



High Spatial Resolution Topsoil Organic Matter Content Mapping Across Desertified Land in Northern China

Yang Junting^{1,2,3}, Li Xiaosong^{1,2*}, Wu Bo⁴, Wu Junjun², Sun Bin⁵, Yan Changzhen⁶ and Gao Zhihai⁵

¹International Research Center of Big Data for Sustainable Development Goals, Beijing, China, ²Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, ³College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, ⁴Institute of Desertification Studies, Chinese Academy of Forestry, Beijing, China, ⁵Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing, China, ⁶Key Laboratory of Desert and Desertification, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China

OPEN ACCESS

Edited by:

Claudio Zucca,
University of Sassari, Italy

Reviewed by:

Sharad Gupta,
Punjab Remote Sensing Centre
(PRSC), India

Ana B. Ruescas,
University of Valencia, Spain

*Correspondence:

Li Xiaosong
lixs@aircas.ac.cn

Specialty section:

This article was submitted to
Environmental Informatics
and Remote Sensing,
a section of the journal
Frontiers in Environmental Science

Received: 17 February 2021

Accepted: 02 August 2021

Published: 25 August 2021

Citation:

Junting Y, Xiaosong L, Bo W, Junjun W, Bin S, Changzhen Y and Zhihai G (2021) High Spatial Resolution Topsoil Organic Matter Content Mapping Across Desertified Land in Northern China. *Front. Environ. Sci.* 9:668912. doi: 10.3389/fenvs.2021.668912

Soil organic matter (SOM) content is an effective indicator of desertification; thus, monitoring its spatial–temporal changes on a large scale is important for combating desertification. However, mapping SOM content in desertified land is challenging owing to the heterogeneous landscape, relatively low SOM content and vegetation coverage. Here, we modeled the SOM content in topsoil (0–20 cm) of desertified land in northern China by employing a high spatial resolution dataset and machine learning methods, with an emphasis on quarterly green and non-photosynthetic vegetation information, based on the Google Earth Engine (GEE). The results show: 1) the machine learning model performed better than the traditional multiple linear regression model (MLR) for SOM content estimation, and the Random Forest (RF) model was more accurate than the Support Vector Machine (SVM) model; 2) the quarterly information regarding green vegetation and non-photosynthetic were identified as key covariates for estimating the SOM content in desertified land, and an obvious improvement could be observed after simultaneously combining the Dead Fuel Index (DFI) and Normalized Difference Vegetation Index (NDVI) of the four quarters (R^2 increased by 0.06, the root mean square error decreased by 0.05, the ratio of prediction deviation increased by 0.2, and the ratio of performance to interquartile distance increased by 0.5). In particular, the effects of the DFI in Q1 (the first quarter) and Q2 (the second quarter) on estimating low SOM content (<1%) were identified; finally, a timely (2019) and high spatial resolution (30 m) SOM content map for the desertified land in northern China was drawn which shows obvious advantages over existing SOM products, thus providing key data support for monitoring and combating desertification.

Keywords: desertified land, soil organic matter content, Sentinel-2, machine learning, Google Earth Engine

INTRODUCTION

Desertification is defined as land degradation in drylands resulting from various factors, including climatic variations and human activities (UNCCD, 1994), which has a major effects on carbon emissions through the loss of soil organic matter (SOM) when soil deplete from their original state (MEA – Millennium Ecosystem Assessment, 2005). China is among the countries most severely affected by desertification, representing 17.93% of its total land area (State Forestry Administration of China, 2015). Thus, desertification in China is likely to considerably impact the global CO₂ status (Feng et al., 2004; An et al., 2019). Accurate and timely SOM content mapping is of great importance for assessing desertification (Sims et al., 2019). However, few researches have focused on the large scale SOM content estimation of the desertified land of Northern China (Li et al., 2018; Sun et al., 2019).

Soil sampling is the traditional method for determining the regional SOM content. Although the method yields more accurate results, it requires considerable effort, material resources, and time, especially for large areas (Wang et al., 2017; Hamzehpour et al., 2019; Lee et al., 2019; Yang et al., 2020). Due to its high-cost soil sampling is not practical for frequent or real-time SOC monitoring. Digital Soil Mapping (DSM) (McBratney et al., 2003) is efficient and convenient for acquiring soil information based on mathematical or statistical relationships between field soil observations and related predictor variables (e.g., climate, vegetation, relief, parent material, and time) (Zhao et al., 2014; Liang et al., 2019b; Keskin et al., 2019).

