Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China

Tao Zhou[a, b, *, 1], Yajun Geng[c, 1], Jie Chen[c], Mengmeng Liu[d], Dagmar Haase[a, b], Angela Lausch[a, b]

[a] Humboldt-Universität zu Berlin, Department of Geography, Unter den Linden 6, 10099 Berlin, Germany

[b] Helmholtz Centre for Environmental Research – UFZ, Department of Computational Landscape Ecology, Permoserstraße 15, 04318 Leipzig, Germany

[c] Nanjing Agricultural University, College of Resources and Environmental Sciences, Weigang 1, 210095 Nanjing, China

[d] Liaoning Technical University, School of Surveying and Geoscience, Zhonghua Road 47, 123000 Fuxin, China

* Corresponding author.

E-mail address: tao.zhou@ufz.de (T. Zhou).

[1] These authors contributed equally to this paper.

Abstract

Soil organic carbon (SOC) has a large impact on soil quality and global climate change. It is therefore important to be able to predict SOC accurately to promote sustainable soil management. Although the synthetic aperture radar (SAR) has many advantages and has been widely used in soil science research, it has rarely been used in previous SOC mapping studies based on remote sensing images. The purpose of this study was to investigate the ability of multi-temporal Sentinel-1A data in SOC prediction, by comparing the predictive performance of random forest (RF) and boosted regression tree (BRT) models in the Heihe River Basin in northwestern China. A set of 162 topsoil (0-20 cm) samples were taken and 15 environmental variables were obtained including land use, topography, climate, and remote sensing images (optical and SAR data). Using a cross-validation procedure to evaluate the performance of the models, three statistical indices were calculated. Overall, both RF and BRT models effectively predicted SOC content, exhibiting similar performance and producing similar spatial distribution patterns of SOC. The results showed that the addition of multi-temporal Sentinel-1A images improved prediction accuracy, with the root mean squared error (RMSE), the mean absolute error (MAE) and the coefficient of determination (R^2) improving by 9.0%, 8.3% and 13.5%, respectively. Furthermore, the combination of all environmental variables had the best prediction performance explaining 75% of SOC variation. The most important environmental variables explaining SOC variation were precipitation, elevation, and temperature. The multi-temporal Sentinel-1A data in RF and BRT models explained 9% and 7%, respectively. The results from our case study highlight the usefulness of multi-temporal Sentinel-1 data in SOC mapping.

Keywords: soil organic carbon, remote sensing, digital soil mapping, random forests, boosted regression tree

1. Introduction

Soil is the largest reservoir of organic carbon on the earth's surface with soil organic carbon (SOC) pools playing an important role in terrestrial ecosystem functioning by affecting soil quality and properties (Schillaci et al., 2017). Even slight changes in SOC storage can have a significant impact on atmospheric carbon concentrations (Chen et al., 2016). In recent decades global warming has attracted widespread attention, and the global average annual air temperature has increased by 0.3-0.6 ℃ over the past 100 years (Cox et al., 2000). SOC pools are susceptible to human disturbance, with land-use change being one of the most important factors affecting soil carbon storage on time scales over several decades (Yang et al., 2016a). Up-to-date SOC maps are essential when trying to understand the spatial variability of SOC, which can help us to maintain soil quality and come up with measures to mitigate global climate change. Therefore, an accurate prediction of SOC content and its distribution patterns is essential.

Taking soil samples from a large number of points for analysis and then performing SOC predictions over large areas is both difficult and costly (Yang et al., 2016b). Digital Soil Mapping (DSM) on the other hand is one way to reduce sampling and analysis costs to predict large-area soil properties and categories from discrete samples (Jeong et al., 2017). Most DSM techniques have been developed from soil landscape models and establish a quantitative relationship between soil observations in the field and readily available environmental variables (Lagacherie et al., 2006; Minasny and McBratney, 2016). Some statistical techniques for predicting SOC have been developed, including mixed linear regression (Doetterl et al., 2013), multiple linear regression (Meersmans et al., 2008), geographically weighted regression (Kumar et al., 2013), and regression kriging based on regression rules (Adhikari et al., 2014). In addition, techniques developed from data mining and machine learning methods to predict SOC have become a popular applied approach (Wang et al., 2018e). Moreover, some studies have reported that tree-based models have better SOC prediction performance, such as boosted regression trees (BRT) (Yang et al., 2016b) and random forests (RF) (Wang et al., 2018b).

The prediction of soil properties requires sufficient environmental data, such as climate, topography, land use/land cover, and satellite imagery, which are commonly used predictors in SOC prediction (Mishra et al., 2010; Ottoy et al., 2017; Zhang et al., 2019). As a data source with broad application prospects, remote sensing data has also been used and made significant contributions to SOC prediction (Grinand et al., 2017). The most commonly used are optical satellite images, whose applications in SOC prediction have been widely developed, although they are affected by clouds. For example, Wang et al. (2018c) used the optical sensor (Landsat 5 TM) to carry out SOC prediction research in Northeast China, and found that optical images provide important information for SOC prediction. Bou Kheir et al. (2010) used different optical image vegetation indices combined with other environmental variables to obtain good prediction results in SOC mapping in Denmark. Similar results were found in other previous SOC prediction studies using different optical images (Forkuor et al., 2017; Mondal et al., 2017; Siewert, 2018). Compared with optical sensors, the synthetic aperture radar (SAR) has the advantages of all-day and all-weather monitoring, but its application in DSM has not been explored or developed to its full potential.

Remote sensing data provide information about soil properties from directly imaging bare soils (Ben‐Dor et al., 2008). However, soils are usually covered by vegetation, which obviously affects the application of remote sensing in soil mapping because remote sensing sensors cannot directly detect soils (Yang et al., 2019). Given this background, many previous DSM studies have incorporated vegetation indices taken from optical images and put into soil prediction models in areas covered by vegetation (Mulder et al., 2011). This approach is promising because vegetation affects the spatial variability of soil properties due to its effects on soil biophysical processes and in turn, the distribution of plant communities is affected by soil properties (Ballabio et al., 2012). Previous studies have demonstrated the effectiveness of using optical remote sensing to characterize soil-vegetation temporal responses to map various soil physicochemical properties (Demattê et al., 2017; Maynard and Levi, 2017). In addition to optical satellite imagery, recent studies have also found that multi-temporal SAR data has the ability to capture soil-vegetation relationships and thus successfully predict soil chemical properties (Ceddia et al., 2017; Yang and Guo, 2019b). The ability to detect vegetation features using SAR data has been demonstrated (Kumar et al., 2019; Wang et al., 2019a). Yang and Guo (2019b) found that the backscatter coefficient of the multi-temporal Sentinel-1 data is a useful indicator to characterize the spatial variability of soil properties in coastal wetlands in eastern China. The recently launched SAR satellites (e.g., Sentinel-1, TerraSAR-X, RISAT-2B and Gaofen-3) have been attracting researchers to predict soil properties using SAR remote sensing techniques, with Sentinel-1 showing good application potential in soil property mapping (Poggio and Gimona, 2017). Although SAR data may provide new opportunities to predict the spatial distribution of soil properties, its application in predicting SOC content is still limited and rarely reported in the literature.

