzi A A 2010 Estimating soil organic carbon from soil reflectance: a review *Precis. Agric.* **11** 82–99
- Lal R 2011 Sequestering carbon in soils of agro-ecosystems *Food Policy* **36** S33–S39
- Lamichhane S, Kumar L and Wilson B 2019 Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: a review *Geoderma* **352** 395–413
- Lek S and Guegan J F 1999 Artificial neural networks as a tool in ecological modelling, an introduction *Ecol. Modelling* **120** 65–73
- Li S M, Li J, Li C, Huang S, Li X, Li S and Ma Y 2016 Testing the RothC and DNDC models against long-term dynamics of soil organic carbon stock observed at cropping field soils in North China *Soil Tillage Res.* **163** 290–7
- Li X, Shang B, Wang D, Wang Z, Wen X and Kang Y 2020a Mapping soil organic carbon and total nitrogen in croplands of the Corn Belt of Northeast China based on geographically weighted regression kriging model *Comput. Geosci.* **135** 104392
- Li Y 2010 Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? *Geoderma* **159** 63–75
- Li Y, Li Z, Chang S X, Cui S, Jagadamma S, Zhang Q and Cai Y 2020b Residue retention promotes soil carbon accumulation in minimum tillage systems: implications for conservation agriculture *Sci. Total Environ.* **740** 140147
- Li Y, Li Z, Cui S, Jagadamma S and Zhang Q 2019 Residue retention and minimum tillage improve physical environment of the soil in croplands: a global meta-analysis *Soil Tillage Res.* **194** 104292
- Liddicoat C, Maschmedt D, Clifford D, Searle R, Herrmann T, Macdonald L M and Baldock J 2015 Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone *Soil Res.* **53** 956–73
- Liu F, Geng X Y, Zhu A X, Fraser W and Waddell A 2012 Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS *Geoderma* **171** 44–52
- Lopez-Granados F, Jurado-Exposito M, Pena-Barragan J M and Garcia-Torres L 2005 Using geostatistical and remote sensing approaches for mapping soil properties *Eur. J. Agron.* **23** 279–89
- Lu P, Wang L, Niu Z, Li L and Zhang W 2013 Prediction of soil properties using laboratory VIS-NIR spectroscopy and Hyperion imagery *J. Geochem. Explor.* **132** 26–33
- Luo C, Zhang X, Wang Y, Men Z and Liu H 2022 Regional soil organic matter mapping models based on the optimal time window, feature selection algorithm and Google Earth Engine *Soil Tillage Res.* **219** 105325
- Luo Z *et al* 2019 Mapping future soil carbon change and its uncertainty in croplands using simple surrogates of a complex farming system model *Geoderma* **337** 311–21
- Ma T W, Wei T, Qin C Z, Zhu A X, Qi F, Liu J, Zhao F and Pan H 2020 *In-situ* recommendation of alternative soil samples during field sampling based on environmental similarity *Earth Sci. Inform.* **13** 39–53
- Malone B P, Minasny B, Odgers N P and McBratney A B 2014 Using model averaging to combine soil property rasters from legacy soil maps and from point data *Geoderma* **232** 34–44
- Mandal A *et al* 2020 Impact of agricultural management practices on soil carbon sequestration and its monitoring through simulation models and remote sensing techniques: a review *Crit. Rev. Environ. Sci. Technol.* **52** 1–49
- Martin M P, Wattenbach M, Smith P, Meersmans J, Jolivet C, Bouillon L and Arrouays D 2011 Spatial distribution of soil organic carbon stocks in France *Biogeosciences* **8** 1053–65
- Martin P D, Malley D F, Manning G and Fuller L 2002 Determination of soil organic carbon and nitrogen at the

- field level using near-infrared spectroscopy *Can. J. Soil Sci.* **82** 413–22
- McBratney A B, Mendonça Santos M L and Minasny B 2003 On digital soil mapping *Geoderma* **117** 3–52
- Meersmans J et al 2012 A high resolution map of French soil organic carbon *Agron. Sustain. Dev.* **32** 841–51