Remote sensing has facilitated the progress of DSM, mainly by providing up-to-date information on land cover and natural vegetation status, and is often replaced by vegetation indices (Lamichhane et al., 2019; Wiesmeier et al., 2019); for example, Normalized Difference Vegetation Index (NDVI) is the most commonly used (Plaza et al., 2018; Wang et al., 2018; Zhang and Huisingh, 2018; Xu et al., 2019; Zeraatpisheh et al., 2019; Yang et al., 2020). However, NDVI is constrained when representing vegetation coverage in semi-arid and arid areas due to the influence of exposed soil, dead vegetation and litter on the spectral response (Maynard and Levi, 2017). While incorporating non-green vegetation information in different seasons could help alleviate this kinds of problem in soil organic carbon prediction (Wang et al., 2018).

Machine learning methods, such as random forest (RF), support vector machine (SVM), enhanced regression trees (ERT), and co-kriging methods, have impacted the success of soil attribute estimations in recent years (Were et al., 2015; Keskin et al., 2019; Lamichhane et al., 2019; Wang et al., 2019; Zeraatpisheh et al., 2019). RF technology has become the most popular machine learning method and can help avoid model overfitting (Zhang et al., 2017; Zeraatpisheh et al., 2019; Yang et al., 2020) and support variable importance evaluation (Ballabio, 2009; Were et al., 2015; Zhi et al., 2018). In addition, the SVM model (Cortes and Vapnik, 1995), based on the principle of structural risk minimization, can avoid the dependence on data scale and data distribution; therefore, it has better learning capabilities and smaller prediction errors (Ballabio, 2009;

Were et al., 2015). Many studies have compared the accuracy of different machine learning models; however, the conclusions are not uniform (Brungard et al., 2015; Liang et al., 2019a; Hamzehpour et al., 2019). For example, Wu et al. (2016) compared Mahalanobis Distance (MD), Maximum Likelihood (ML), Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests (RFs) for land degradation mapping, and found ML classification is a suitable candidate. Were et al. (2015) found SVR to be the best predictor of SOC, while Wu et al. (2018) showed that the RF model is superior to other models such as SVM. Nevertheless, there is a consensus that machine learning is more suitable for soil attribute estimation than traditional multiple regression (Chagas et al., 2016; Zeraatpisheh et al., 2019).

The relatively coarse resolution (250–1,000 m) of existing SOM products [e.g., the Harmonized World Soil database (WIEDER, 2014) with 1000-m resolution and Soil Grids (Hengl et al., 2017) with a higher 250-m resolution] is another limitation for the desertified land in northern China, in which high spatial heterogeneity exist and directly affect the spatial and temporal distribution of topsoil organic matter (Tongway and Ludwig, 2007). With the open access of the multi-spectral satellites (e.g., Landsat 8 and Sentinel-2), researches on SOC estimation with high spatial resolution were gradually carried out; for example, Wang et al. used Landsat 5 TM data to estimate SOC and nitrogen in Liaoning Province, China (Wang et al., 2017); Forkuor et al. (2017) found that Landsat eight data provides acceptable SOC digital map of south-western Burkina Faso, while in the study of Castaldi et al. (2019), Sentinel-2 data was sufficient for describing the variability of SOC. However, the majority of previous studies have only focused on the field or local scales, mainly due to the limitation associated with the processing of extensive high spatial resolution data. In recent years, the emergence of cloud platforms (such as GEE), characterized by massive amounts of available high spatial resolution datasets and powerful computing capabilities, has aided in collecting multi-temporal, high spatial resolution vegetation information for SOM modeling (Dong et al., 2016; Gorelick et al., 2017; Beaton et al., 2019).

The main purpose of this study was to estimate the high spatial resolution topsoil SOM content in the desertified land of northern China based on the GEE platform by employing Sentinel-2 data, field-measured SOM samples, and ancillary factors with advanced machine learning methods (SVM and RF) and traditional statistical methods (MLR). In particular, the potential of high spatial and quarterly green and non-photosynthetic information generated from Sentinel-2 data was explored.

MATERIAL AND METHODS

Study Area

The extent of the desertified land was from the “Atlas of Desert and Aeolian Desertification in Northern China” which was produced through visual interpretation of Landsat images and has been adopted by many studies (Wang et al., 2011). The