Therefore, the purpose of this study was to evaluate the capability of multi-temporal Sentinel-1A data in SOC prediction by comparing the predictive performance of different tree-based models in Northwest China under complex terrain conditions. For this purpose, 15 environmental variables were used to construct SOC content models with and without multi-temporal Sentinel-1A images, with remote sensing variables including optical (Landsat-8) and SAR (Sentinel-1A) images. The SOC prediction performance from different combinations of 15 environmental variables was then compared based on RF and BRT techniques to explore the contribution and potential of different environmental variables. We then explored the role of including multi-temporal Sentinel-1A images into SOC mapping and quantified the effects of various environmental variables on SOC variation.

2. Materials and methods

2.1. Study area

The study area was the middle and upper reaches of the Heihe River Basin (HRB) in northwestern China (latitude 37°50′–42°40′ North, longitude 98°–102 °East), which is the second largest inland river in China (Gaofeng et al., 2010). It covers 142,900 km^2, but this study was mainly located in Gansu Province and a small part of Qinghai Province (Fig. 1). The Heihe River originates from the Qilian Mountains (Zang et al., 2012), where the elevations in the upper and middle reaches are between 1700 and 5000 m and between 1300 and 1700 m, respectively, while the elevations in the lower reaches are between 900 and 1300 m. The HRB has an arid continental monsoon climate (Luo et al., 2016). The upper reaches are cold and humid with a mean annual temperature (MAT) of 0.7 ℃ and a mean annual precipitation (MAP) of 400–500 mm. In the middle reaches, the MAT is 7.1 ℃ and the MAP is 160.1 mm, whereas in the lower reaches, the MAT is 8.4 ℃, and the MAP is 40.1 mm. The main land cover types in the HRB are farmland, forests, grassland and barren land (Hu et al., 2015). The main soil types are as follows: Leptosols, Arenosols, Fluvisols, Solonchaks, Greyzems, Gypsisols, Kastanozems, Phaeozems, Anthrosols, Gleysols and Calcisols, etc. (Song et al., 2016).

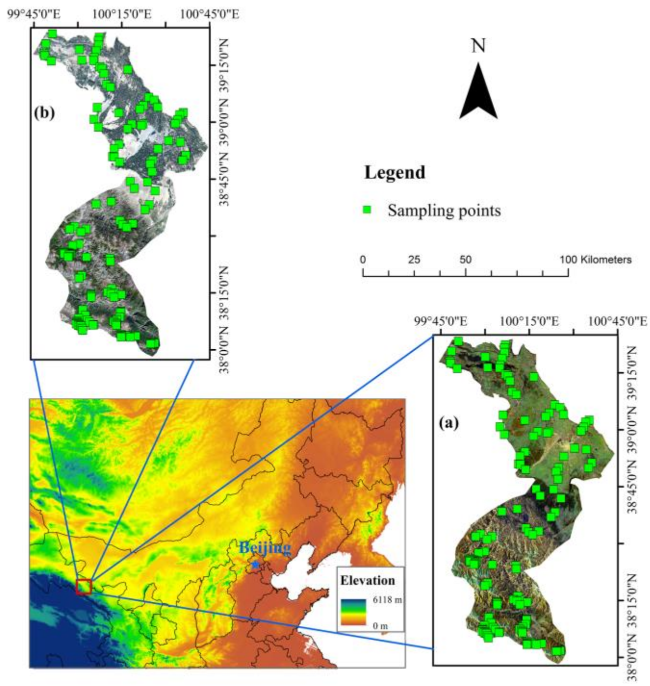

Fig. 1. The study area is located in the Heihe River Basin of China, and the sampling points are displayed in the Sentinel-1A (a) and Landsat-8 (b) composite images of the study area.

2.2. Soil data

Part of the soil data was obtained from the Cold and Arid Regions Science Data Center at Lanzhou (CARD) including 162 topsoil (0–20 cm) samples (Fig. 1), and soil surveys were conducted between 2011 and 2014. The latitude and longitude of all sampling points were located by a hand-held GPS (global positioning system) receiver, and a 1–1.5 m deep soil pit was dug at each soil sample site. In the laboratory, the soil samples were air-dried, ground, and sieved with a 2-mm sieve. The SOC content of the soil samples was then measured by the Walkley–Black method (Nelson and Sommers, 1996). The soil survey dataset recorded the physicochemical properties of soil data such as soil bulk density, soil pH, nutrient concentrations, SOC content etc. In addition, this field survey also investigated the environmental factors of the sites (including landforms, slope, vegetation types, and land-use types, among others). In this study, we only focused on the SOC content of the topsoil because some soil data only investigated the properties of the topsoil.

2.3. Environmental data

2.3.1. Topographic variables

Topographic variables are one of the most commonly used predictor variables for SOC content prediction (Wang et al., 2018a). Topographic variables were extracted from an ASTER GDEM product at 30 m resolution including elevation, slope, aspect, and topographic wetness index (TWI). Among these variables, TWI has become a popular method for characterizing hydrological conditions such as soil moisture and groundwater flow (Naito and Cairns, 2011) and has been widely used in SOC mapping (Lamichhane et al., 2019; Pei et al., 2010). With the use of ArcGIS 10.2 and SAGA GIS, we were able to perform ASTER GDEM processing and calculate these four topographic variables. TWI was calculated using SAGA GIS (Conrad et al., 2015), and the remaining topographic variables were obtained using ArcGIS 10.2.

2.3.2. Land-use data

The land use data used in this study was a raster map provided by the CARD. The land use types in the study area were divided into croplands, forests, grasslands, wetlands, barren lands, and villages.

2.3.3. Climate variables

The climate variables used in this study included the MAP and the MAT of the study area over 50 years (1961-2010), which were provided by the CARD (Yue et al., 2013). The climate variables data were derived as a 500 m grid obtained by interpolating observational data from 34 meteorological observation stations in the HRB, including 21 meteorological observation stations in the HRB and surrounding areas, and 13 national reference stations around the HRB.

2.3.4. Remote sensing variables

Multi-source remote sensing variables extracted from SAR (Sentinel-1A) and optical images (Landsat-8 OLI) were used for SOC prediction. As one of the two satellites of Sentinel-1, Sentinel-1A is equipped with a C-band SAR instrument and was launched on 3rd April, 2014 (Navarro et al., 2016; Zhou et al., 2018). Four Sentinel-1A images (single-look complex (SLC) products) from 2014 and 2015 covering the study area were downloaded from the European Space Agency and are in IW (interferometric wide swath) mode (Table 1). The Landsat-8 OLI images from 13th September, 2015 were collected from the Earth Explorer website with cloud cover < 10%. We used ENVI 5.3 software to pre-process Landsat-8 OLI data, including radiance calibration and atmospheric correction (Jia et al., 2014). Atmospheric correction was performed by the ENVI FLAASH model. Preprocessing of all SAR images was performed in SARscape 5.2 software, including multi-look, co-registration, speckle filtering, geocoding, and radiometric calibration. In this study, the Lee filter (Lee, 1986) was used to filter the speckle noise in the SAR data. The digital numbers (DN) of the SAR data were converted to a decibel (dB) scale backscatter coefficient and the images were geocoded using the ASTER GDEM.

Table 1 The parameter information of the Sentinel-1A data obtained in this study.