- Meersmans J, De Ridder F, Canters F, De Baets S and Van Molle M 2008 A multiple regression approach to assess the spatial distribution of soil organic carbon (SOC) at the regional scale (Flanders, Belgium) *Geoderma* **143** 1–13
- Messori G, Ruiz-Perez G, Manzoni S and Vico G 2019 Climate drivers of the terrestrial carbon cycle variability in Europe *Environ. Res. Lett.* **14** 063001
- Miklos M, Short M G, McBratney A B and Minasny B 2010 Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, north-west New South Wales *Aust. J. Soil Res.* **48** 248–57
- Miller B A and Schaetzl R J 2014 The historical role of base maps in soil geography *Geoderma* **230–231** 329–39
- Minasny B et al 2017 Soil carbon 4 per mille *Geoderma* **292** 59–86
- Minasny B and McBratney A B 2002 Uncertainty analysis for pedotransfer functions *Eur. J. Soil Sci.* **53** 417–29
- Minasny B and McBratney A B 2006 A conditioned Latin hypercube method for sampling in the presence of ancillary information *Comput. Geosci.* **32** 1378–88
- Minasny B and McBratney A B 2008 Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy *Chemometr. Intell. Lab. Syst.* **94** 72–79
- Minasny B and McBratney A B 2016 Digital soil mapping: a brief history and some lessons *Geoderma* **264** 301–11
- Minasny B, McBratney A B, Malone B P and Ichani W 2013 Digital mapping of soil carbon *Adv. Agron.* **113** 1–47
- Minasny B, McBratney A B, Pichon L, Sun W and Short M G 2009 Evaluating near infrared spectroscopy for field prediction of soil properties *Aust. J. Soil Res.* **47** 664–73
- Mirzaee S, Ghorbani-Dashtaki S, Mohammadi J, Asadi H and Asadzadeh F 2016 Spatial variability of soil organic matter using remote sensing data *Catena* **145** 118–27
- Mishra G, Jangir A and Francaviglia R 2019 Modeling soil organic carbon dynamics under shifting cultivation and forests using RothC model *Ecol. Modelling* **396** 33–41
- Mora-Vallejo A, Claessens L, Stoorvogel J and Heuvelink G B M 2008 Small scale digital soil mapping in Southeastern Kenya *Catena* **76** 44–53
- Mosleh Z, Salehi M H, Jafari A, Borujeni I E and Mehnatkesh A 2016 The effectiveness of digital soil mapping to predict soil properties over low-relief areas *Environ. Monit. Assess.* **188** 195
- Mponela P, Snapp S, Villamor G B, Tamene L, Le Q B and Borgemeister C 2020 Digital soil mapping of nitrogen, phosphorus, potassium, organic carbon and their crop response thresholds in smallholder managed escarpments of Malawi *Appl. Geogr.* **124** 102299
- Nabiollahi K, Golmohamadi F, Taghizadeh-Mehrjardi R, Kerry R and Davari M 2018 Assessing the effects of slope gradient and land use change on soil quality degradation through digital mapping of soil quality indices and soil loss rate *Geoderma* **318** 16–28
- Naspendra Z, Aprisal A, Hijri N, Harianti M and Junaidi 2021 Digital mapping and soil carbon stock distribution on various landuse of tropical peatland in Pesisir Selatan, West Sumatra *IOP Conf. Ser.: Earth Environ. Sci.* **741** 012024
- Nawar S, Buddenbaum H and Hill J 2015 Digital mapping of soil properties using multivariate statistical analysis and ASTER data in an arid region *Remote Sens.* **7** 1181–205
- Nguyen T T, Pham T D, Nguyen C T, Delfos J, Archibald R, Dang K B, Hoang N B, Guo W and Ngo H H 2022 A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion *Sci. Total Environ.* **804** 150187
- Nikolaidis N P 2011 Human impacts on soils: tipping points and knowledge gaps *Appl. Geochem.* **26** S230–3
- Pahlavan-Rad M R, Dahmardeh K and Brungard C 2018 Predicting soil organic carbon concentrations in a low relief landscape, eastern Iran *Geoderma Reg.* **15** e00195
- Paul S S, Coops N C, Johnson M S, Krzic M, Chandna A and Smukler S M 2020a Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation *Geoderma* **363** 114177