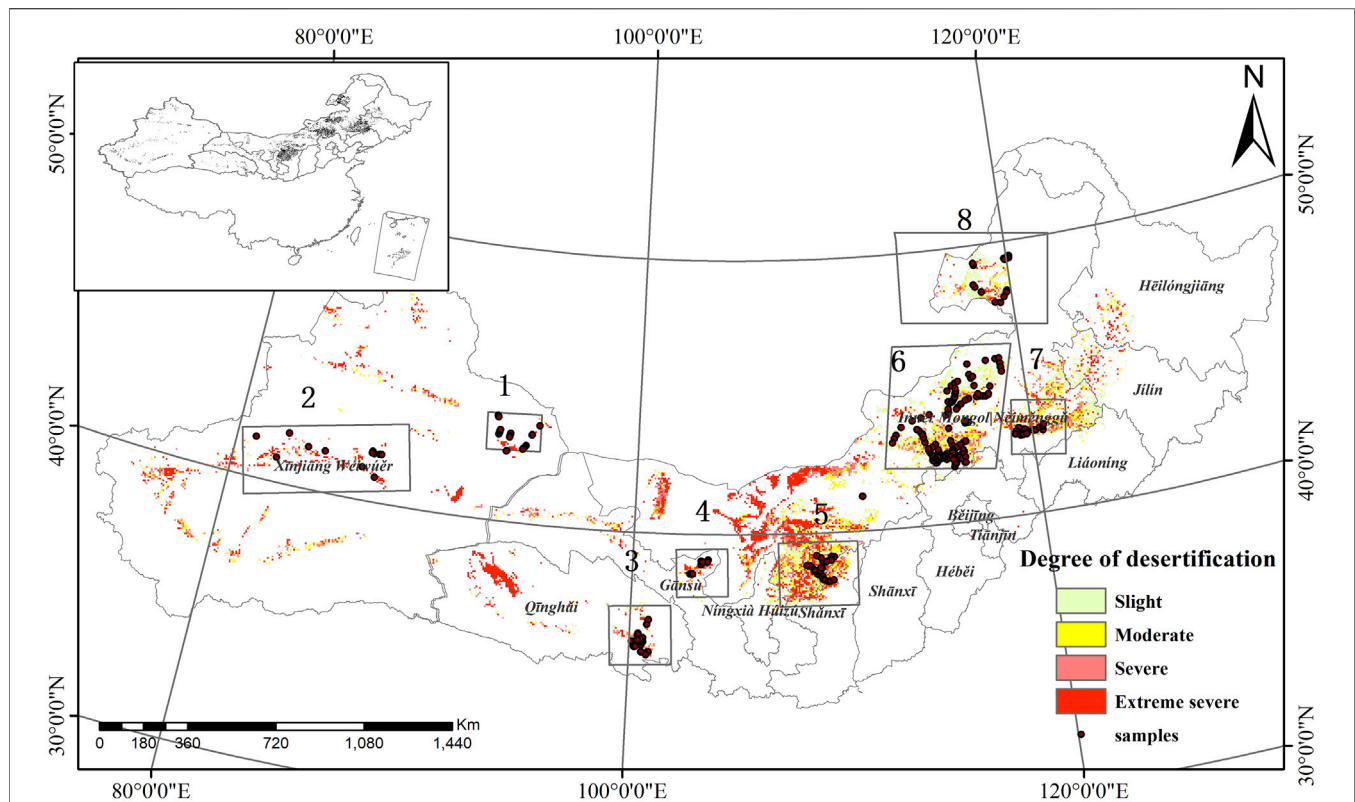


FIGURE 1 | Study area range and sampling locations (1: Gurbantunggut Desert, 2: Taklimakan Desert, 3: Gonghe alpine desertified grassland, 4: Minqin Oasis-Desert transition belt, 5: Mu Us Sandy Land, 6: Otindag Sandy Land, 7: Horqin Sandy Land, 8: Hulunbuir Sandy Land).

product includes multi-temporal results from 1975–2010 with 4° of desertification: slight, moderate, severe, and extremely severe. In this study, desertification monitoring results from 2010 were finally utilized (Figure 1). The total area is 375,900 km² and the climate difference is significant in the area, which spans arid, semi-arid, and sub-humid climate regions (Tao, 2014). Natural vegetation types include woodlands, shrublands, grasslands, and barren land from east to west (Zhang and Huisingh, 2018). The major soil types are Gray-brown desert soils (WRB reference soil group: Cambisols), brown calcic soil (WRB reference soil group: Cambisols), and aeolian sandy soil (WRB reference soil group: Arenosols). All desertification processes generally lead to an obvious reduction of SOM in topsoil; thus, the SOM content of the study area was lower than that in non-desertified areas (Allington and Valone, 2010; Tao, 2014).