Date	Imaging model	Polarization	Incident Angle (◦)	Direction
26th October 2015	IW	VH	39.36	Ascending
2nd October 2015	IW	VV	39.35	Ascending
11th May 2015	IW	VV	39.35	Ascending
12th November 2014	IW	VV	39.35	Ascending

A total of eight environmental variables were derived from remote sensing images, including four variables from Sentinel-1A data, and four variables from Landsat-8 OLI images. The four SAR data predictors were constructed using backscatter coefficients from four Sentinel-1A images, respectively. Landsat imagery has been effectively used to predict SOC, where visible red, near-infrared, and short-wave infrared bands (representing vegetation growth, cover, and biomass, respectively) have been selected and reported as key variables for predicting SOC in previous studies (Qi et al., 2019; Yang et al., 2016c). For example, Bian et al. (2019) selected these three bands combined with terrain variables to predict SOC in the coastal areas of Northeast China, and found these bands to be the main predictor variables for mapping SOC. In this study, these three bands from Landsat-8 OLI (red band 4, near-infrared band 5, and shortwave infrared band 6) were also selected as environmental variables. In addition, the normalized difference vegetation index (NDVI) was calculated as a predictor using Landsat-8 OLI data.

2.4. Modelling techniques

This study selected and compared two commonly used machine learning techniques for SOC mapping: RF and BRT. Both modeling techniques have the ability and the advantage of being able to measure the relative importance of environmental variables (see Triviño et al. (2011) for the evaluation methods of importance provided by RF and BRT) and were also used to assess the importance of the variables in this study. They have been reported to perform well in SOC prediction for various types of landscapes (Martin et al., 2014; Veronesi and Schillaci, 2019; Wang et al., 2018a).

2.4.1. Random forest

RF is a tree-based ensemble learning method introduced by Breiman (2001) for classification and regression purposes (Rial et al., 2017; Zhou et al., 2017), which generates multiple trees without pruning (Grimm et al., 2008). During training, each tree is produced based on a unique bootstrap sample (with replacement) from the entire training sample dataset (Wang et al., 2018b). Compared to decision trees, the bootstrap sampling method makes RF less sensitive to over-fitting (Heung et al., 2014). In the RF model, the data are classified into "in-bag" data and "out-of-bag (OOB)" data (Wang et al., 2018d). The OOB samples were the training data left over from the bootstrap samples, which were used to estimate general errors and the importance of the variables (Were et al., 2015). The "in-bag" samples were used for model training.

The RF algorithm requires the following three parameters to generate a prediction model: (i) the number of regression trees (ntree), (ii) the number of randomly selected variables at each node (mtry), and (iii) the minimum number of terminal nodes (node size) (Friedman and Meulman, 2003). To optimize these parameters, we tested and compared different combinations of these parameters. In this study, these combinations finally determined the optimal values for the ntree, mtry and node size by performing a grid-search approach using the "caret" package (Siewert, 2018; Wang et al., 2018b).

2.4.2. Boosted regression trees

The BRT method is a machine learning algorithm developed by Friedman et al. (2000), which combines many regression trees and a boosting technique to improve the predictive performance of many single models (Wang et al., 2018a). By using an iterative method, the boosting algorithm develops a final model and gradually adds trees to the model (Muller et al., 2013). BRT relies on a stochastic gradient boosting procedure that can improve model performance and reduce the risk of over-fitting through numerical optimization and regularization (Friedman, 2002).

BRT modeling is controlled by the following four parameters: the learning rate (LR), tree complexity (TC), the number of trees (NT) and the bag fraction (BF). LR determines the contribution of each tree to the growing model. BF sets the proportion of data selected in each step of the modeling process, TC controls the number of splits, while NT is determined by the combination of LR and TC (Wang et al., 2016). The BRT model needs to be adjusted by setting the parameters before making predictions (Elith et al., 2008). To optimize these four parameters, some combinations of parameter values (LR, TC, NT and BF) were tested. We also performed a grid-search approach using the “caret” package in the R software to determine the optimal values for these parameters. The combination with the minimum predictive deviance was determined as the best combination of parameters.

2.5. Statistical analyses

We used SPSS 24.0 software to conduct a descriptive statistical analysis of SOC data and environmental variables. The RF and BRT models were developed using R-software. In this study, we fitted the RF and BRT models using the “randomForest”-package and the “gbm”-package, respectively.

2.6. Model evaluation

We used the RF and BRT methods to construct SOC content models with and without multi-temporal Sentinel-1A images (Fig. 2), allowing an evaluation of the contribution of the multi-temporal Sentinel-1A images to the SOC content prediction. The above models were constructed from different combinations of environmental variables: Model I and Model II were constructed using optical (Landsat-8 OLI) and SAR (Sentinel-1A) images, respectively, while Model III was a combination of optical and SAR images; Model IV included climate, land use, optical images, and topography, while Model V was constructed by adding SAR images to Model IV (Table 2). The predictive quality of these models was evaluated by a 10-fold cross-validation procedure and the following three common statistical indices were calculated: the mean absolute error (MAE), the root mean square error (RMSE) and the coefficient of determination (R^2). Cross-validation avoids testing training data and is a useful technique for evaluating model performance. Recent studies using the DSM model for SOC prediction have been reviewed and found cross-validation techniques to be one of the most commonly used methods for evaluating results (Lamichhane et al., 2019). These indicators were calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right| \quad (1)$$

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

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

where n represents the number of samples and P_i and O_i represent the predicted and observed SOC content at site i, respectively.

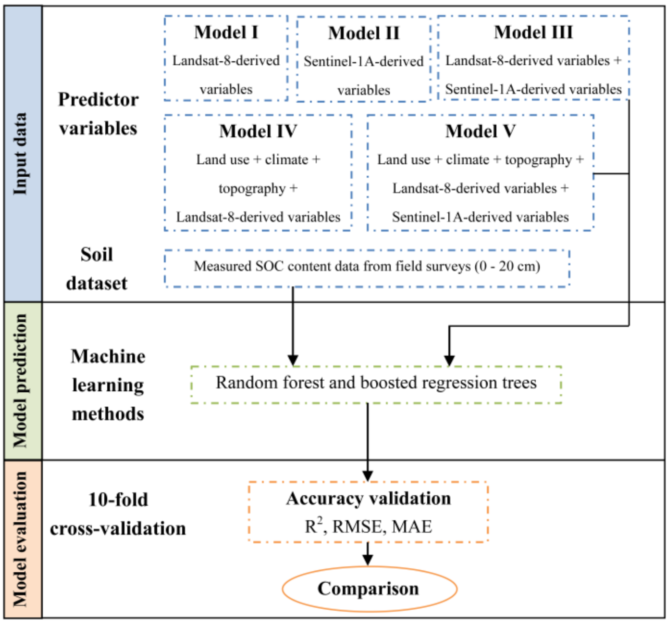

Fig. 2. Workflow diagram for predicting SOC content in this study.

Table 2 Different combinations of predictor variables for SOC content.

NO.	Model	Variables
a	Model I	Optical imagery
b	Model II	SAR imagery
c	Model III	SAR and optical images
d	Model IV	Land use + climate + topography + optical imagery
e	Model V	Land use + climate + topography + remote sensing data (including SAR and optical images)

3. Results and Discussion

3.1. Descriptive statistics of sampled SOC content

Table 3 Summary statistics of measured SOC content and values of environmental variables at sample locations.