- Paul S S, Dowell L, Coops N C, Johnson M S, Krzic M, Geesing D and Smukler S M 2020b Tracking changes in soil organic carbon across the heterogeneous agricultural landscape of the Lower Fraser Valley of British Columbia *Sci. Total Environ.* **732** 138994
- Paustian K, Lehmann J, Ogle S, Reay D, Robertson G P and Smith P 2016 Climate-smart soils *Nature* **532** 49–57
- Piao S L et al 2013 Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO₂ trends *Glob. Change Biol.* **19** 2117–32
- Piao S L, Fang J, Ciais P, Peylin P, Huang Y, Sitch S and Wang T 2009 The carbon balance of terrestrial ecosystems in China *Nature* **458** 1009–13
- Piccini C, Marchetti A and Francaviglia R 2014 Estimation of soil organic matter by geostatistical methods: use of auxiliary information in agricultural and environmental assessment *Ecol. Indic.* **36** 301–14
- Piikki K and Söderström M 2019 Digital soil mapping of arable land in Sweden—validation of performance at multiple scales *Geoderma* **352** 342–50
- Poggio L et al 2021 SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty *Soil* **7** 217–40
- Pouladi N, Moller A B, Tabatabai S and Greve M H 2019 Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging *Geoderma* **342** 85–92
- Priori S, Fantappie M, Bianconi N, Ferrigno G, Pellegrini S and Costantini E A C 2016 Field-scale mapping of soil carbon stock with limited sampling by coupling gamma-ray and Vis-NIR spectroscopy *Soil Sci. Soc. Am. J.* **80** 954–64
- Rizzo R, Demattè J A M, Lepsch I E, Gallo B C and Fongaro C T 2016 Digital soil mapping at local scale using a multi-depth Vis-NIR spectral library and terrain attributes *Geoderma* **274** 18–27
- Roberts D, Wilford J and Ghattas O 2019 Exposed soil and mineral map of the Australian continent revealing the land at its barest *Nat. Commun.* **10** 5297
- Roudier P, Malone B P, Hedley C B, Minasny B and McBratney A B 2017 Comparison of regression methods for spatial downscaling of soil organic carbon stocks maps *Comput. Electron. Agric.* **142** 91–100
- Rumpel C, Amiraslani F, Koutika L S, Smith P, Whitehead D and Wollenberg E 2018 Put more carbon in soils to meet Paris climate pledges *Nature* **564** 32
- Safanelli J L et al 2021 Leveraging the application of Earth observation data for mapping cropland soils in Brazil *Geoderma* **396** 115042
- Sanchez P A et al 2009 Digital soil map of the world *Science* **325** 680–1
- Schillaci C, Acutis M, Lombardo L, Lipani A, Fantappiè M, Märker M and Saia S 2017 Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling *Sci. Total Environ.* **601** 821–32
- Schlesinger W H and Amundson R 2019 Managing for soil carbon sequestration: let's get realistic *Glob. Change Biol.* **25** 386–9
- Shi J, Yang L, Zhu A-X, Qin C, Liang P, Zeng C and Pei T 2018 Machine-learning variables at different scales vs. knowledge-based variables for mapping multiple soil properties *Soil Sci. Soc. Am. J.* **82** 645–56
- Shrestha D L and Solomatine D P 2006 Machine learning approaches for estimation of prediction interval for the model output *Neural Netw.* **19** 225–35

- Simbahan G C, Dobermann A, Goovaerts P, Ping J and Haddix M L 2006 Fine-resolution mapping of soil organic carbon based on multivariate secondary data *Geoderma* **132** 471–89
- Sindayihebura A, Ottoy S, Dondeyne S, Van Meirvenne M and Van Orshoven J 2017 Comparing digital soil mapping techniques for organic carbon and clay content: case study in Burundi's central plateaus *Catena* **156** 161–75
- Song M, Yang L, Zhu A-X and Qin C-Z 2017 Mapping soil organic matter in farming areas with crop rotation (in Chinese) *Chin. J. Soil Sci.* **48** 778–85
- Sreenivas K et al 2016 Digital mapping of soil organic and inorganic carbon status in India *Geoderma* **269** 160–73
- Stevens A, Miralles I and van Wesemael B 2012 Soil organic carbon predictions by airborne imaging spectroscopy: comparing cross-validation and validation *Soil Sci. Soc. Am. J.* **76** 2174–83