	Minimum	Maximum	Mean	Median	Standard deviation (SD)	Skewness
SOC (g/kg)	1.75	139.83	41.44	30.55	36.68	0.94
LnSOC (g/kg)	0.56	4.94	3.24	3.41	1.08	-0.31
Elevation (m)	1347.00	4329.00	2604.44	2927.00	838.19	-0.34
Aspect (degree)	0.00	347.32	173.98	193.39	109.58	-0.12
Slope (degree)	0.00	44.89	14.60	13.14	11.33	0.65
TWI	2.38	9.32	4.45	4.31	1.18	0.73
BC_1 (dB)	-21.47	-0.51	-11.99	-11.78	3.75	0.14
BC_2 (dB)	-20.02	-1.29	-10.16	-10.08	3.54	-0.01
BC_3 (dB)	-18.63	-2.50	-10.04	-9.44	3.42	-0.22
BC_4 (dB)	-21.72	-2.88	-16.55	-16.59	2.82	0.77
band_4 (digital number)	156.00	2420.00	813.75	549.50	592.07	1.23
band_5 (digital number)	1506.00	5469.00	3049.88	2997.00	787.85	0.28
band_6 (digital number)	547.00	3795.00	2058.16	1964.00	736.27	0.25
NDVI	0.03	0.85	0.59	0.69	0.23	-1.01
MAP (mm)	108.31	754.35	389.84	449.32	203.28	-0.14
MAT (degree celsius)	-8.61	8.03	0.79	-0.63	5.05	0.29

Notes: LnSOC, log-transformed SOC; BC_1, BC_2, BC_3, and BC_4 correspond to the backscatter coefficients of Sentinel-1A images from different acquisition dates: 12th November 2014, 11th May 2015, 2nd October 2015, and 26th October 2015, respectively; Band_4, band_5, and band_6 correspond to bands 4 to 6 of the Landsat-8 OLI image (September 13th, 2015), respectively.

The statistical characteristics of the measured SOC content and the values of the environmental variables at the sample location are shown in Table 3. The measured SOC content showed a slightly skewed distribution (with a skewness value of 0.94), varying from 1.75 to 139.83 g kg^{-1}, with an average of 41.44 g kg^{-1}. Therefore, for all prediction models in this paper, the SOC content was converted by using the natural logarithm of SOC (LnSOC) to reduce the rightward skew of the untransformed SOC (reducing skewness from 0.94 to 0.31). The standard deviation of raw SOC and LnSOC were 36.68 and 1.08 g kg^{-1}, respectively, which were both less than their mean value.

3.2. Evaluation of model predictions

To assess the ability of multi-temporal Sentinel-1A data for predicting SOC content, we used RF and BRT methods to construct SOC content models: Model I and Model II included only optical and SAR data, respectively, while Model III used both optical and SAR data. Model V (land use, climate, topography, SAR and optical images) and Model IV (land use, climate, topography, and optical imagery) were combinations of environmental variables with and without SAR data, respectively. Table 4 shows the predictive power of each of the above models.

Table 4 Comparison of the predictive power using different combinations of predictor variables.

Modeling technique	Model	MAE	RMSE	R^2
RF	Model I	0.63	0.78	0.50
	Model II	0.86	1.01	0.19
	Model III	0.58	0.73	0.56
	Model IV	0.46	0.57	0.74
	Model V	0.44	0.55	0.75
BRT	Model I	0.60	0.77	0.52
	Model II	0.85	1.00	0.22
	Model III	0.55	0.70	0.59
	Model IV	0.46	0.57	0.74
	Model V	0.45	0.55	0.75

Notes: Model I, optical imagery alone; Model II, SAR imagery alone; Model III, SAR and optical images; Model IV, Land use + climate + topography + optical imagery; Model V, Land use + climate +

topography + remote sensing data (including SAR and optical images).

Overall, our results showed that the prediction accuracy levels of the BRT and RF methods were similar; for the RF model, the range of the verification indices were: MAE (in the range of 0.44 to 0.86), RMSE (in the range of 0.55 to 1.01) and R^2 (in the range of 0.19 to 0.75); the range of the verification indices for the BRT model were: MAE (in the range of 0.45 to 0.85), RMSE (in the range of 0.55 to 1.00) and R^2 (in the range of 0.22 to 0.75). The RF and BRT models have been reported to have a stable predictive power for SOC mapping (Yang et al., 2016b). There are some uncertainties in SOC research, including the variability of measured SOC values, sampling errors, laboratory analysis errors and modeling errors (Krishnan et al., 2007; Terribile et al., 2011). Although we did not compare RF and BRT models with other predictive models, some previous studies have shown better performance for BRT and RF models in SOC prediction compared to other models, such as multiple linear models (Razakamanarivo et al., 2011) or support vector machine (SVM) models (Forkuor et al., 2017). However, other studies have reported opposing performance results from the tree-based model, with SVM showing better performance in SOC prediction than RF (Rossel and Behrens, 2010; Were et al., 2015). The results of this study were consistent with the results of Yang et al. (2016b), who reported that BRT and RF models have a similar ability to predict SOC concentrations in the Qinghai-Tibet Plateau. Lamichhane et al. (2019) reviewed the SOC mapping studies from 2013 to February 2019 and found that 13 out of 17 studies using RF models showed that RF models obtained better predictions than other DSM techniques. It would appear that there is no single machine learning algorithm that works best for all ecosystems. Hence, it is important to evaluate the performance of different models under different conditions and environmental input variables.

For both RF and BRT techniques, Model I performed better than Model II, indicating that the predictive power of optical images was superior to that of SAR images in this study area. The predictive performance of SOC improved when optical images were combined with multi-temporal Sentinel-1A data. The addition of multi-temporal Sentinel-1A data using the BRT model compared to the use of only optical images improved RMSE (from 0.77 to 0.70), MAE (from 0.60 to 0.55) and R^2 (from 0.52 to 0.59) by 9.0%, 8.3% and 13.5%, respectively. Similar improvements were also observed for the RF model. This was expected because the prediction accuracy improved when more useful information was added. Previous studies have also explored the usefulness of other remote sensing variables in SOC prediction. For example, Wang et al. (2018b) looked at the effect of adding seasonal fractional cover data on SOC prediction and was able to improve the RMSE by 2.8–5.9% at 0–30 cm soil depths. The results of Yang et al. (2015) showed how useful optical imagery (Landsat TM) was in predicting SOC content. However, previous studies on SOC prediction mainly used optical images such as Landsat and MODIS, ignoring the potential of SAR data. Compared with a single sensor, the method of multi-sensor (i.e., Landsat-8 OLI and Sentinel-1A sensors) SOC mapping in this study improved the prediction accuracy, indicating that multi-temporal Sentinel-1A images are useful for SOC prediction in the study area. This was also supported by Poggio and Gimona (2017) who used multi-source remote sensing data to predict soil properties, proving that Sentinel-1 is useful for predicting soil physical and chemical properties.

Similar to the improvement in accuracy between Model I and Model III, an improvement in prediction performance was also observed between Model IV and Model V due to the addition of multi-temporal Sentinel-1A data. However, the latter only observed a relatively slight improvement in accuracy, which was lower than the former. These results further demonstrated the potential of multi-temporal Sentinel-1A data as a predictor to improve SOC prediction accuracy in the study area. A previous study found that the backscatter coefficient of C-band SAR data under frozen conditions can represent vegetation and surface structure properties associated with soil properties, particularly SOC (Bartsch et al., 2016). In a soil mapping study of coastal wetlands in eastern China, Yang and Guo (2019a) found that multi-temporal Sentinel-1 data can capture the dynamic characteristics of vegetation and the relationship between soil properties and vegetation to help predict soil properties. Model V which combined all environmental variables had the highest value R^2 (0.75) and achieved the lowest values for MAE (0.44) and RMSE (0.55). This R^2 value revealed that Model V could explain 75% of SOC variation. Compared with previous studies conducted in this study area, the method based on multi-source remote sensing variables in this study yielded a more promising SOC prediction performance. For example, Zhang and Shao (2014) also conducted a SOC mapping study in the HRB, explaining only 47% of SOC variation. Wang et al. (2014) carried out SOC mapping based on MODIS data and climate variables, explaining 69% of SOC variation. Yang et al. (2015) also developed a BRT model to map SOC content near the HRB using Landsat 5 TM combined with topographic and climate variables, explaining 71% of SOC variation.

Although the prediction accuracy obtained with multi-source remote sensing variables was satisfactory for this study, further improvement is still needed. The predictor variables used in this study had different spatial scales, but we used a single analytical scale typically performed in DSM. It is well known that the spatial scale of predictor variables can have a significant impact on prediction accuracy (Drăguţ et al., 2009). Siewert (2018) used different environmental variables combined with machine learning algorithms to predict SOC in the northernmost part of Sweden, and found that the power of predicting SOC dropped significantly between 30 and 100 m resolution. Chi et al. (2019) compared the prediction accuracy of soil total nitrogen using environmental variables with different spatial scales (100 m, 200 m, 400 m, and 800 m) on Chongming Island in China, and found that the 100 m scale obtained the highest accuracy. At the same time, they found that soil total nitrogen prediction models with different spatial resolutions produced similar spatial patterns of soil total nitrogen. In addition, land surface characteristics change with time and the acquisition time of remote sensing data for predicting soil properties also affects the prediction accuracy (Forkuor et al., 2017). Therefore, it would be worth exploring these avenues further to improve prediction accuracy.

3.3. The relative importance of environmental data

The relative importance of each environmental factor obtained from the RF and BRT methods in Model V is shown in Figure 3, and we increased the factor comparability by normalizing the environmental factors to 100%. For both RF and BRT models, the four most important variables were MAP, elevation, MAT, and band_4, with MAP being the most important environmental variable. This indicated that these environmental variables were the main environmental variables affecting SOC variation in the study area. For the BRT model, climate variables (relative importance of 51%) were the main explanatory variables for SOC variation, followed by topographic variables (27%) and remote sensing variables (21%). In the RF model, remote sensing variables, topographic and climate variables explained 34%, 32% and 30% of SOC variation, respectively. In addition, the SAR remote sensing variables in the RF and BRT models explained 9% and 7% SOC variation, respectively.

Precipitation and temperature were the main climate variables affecting SOC distribution. For both RF and BRT models, both MAP and MAT were located in the top three most important environmental variables affecting SOC spatial variation. This is mainly due to the close relationship between climate variables and soil moisture, affecting plant growth and net primary productivity (Wang et al., 2018b). Climate variables have a profound impact on the decomposition and accumulation of SOC. In the Alpine ecosystem, climate variables affect hydrological and ecological functions, which in turn affect SOC variation. This was consistent with previous studies that also emphasize the importance of climate variables in predicting SOC (Hobley et al., 2015; Richardson et al., 2017).

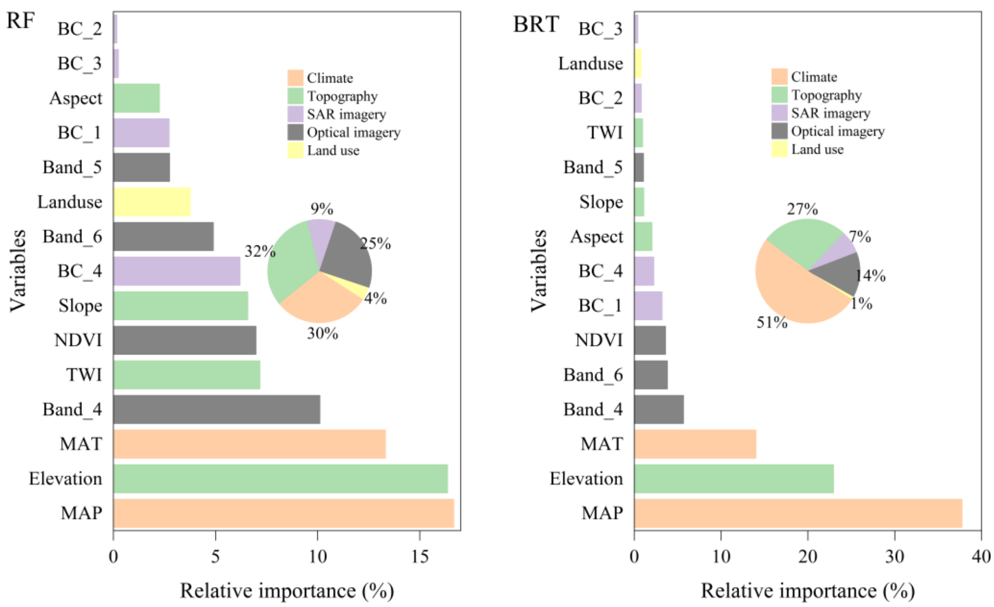

Fig. 3. The relative importance of each environmental factor obtained from the RF and the BRT methods in Model V (increasing factor comparability by normalizing the environmental factors to 100%). Model V, Land use + climate + topography + remote sensing data (including SAR and optical images); BC_1, BC_2, BC_3, and BC_4 correspond to the backscatter coefficients of Sentinel-1A images from different acquisition dates: 12^{th} November 2014, 11^{th} May 2015, 2^{nd} October 2015, and 26^{th} October 2015, respectively; Band_4, band_5, and band_6 correspond to bands 4 to 6 of the Landsat-8 OLI image (September 13^{th}, 2015), respectively.

Our research showed that remote sensing images were important variables in SOC prediction, including optical remote sensing data and multi-temporal Sentinel-1A data. Therefore, the spectral reflectance, the backscatter coefficient of multi-temporal Sentinel-1A data, and the derived vegetation index are practical indicators of SOC prediction. Zhong et al. (2018) found that SOC dynamics were mainly affected by vegetation and soil characteristics under similar climate conditions. Derived vegetation indices can significantly represent vegetation biomass and density (Li et al., 2019; Zhao et al., 2016), so they were important environmental variables for predicting SOC. Using SAR data to predict soil properties depends on the sensitivity of the backscatter coefficient to changes in land surface conditions and soil moisture (Kasischke et al., 1997). Yang et al. (2019) reported that a successful use of SAR data to predict soil properties can be explained by the relationship observed in the soil-vegetation system based on multi-temporal Sentinel-1 data. The relationship in the soil-vegetation system observed by remote-sensing techniques can help explain the spatial variation of SOC (Yang and Guo, 2019b) and is supported by other previous studies (Ceddia et al., 2017; Maynard and Levi, 2017). In addition, some studies have reported that Sentinel-1 images can provide useful information for detecting vegetation (e.g., Castillo et al., 2017; Muro et al., 2016). The results showed that not only optical remote sensing images can be used for SOC prediction, but also the backscattering coefficients of multi-temporal Sentinel-1 images are useful for SOC mapping. However, this finding using RF model was different from the results of previous studies, which reported that topographic variables were more important environmental variables than remote sensing variables for predicting SOC (Wang et al., 2018c; Wang et al., 2017). This difference was mainly due to the fact that the use of multi-temporal Sentinel-1A data in this study improved the predictive power of remote sensing variables compared to previous studies using only optical remote sensing data.