- Stevens A, Udelhoven T, Denis A, Tychon B, Lioy R, Hoffmann L and van Wesemael B 2010 Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy *Geoderma* **158** 32–45
- Stevens F, Bogaert P, Van Oost K, Doetterl S and Van Wesemael B 2014 Regional-scale characterization of the geomorphic control of the spatial distribution of soil organic carbon in cropland *Eur. J. Soil Sci.* **65** 539–52
- Stevens F, Bogaert P and van Wesemael B 2015 Detecting and quantifying field-related spatial variation of soil organic carbon using mixed-effect models and airborne imagery *Geoderma* **259** 93–103
- Suleymanov A, Abakumov E, Suleymanov R, Gabbasova I and Komissarov M 2021 The soil nutrient digital mapping for precision agriculture cases in the Trans-Ural steppe zone of Russia using topographic attributes *ISPRS Int. J. Geo-Inf.* **10** 243
- Summers D, Lewis M, Ostendorf B and Chittleborough D 2011 Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties *Ecol. Indic.* **11** 123–31
- Sun W J, Canadell J G, Yu L, Yu L, Zhang W, Smith P, Fischer T and Huang Y 2020 Climate drives global soil carbon sequestration and crop yield changes under conservation agriculture *Glob. Change Biol.* **26** 3325–35
- Taghizadeh-Mehrjardi R, Nabiollahi K and Kerry R 2016 Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran *Geoderma* **266** 98–110
- Takata Y, Funakawa S, Akshalov K, Ishida N and Kosaki T 2007 Spatial prediction of soil organic matter in northern Kazakhstan based on topographic and vegetation information *Soil Sci. Plant Nutr.* **53** 289–99
- Tautges N E, Chiartas J L, Gaudin A C M, O'Geen A T, Herrera I and Scow K M 2019 Deep soil inventories reveal that impacts of cover crops and compost on soil carbon sequestration differ in surface and subsurface soils *Glob. Change Biol.* **25** 3753–66
- Tayebi M et al 2021 Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series *Remote Sens.* **13** 2223
- Trost B, Prochnow A, Drastig K, Meyer-Aurich A, Ellmer F and Baumecker M 2013 Irrigation, soil organic carbon and N₂O emissions. A review *Agron. Sustain. Dev.* **33** 733–49
- Tu C L, He T B, Lu X H, Luo Y and Smith P 2018 Extent to which pH and topographic factors control soil organic carbon level in dry farming cropland soils of the mountainous region of Southwest China *Catena* **163** 204–9
- van Meij H et al 2018 Comparing impacts of climate change and mitigation on global agriculture by 2050 *Environ. Res. Lett.* **13** 064021
- Varney R M, Chadburn S E, Friedlingstein P, Burke E J, Koven C D, Hugelius G and Cox P M 2020 A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming *Nat. Commun.* **11** 5544
- Vaudour E, Bel L, Gilliot J M, Coquet Y, Hadjar D, Cambier P, Michelin J and Houot S 2013 Potential of SPOT multispectral satellite images for mapping topsoil organic carbon content over peri-urban croplands *Soil Sci. Soc. Am. J.* **77** 2122–39
- Vaudour E, Gilliot J M, Bel L, Lefevre J and Chehdi K 2016 Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra *Int. J. Appl. Earth Obs. Geoinf.* **49** 24–38
- Vaudour E, Gomez C, Lagacherie P, Loiseau T, Baghdadi N, Urbina-Salazar D, Loubet B and Arrouays D 2021 Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands *Int. J. Appl. Earth Obs. Geoinf.* **96** 102277
- Viscarra Rossel R A et al 2016 A global spectral library to characterize the world's soil *Earth-Sci. Rev.* **155** 198–230
- Wadoux A M J C 2019 Using deep learning for multivariate mapping of soil with quantified uncertainty *Geoderma* **351** 59–70
- Wadoux A M J C et al 2021 Ten challenges for the future of pedometrics *Geoderma* **401** 115155
- Wadoux A M J C, Brus D J and Heuvelink G B M 2019 Sampling design optimization for soil mapping with random forest *Geoderma* **355** 113913