As one of the five soil forming factors, topography can affect water temperature conditions and the distribution of soil-forming materials. Among the four terrain variables, elevation had the highest relative importance. The topographic variables determine the direction and the rate of material migration, where elevation affects the vertical distribution of water heat and affects the decomposition and transformation of SOC (Martin et al., 2014). Similar to this study, the results of Wang et al. (2019b) and Dong et al. (2019) also found that elevation was the most important topographic variable for predicting SOC. However, the relative importance of topographic variables was lower than climate variables in the BRT model. This was consistent with the results of the SOC prediction study conducted by Yang et al. (2016b) on the Qinghai-Tibet Plateau, which explained that the influence of topography was mediated by vegetation.

3.4. The spatial prediction of SOC content

The SOC content maps were obtained in Model V based on RF and BRT techniques, respectively (Figure 4). For Model V, the average and SD values of the predicted SOC content obtained by the RF method were 24.14 and 20.12 g kg^{-1}, respectively, whereas the average and SD values of the predicted SOC content obtained by the BRT method were 24.91 and 21.67 g kg^{-1}, respectively. The average and SD values of the predicted SOC content obtained by all models were lower than the observed SOC. The predicted SOC variation was less than the measured value. These results were consistent with the results of previous studies that had conducted SOC predictions (Adhikari and Hartemink, 2015; Wang et al., 2018c).

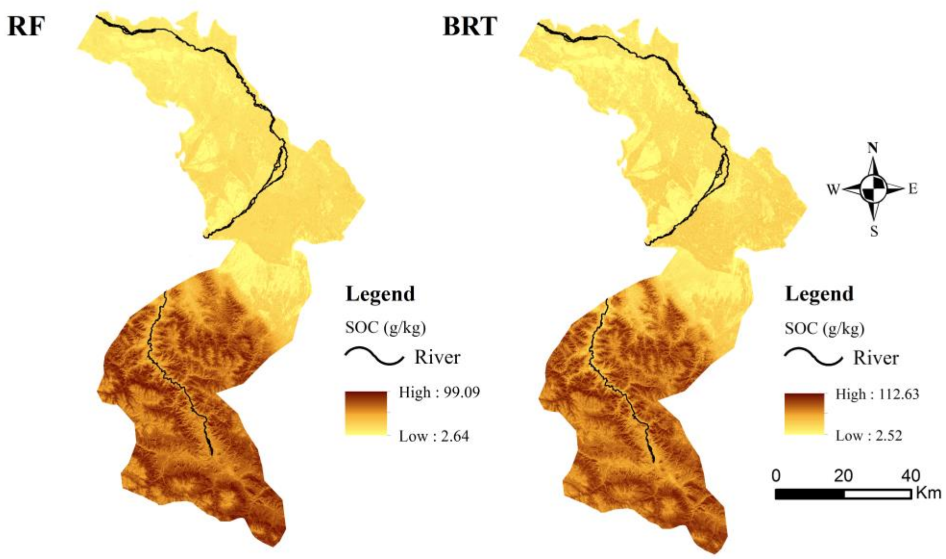

Fig. 4. SOC content maps obtained from the Model V based on RF and BRT techniques (Model V: Land use + climate + topography + remote sensing data (including SAR and optical images)).

The spatial distribution maps of SOC content obtained from different models were similar, and a strong SOC spatial variation was observed on all of the distribution maps. The predicted SOC content in the southern part of the study area was the highest. The main land type in the southern part of the study area was the plateau forests and grasslands, displaying higher elevations. This part of the study area also had higher rainfall and lower average temperatures, which favored low SOC turnover and explained the high SOC content. This was similar to the results of the SOC prediction study conducted by Song et al. (2016) in the HRB. In addition, previous studies reported an increase in SOC content as altitude increased (Tsui et al., 2013; Wang et al., 2018c). Correspondingly, the northern region dominated by cultivated land, barren and urban areas had a low SOC content. Compared with the southern part of the study area, the precipitation in the north was lower and the temperature was higher. Agro-ecosystems near rivers had a relatively high predicted SOC.

4. Conclusions

We applied RF and BRT models to predict the SOC content in the HRB of China using multi-source remote sensing variables. The following main conclusions can be drawn from this study: (1) Both BRT and RF models effectively and accurately predicted SOC, showing similar performance. (2) The addition of the multi-temporal Sentinel-1A data improved the predictive performance, with RMSE, MAE and R^2 improving by 9.0%, 8.3% and 13.5%, respectively. The combination of all environmental variables achieved the best results with the highest value of R^2 (0.75) and the lowest values of MAE (0.44) and RMSE (0.55). (3) Precipitation, elevation, and temperature were the main variables explaining SOC variation. (4) RF and BRT models produced similar spatial distribution maps of SOC content, with SOC content levels in the southern regions significantly higher than elsewhere. In future research it would be worth exploring the implementation of other remote sensing sensors to predict other soil properties.

Acknowledgments

Tao Zhou would like to thank the China Scholarship Council for its financial support (NO. 201806850090). We are very grateful for the data support from CARD.

References

Adhikari, K., Hartemink, A.E., 2015. Digital Mapping of Topsoil Carbon Content and Changes in the Driftless Area of Wisconsin, USA. Soil Science Society of America Journal 79, 155-164.

Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. PLOS ONE 9, e105519.

Ballabio, C., Fava, F., Rosenmund, A., 2012. A plant ecology approach to digital soil mapping, improving the prediction of soil organic carbon content in alpine grasslands. Geoderma 187-188, 102-116.

Bartsch, A., Widhalm, B., Kuhry, P., Hugelius, G., Palmtag, J., Siewert, M.B., 2016. Can C-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra? Biogeosciences 13, 5453-5470.

Ben‐Dor, E., Taylor, R.G., Hill, J., Demattê, J.A.M., Whiting, M.L., Chabrillat, S., Sommer, S., 2008. Imaging Spectrometry for Soil Applications, Advances in Agronomy. Academic Press, pp. 321-392.

Bian, Z., Guo, X., Wang, S., Zhuang, Q., Jin, X., Wang, Q., Jia, S., 2019. Applying statistical methods to map soil organic carbon of agricultural lands in northeastern coastal areas of China. Archives of Agronomy and Soil Science, 1-13.

Bou Kheir, R., Greve, M.H., Bøcher, P.K., Greve, M.B., Larsen, R., McCloy, K., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. Journal of Environmental Management 91, 1150-1160.

Breiman, L., 2001. Random forests. Machine Learning 45, 5-32.

Castillo, J.A.A., Apan, A.A., Maraseni, T.N., Salmo, S.G., 2017. Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. ISPRS Journal of Photogrammetry and Remote Sensing 134, 70-85.

Ceddia, M.B., Gomes, A.S., Vasques, G.M., Pinheiro, É.F.M., 2017. Soil Carbon Stock and Particle Size Fractions in the Central Amazon Predicted from Remotely Sensed Relief, Multispectral and Radar Data. Remote Sensing 9, 124.

Chen, L.-F., He, Z.-B., Du, J., Yang, J.-J., Zhu, X., 2016. Patterns and environmental controls of soil organic carbon and total nitrogen in alpine ecosystems of northwestern China. CATENA 137, 37-43.

Chi, Y., Zhao, M., Sun, J., Xie, Z., Wang, E., 2019. Mapping soil total nitrogen in an estuarine area with high landscape fragmentation using a multiple-scale approach. Geoderma 339, 70-84.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geoscientific Model Development 8, 1991-2007.

Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A., Totterdell, I.J., 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 184-187.

Demattê, J.A.M., Sayão, V.M., Rizzo, R., Fongaro, C.T., 2017. Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing. Geoderma 302, 39-51.

Doetterl, S., Stevens, A., van Oost, K., Quine, T.A., van Wesemael, B., 2013. Spatially-explicit regional-scale prediction of soil organic carbon stocks in cropland using environmental variables and mixed model approaches. Geoderma 204-205, 31-42.

Dong, W., Wu, T., Luo, J., Sun, Y., Xia, L., 2019. Land parcel-based digital soil mapping of soil nutrient properties in an alluvial-diluvia plain agricultural area in China. Geoderma 340, 234-248.

Drăguţ, L., Schauppenlehner, T., Muhar, A., Strobl, J., Blaschke, T., 2009. Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling. Computers & Geosciences 35, 1875-1883.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77, 802-813.

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. PLOS ONE 12, e0170478.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics 28, 337-407.

Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis 38, 367-378.

Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine 22, 1365-1381.

Gaofeng, Z., Yonghong, S., Chunlin, H., Qi, F., Zhiguang, L., 2010. Hydrogeochemical processes in the groundwater environment of Heihe River Basin, northwest China. Environmental Earth Sciences 60, 139-153.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102-113.

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

Heung, B., Bulmer, C.E., Schmidt, M.G., 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma 214-215, 141-154.

Hobley, E., Wilson, B., Wilkie, A., Gray, J., Koen, T., 2015. Drivers of soil organic carbon storage and vertical distribution in Eastern Australia. Plant and Soil 390, 111-127.

Hu, X., Lu, L., Li, X., Wang, J., Guo, M., 2015. Land Use/Cover Change in the Middle Reaches of the Heihe River Basin over 2000-2011 and Its Implications for Sustainable Water Resource Management. PLOS ONE 10, e0128960.

Jeong, G., Oeverdieck, H., Park, S.J., Huwe, B., Ließ, M., 2017. Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain. CATENA 154, 73-84.

Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., Li, B., 2014. Land cover classification using Landsat 8

Operational Land Imager data in Beijing, China. Geocarto International 29, 941-951.

Kasischke, E.S., Melack, J.M., Craig Dobson, M., 1997. The use of imaging radars for ecological applications—A review. Remote Sensing of Environment 59, 141-156.

Krishnan, P., Bourgeon, G., Seen, D.L., Nair, K.M., Prasanna, R., Srinivas, S., Muthusankar, G., Dufy, L., Ramesh, B.R., 2007. Organic carbon stock map for soils of southern India: A multifactorial approach. Current Science 93, 706-710.

Kumar, A., Kishore, B.S.P.C., Saikia, P., Deka, J., Bharali, S., Singha, L.B., Tripathi, O.P., Khan, M.L., 2019. Tree diversity assessment and above ground forests biomass estimation using SAR remote sensing: A case study of higher altitude vegetation of North-East Himalayas, India. Physics and Chemistry of the Earth, Parts A/B/C 111, 53-64.

Kumar, S., Lal, R., Liu, D., Rafiq, R., 2013. Estimating the spatial distribution of organic carbon density for the soils of Ohio, USA. Journal of Geographical Sciences 23, 280-296.

Lagacherie, P., McBratney, A., Voltz, M., 2006. Digital soil mapping: an introductory perspective. Elsevier.

Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma 352, 395-413.

Lee, J.-S., 1986. Speckle suppression and analysis for synthetic aperture radar images. Optical Engineering 25, 255636.

Li, C., Li, Y., Li, M., 2019. Improving Forest Aboveground Biomass (AGB) Estimation by Incorporating Crown Density and Using Landsat 8 OLI Images of a Subtropical Forest in Western Hunan in Central China. Forests 10, 104.

Luo, K., Tao, F., Moiwo, J.P., Xiao, D., 2016. Attribution of hydrological change in Heihe River Basin to climate and land use change in the past three decades. Scientific Reports 6, 33704.

Martin, M.P., Orton, T.G., Lacarce, E., Meersmans, J., Saby, N.P.A., Paroissien, J.B., Jolivet, C., Boulonne, L., Arrouays, D., 2014. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma 223-225, 97-107.

Maynard, J.J., Levi, M.R., 2017. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. Geoderma 285, 94-109.

Meersmans, J., De Ridder, F., Canters, F., De Baets, S., 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.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301-311.

Mishra, U., Lal, R., Liu, D., Van Meirvenne, M., 2010. Predicting the Spatial Variation of the Soil Organic Carbon Pool at a Regional Scale. Soil Science Society of America Journal 74, 906-914.

Mondal, A., Khare, D., Kundu, S., Mondal, S., Mukherjee, S., Mukhopadhyay, A., 2017. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. The Egyptian Journal of Remote Sensing and Space Science 20, 61-70.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1-19.

Muller, D., Leitao, P.J., Sikor, T., 2013. Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees. Agricultural Systems 117, 66-77.

Muro, J., Canty, M., Conradsen, K., Hüttich, C., Nielsen, A.A., Skriver, H., Remy, F., Strauch, A., Thonfeld, F., Menz, G., 2016. Short-Term Change Detection in Wetlands Using Sentinel-1 Time Series. Remote Sensing 8, 795.

Naito, A.T., Cairns, D.M., 2011. Relationships between Arctic shrub dynamics and topographically derived hydrologic characteristics. Environmental Research Letters 6, 045506.

Navarro, A., Rolim, J., Miguel, I., Catalão, J., Silva, J., Painho, M., Vekerdy, Z., 2016. Crop Monitoring Based on SPOT-5 Take-5 and Sentinel-1A Data for the Estimation of Crop Water Requirements. Remote Sensing 8, 525.

Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. Methods of soil analysis part 3—chemical methods, 961-1010.

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

Pei, T., Qin, C.-Z., Zhu, A.X., Yang, L., Luo, M., Li, B., Zhou, C., 2010. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. Ecological Indicators 10, 610-619.

Poggio, L., Gimona, A., 2017. Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas. Science of The Total Environment 579, 1094-1110.