- Wadoux A M J C, Minasny B and McBratney A B 2020 Machine learning for digital soil mapping: applications, challenges and suggested solutions *Earth-Sci. Rev.* **210** 103359
- Wang K, Qi Y B, Guo W J, Zhang J L and Chang Q R 2021a Retrieval and mapping of soil organic carbon using Sentinel-2A spectral images from bare cropland in autumn *Remote Sens.* **13** 1072
- Wang L J, Cheng H, Yang L C and Zhao Y G 2021b Soil organic carbon mapping in cultivated land using model ensemble methods *Arch. Agron. Soil Sci.* **68** 1711–25
- Wang S, Adhikari K, Wang Q, Jin X and Li H 2018 Role of environmental variables in the spatial distribution of soil carbon (C), nitrogen (N), and C:N ratio from the northeastern coastal agroecosystems in China *Ecol. Indic.* **84** 263–72
- Wang S, Zhuang Q, Wang Q, Jin X and Han C 2017 Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China *Geoderma* **305** 250–63
- Wang Y, Wang S, Adhikari K, Wang Q, Sui Y and Xin G 2019 Effect of cultivation history on soil organic carbon status of arable land in northeastern China *Geoderma* **342** 55–64
- Wang Z, Li R, Chen Z, Yao Q, Gao B, Xu M, Yang L, Li M and Zhou C 2022 The estimation of hourly PM_{2.5} concentrations across China based on a Spatial and Temporal Weighted Continuous Deep Neural Network (STWC-DNN) *ISPRS J. Photogramm. Remote Sens.* **190** 38–55
- Weiss M, Jacob F and Duveiller G 2020 Remote sensing for agricultural applications: a meta-review *Remote Sens. Environ.* **236** 111402
- Wetterlind J, Piikki K, Stenberg B and Soderstrom M 2015 Exploring the predictability of soil texture and organic matter content with a commercial integrated soil profiling tool *Eur. J. Soil Sci.* **66** 631–8
- Wiesmeier M et al 2019 Soil organic carbon storage as a key function of soils—a review of drivers and indicators at various scales *Geoderma* **333** 149–62
- Wu C, Wu J, Luo Y, Zhang L and DeGloria S D 2009 Spatial prediction of soil organic matter content using cokriging with remotely sensed data *Soil Sci. Soc. Am. J.* **73** 1202–8
- Wu L, Li L, Yao Y, Qin F, Guo Y, Gao Y and Zhang M 2017 Spatial distribution of soil organic carbon and its influencing factors at different soil depths in a semiarid region of China *Environ. Earth Sci.* **76** 654
- Wu Z H, Liu Y L, Han Y R, Zhou J A, Liu J M and Wu J A 2021 Mapping farmland soil organic carbon density in plains with combined cropping system extracted from NDVI time-series data *Sci. Total Environ.* **754** 142120
- Wu Z, Liu Y, Han Y, Zhou J, Liu J and Wu J 2020 Mapping farmland soil organic carbon density in plains with

- combined cropping system extracted from NDVI time-series data *Sci. Total Environ.* **754** 142120
- Wu Z, Wang B, Huang J, An Z, Jiang P, Chen Y and Liu Y 2019 Estimating soil organic carbon density in plains using landscape metric-based regression Kriging model *Soil Tillage Res.* **195** 104381
- Xie E, Zhang X, Lu F, Peng Y, Chen J and Zhao Y 2021 Integration of a process-based model into the digital soil mapping improves the space-time soil organic carbon modelling in intensively human-impacted area *Geoderma* **409** 115599
- Xu S, Wang M and Shi X 2020 Hyperspectral imaging for high-resolution mapping of soil carbon fractions in intact paddy soil profiles with multivariate techniques and variable selection *Geoderma* **370** 114358
- Yang L, Cai Y Y, Zhang L, Guo M, Li A Q and Zhou C H 2021 A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables *Int. J. Appl. Earth Obs. Geoinf.* **102** 102428
- Yang L, He X, Shen F, Zhou C, Zhu A-X, Gao B, Chen Z and Li M 2020 Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data *Soil Tillage Res.* **196** 104465
- Yang L, Song M, Zhu A-X, Qin C, Zhou C, Qi F, Li X, Chen Z and Gao B 2019 Predicting soil organic carbon content in croplands using crop rotation and Fourier transform decomposed variables *Geoderma* **340** 289–302