Qi, L., Wang, S., Zhuang, Q., Yang, Z., Bai, S., Jin, X., Lei, G., 2019. Spatial-Temporal Changes in Soil Organic Carbon and pH in the Liaoning Province of China: A Modeling Analysis Based on Observational Data. Sustainability 11, 3569.

Razakamanarivo, R.H., Grinand, C., Razafindrakoto, M.A., Bernoux, M., Albrecht, A., 2011. Mapping organic carbon stocks in eucalyptus plantations of the central highlands of Madagascar: A multiple regression approach. Geoderma 162, 335-346.

Rial, M., Martínez Cortizas, A., Rodríguez-Lado, L., 2017. Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils. Science of The Total Environment 609, 1411-1422.

Richardson, H.J., Hill, D.J., Denesiuk, D.R., Fraser, L.H., 2017. A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada). Giscience & Remote Sensing 54, 573-591.

Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46-54.

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., 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. Science of The Total Environment 601-602, 821-832.

Siewert, M.B., 2018. High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment. Biogeosciences 15, 1663-1682.

Song, X.-D., Brus, D.J., Liu, F., Li, D.-C., Zhao, Y.-G., Yang, J.-L., Zhang, G.-L., 2016. Mapping soil organic carbon content by geographically weighted regression: A case study in the Heihe River Basin, China. Geoderma 261, 11-22.

Terribile, F., Coppola, A., Langella, G., Martina, M., Basile, A., 2011. Potential and limitations of using soil mapping information to understand landscape hydrology. Hydrology and Earth System

Sciences 15, 3895-3933.

Triviño, M., Thuiller, W., Cabeza, M., Hickler, T., Araújo, M.B., 2011. The Contribution of Vegetation and Landscape Configuration for Predicting Environmental Change Impacts on Iberian Birds. PLOS ONE 6, e29373.

Tsui, C.-C., Tsai, C.-C., Chen, Z.-S., 2013. Soil organic carbon stocks in relation to elevation gradients in volcanic ash soils of Taiwan. Geoderma 209-210, 119-127.

Veronesi, F., Schillaci, C., 2019. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. Ecological Indicators 101, 1032-1044.

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I., Sides, T., 2018a. Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. Ecological Indicators 88, 425-438.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018b. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Science of The Total Environment 630, 367-378.

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019a. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. ISPRS Journal of Photogrammetry and Remote Sensing 154, 189-201.

Wang, M., Su, Y., Yang, X., 2014. Spatial Distribution of Soil Organic Carbon and Its Influencing Factors in Desert Grasslands of the Hexi Corridor, Northwest China. PLOS ONE 9, e94652.

Wang, S., Adhikari, K., Wang, Q., Jin, X., Li, H., 2018c. 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. Ecological Indicators 84, 263-272.

Wang, S., Jin, X., Adhikari, K., Li, W., Yu, M., Bian, Z., Wang, Q., 2018d. Mapping total soil nitrogen from a site in northeastern China. CATENA 166, 134-146.

Wang, S., Wang, Q.B., Adhikari, K., Jia, S.H., Jin, X.X., Liu, H.B., 2016. Spatial-Temporal Changes of Soil Organic Carbon Content in Wafangdian, China. Sustainability 8, 16.

Wang, S., Zhuang, Q., Wang, Q., Jin, X., Han, C., 2017. Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. Geoderma 305, 250-263.

Wang, Y., Deng, L., Wu, G., Wang, K., Shangguan, Z., 2018e. Large-scale soil organic carbon mapping based on multivariate modelling: The case of grasslands on the Loess Plateau. Land Degradation & Development 29, 26-37.

Wang, Y., Wang, S., Adhikari, K., Wang, Q., Sui, Y., Xin, G., 2019b. Effect of cultivation history on soil organic carbon status of arable land in northeastern China. Geoderma 342, 55-64.

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators 52, 394-403.

Yang, L., Luo, P., Wen, L., Li, D., 2016a. Soil organic carbon accumulation during post-agricultural succession in a karst area, southwest China. Scientific Reports 6, 37118.

Yang, R.-M., Guo, W.-W., 2019a. Modelling of soil organic carbon and bulk density in invaded coastal wetlands using Sentinel-1 imagery. International Journal of Applied Earth Observation and Geoinformation 82, 101906.

Yang, R.-M., Guo, W.-W., 2019b. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. Environmental Monitoring and Assessment 191, 462.

Yang, R.-M., Guo, W.-W., Zheng, J.-B., 2019. Soil prediction for coastal wetlands following Spartina alterniflora invasion using Sentinel-1 imagery and structural equation modeling. CATENA 173, 465-470.

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

Yang, R.-M., Zhang, G.-L., Yang, F., Zhi, J.-J., Yang, F., Liu, F., Zhao, Y.-G., Li, D.-C., 2016c. Precise estimation of soil organic carbon stocks in the northeast Tibetan Plateau. Scientific Reports 6, 21842.

Yang, R., Rossiter, D.G., Liu, F., Lu, Y., Yang, F., Yang, F., Zhao, Y., Li, D., Zhang, G., 2015. Predictive Mapping of Topsoil Organic Carbon in an Alpine Environment Aided by Landsat TM. PLOS ONE 10, e0139042.

Yue, T.-X., Zhao, N., Ramsey, R.D., Wang, C.-L., Fan, Z.-M., Chen, C.-F., Lu, Y.-M., Li, B.-L., 2013. Climate change trend in China, with improved accuracy. Climatic Change 120, 137-151.

Zang, C.F., Liu, J., van der Velde, M., Kraxner, F., 2012. Assessment of spatial and temporal patterns of green and blue water flows under natural conditions in inland river basins in Northwest China. Hydrology and Earth System Sciences 16, 2859-2870.

Zhang, P., Shao, M.a., 2014. Spatial Variability and Stocks of Soil Organic Carbon in the Gobi Desert of Northwestern China. PLOS ONE 9, e93584.

Zhang, Y., Guo, L., Chen, Y., Shi, T., Luo, M., Ju, Q., Zhang, H., Wang, S., 2019. Prediction of Soil Organic Carbon based on Landsat 8 Monthly NDVI Data for the Jianghan Plain in Hubei Province, China. Remote Sensing 11, 1683.

Zhao, P., Lu, D., Wang, G., Liu, L., Li, D., Zhu, J., Yu, S., 2016. Forest aboveground biomass estimation in Zhejiang Province using the integration of Landsat TM and ALOS PALSAR data. International Journal of Applied Earth Observation and Geoinformation 53, 1-15.

Zhong, Z., Chen, Z., Xu, Y., Ren, C., Yang, G., Han, X., Ren, G., Feng, Y., 2018. Relationship between Soil Organic Carbon Stocks and Clay Content under Different Climatic Conditions in Central China. Forests 9, 598.

Zhou, T., Pan, J., Zhang, P., Wei, S., Han, T., 2017. Mapping Winter Wheat with Multi-Temporal SAR and Optical Images in an Urban Agricultural Region. Sensors 17, 1210.

Zhou, T., Zhao, M., Sun, C., Pan, J., 2018. Exploring the Impact of Seasonality on Urban Land-Cover Mapping Using Multi-Season Sentinel-1A and GF-1 WFV Images in a Subtropical Monsoon-Climate Region. ISPRS International Journal of Geo-Information 7, 3.