- Yang L, Zhu A X, Qi F, Qin C Z, Li B L and Pei T 2013 An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping *Int. J. Geogr. Inf. Sci.* **27** 1–23
- Zeng C Y, Zhu A X, Qi F, Liu J Z, Yang L, Liu F and Li F L 2019 Construction of land surface dynamic feedbacks for digital soil mapping with fusion of multisource remote sensing data *Eur. J. Soil Sci.* **70** 174–84
- Zeng C, Qi F, Zhu A-X and Liu F 2020a Construction of land surface dynamic feedback for digital soil mapping considering the spatial heterogeneity of rainfall magnitude *Catena* **191** 104576
- Zeng C, Yang L, Zhu A-X, Rossiter D G, Liu J, Liu J, Qin C and Wang D 2016 Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method *Geoderma* **281** 69–82
- Zeng L L, Wardlow B D, Xiang D X, Hu S and Li D R 2020b A review of vegetation phenological metrics extraction using time-series, multispectral satellite data *Remote Sens. Environ.* **237** 111511
- Zepp S, Heiden U, Bachmann M, Wiesmeier M, Steininger M and van Wesemael B 2021 Estimation of soil organic carbon contents in croplands of Bavaria from SCMaP soil reflectance composites *Remote Sens.* **13** 3141
- Zeraatpisheh M, Ayoubi S, Jafari A, Tajik S and Finke P 2019 Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran *Geoderma* **338** 445–52
- Zeraatpisheh M, Garosi Y, Reza Owliaie H, Ayoubi S, Taghizadeh-Mehrjardi R, Scholten T and Xu M 2021 Improving the spatial prediction of soil organic carbon using environmental covariates selection: a comparison of a group of environmental covariates *Catena* **208** 105723
- Zhang G L, Liu F and Song X D 2017a Recent progress and future prospect of digital soil mapping: a review *J. Integr. Agric.* **16** 2871–85
- Zhang H, Wu P B, Yin A J, Yang X H, Zhang M and Gao C 2017b Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: a comparison of multiple linear regressions and the random forest model *Sci. Total Environ.* **592** 704–13
- Zhang L, Yang L, Cai Y Y, Huang H L, Shi J J and Zhou C H 2022 A multiple soil properties oriented representative sampling strategy for digital soil mapping *Geoderma* **406** 115531
- Zhang Y K, Hartemink A E, Huang J Y and Townsend P A 2021 Synergistic use of hyperspectral imagery, Sentinel-1 and LiDAR improves mapping of soil physical and geochemical properties at the farm-scale *Eur. J. Soil Sci.* **72** 1690–717
- Zhang Y, Ji W, Saurette D D, Easher T H, Li H, Shi Z, Adamchuk V I and Biswas A 2020 Three-dimensional digital soil mapping of multiple soil properties at a field-scale using regression kriging *Geoderma* **366** 114253
- Zhao M-S, Rossiter D G, Li D-C, Zhao Y-G, Liu F and Zhang G-L 2014 Mapping soil organic matter in low-relief areas based on land surface diurnal temperature difference and a vegetation index *Ecol. Indic.* **39** 120–33
- Zhou T, Geng Y, Chen J, Pan J, Haase D and Lausch A 2020 High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms *Sci. Total Environ.* **729** 138244
- Zhou Y, Hartemink A E, Shi Z, Liang Z and Lu Y 2019 Land use and climate change effects on soil organic carbon in North and Northeast China *Sci. Total Environ.* **647** 1230–8
- Zhu A X, Liu J, Du F, Zhang S J, Qin C Z, Burt J, Behrens T and Scholten T 2015 Predictive soil mapping with limited sample data *Eur. J. Soil Sci.* **66** 535–47
- Zhu A X, Lu G N, Liu J, Qin C Z and Zhou C H 2018 Spatial prediction based on Third Law of Geography *Ann. GIS* **24** 225–40
- Zhu A X, Qi F, Moore A and Burt J E 2010 Prediction of soil properties using fuzzy membership values *Geoderma* **158** 199–206
- Zhu A X, Yang L, Li B L, Qin C Z, English E, Burt J E and Zhou C H 2008 Purposive sampling for digital soil mapping for areas with limited data *Workshop on Digital Soil Mapping for Regions and Countries with Sparse Soil Spatial Data Infrastructures (Rio de Janeiro, Brazil, July 2006)* p 233