Soil Sample Collection and Chemical Analysis

Based on the information on the type and extent of desertified land, and considering elevation, climate, and soil type, soil samples were collected from July to August 2017–2019. The sample collection was mainly concentrated in eight representative areas: Mu Us Sandy Land, Otindag Sandy Land, Horqin Sandy Land, Hulunbuir Sandy Land, Gonghe alpine desertified grassland, Minqin Oasis-Desert transition belt, the

northern edge of the Taklimakan Desert, and the southern margin of the Gurbantunggut Desert (Figure 1). The total number of samples was 243 (Table 1). The size of the sampled square is 30 m*30 m, the soil profile depth was 0–20 cm and then nine 10*10 m grids are laid out in the square. Take one soil sample at the center of each grid, and get nine soil samples. The nine soil samples were thoroughly mixed and analyzed in the laboratory, and the SOM content was obtained by the analysis. At the same time, a handheld GPS (Trimble GPS, precision <1 m) recorder was used to record the geographic coordinates of the center point of the plot.

Soil samples were stored in a sealed package for SOM analysis in the laboratory. Subsequently, the bulk samples were air-dried and sieved to retrieve the fine earth fraction (<2 mm) (O’Kelly, 2004). SOM was measured by the dichromate oxidation method as described by the Agriculture Chemistry Speciality Council, Soil Science Society of China (Agriculture Chemistry Speciality Council, 1983).

Predictor Variables

Jenny (1941) defined soil development as a function of climate, organic matter/organic activity, relief, parent materials and time. Later, McBratney et al. (2003) proposed digital soil mapping based on this, using the SCORPAN equation as a framework to quantitatively estimate soil properties through soil samples and environmental covariates. The latest SOC digital mapping review

TABLE 1 | Distribution of sample points.

	Land cover				Stages of desertification				Total
	Barren land	Shrublands	Grassland	Woodland	Slight	Moderate	Severe	Extremely severe	
Gurbantunggut Desert	15	0	7	1	3	1	3	15	23
Otindag Sandy Land	3	4	61	4	35	20	12	5	72
Taklimakan Desert	14	0	0	0	0	0	0	14	14
Minqin	0	18	0	2	3	5	6	6	20
Gonghe alpine	0	4	17	0	14	2	5	0	21
Mu Us Sandy Land	0	33	0	1	9	15	6	3	33
Horqin Sandy Land	1	33	5	5	1	3	6	24	34
Hulunbuir Sandy Land	0	7	14	5	15	8	3	0	26

article shows that covariates representing organism/organic activities were among the most frequent among top five covariates, followed by the variables representing climate and topography (Lamichhane et al., 2019).

Organism/organic activities including land cover and natural vegetation, but real-time vegetation cover information is difficult to obtain. Vegetation index (Mahmoudabadi et al., 2017; Guo et al., 2020) and land cover or land use data are often used as vegetation indicators (Meersmans et al., 2008; Li et al., 2013). Especially, NDVI (Tucker, 1979) has been widely used in the estimation of SOC content, and it has been found to have a significant positive correlation with soil carbon storage (Wang et al., 2019), and in several soil carbon content estimation studies, NDVI is important predictors (Guo et al., 2020; Mahmoudabadi et al., 2017; Wang et al., 2020, 2017). At present, studies using NDVI for DSM (digital soil mapping), the acquisition of NDVI is mostly by an annual mean or alternatively by a single image acquired during a specific time of year (the peak of the growing season or the dry season, etc.). However, more and more studies support the view that the temporal variability of vegetation spectra driven by soil and climate feedback is an important predictor of soil variability (Ballabio et al., 2012; Maynard and Levi, 2017). Guo et al. (2020) used three NDVI images in different periods to successfully predict the SOC content at the field scale. Clark (2017) found that using NDVI in different seasons at the same time improved the accuracy of land cover classification.

In studies of digital mapping of SOC/SOM, the variable representing the “C” factor is the second most common (Lamichhane et al., 2019). Climate is the main controlling factor for SOC storage from a regional to global scale (Lamichhane et al., 2019; McBratney et al., 2003; Plaza et al., 2018; Wiesmeier et al., 2019). Therefore, the average annual temperature and annual rainfall are used as predictors for estimating SOC content in many studies (Liang et al., 2019b; Schillaci et al., 2017). With the development of remote sensing technology, climate indicators such as surface temperature and surface evapotranspiration have also been incorporated into the SOC content prediction index system (Sayão et al., 2018; Zhao et al., 2014).

Topography plays a vital role in the spatial distribution of SOC, because topography affects precipitation, water flow paths,

stagnant water and discharge, and has a significant impact on the erosion process. Based on the terrain index extracted from DEM data, elevation, slope, aspect, TWI... are all commonly used predictors in SOC prediction (Wang et al., 2017; Zeraatpisheh et al., 2019).

We also selected 14 indicators that can represent climate, organism, and topography to predict the soil organic matter content in the study area (Table 2), included two climatic factors: 10-years average annual temperature (MAT) and annual average precipitation (MAP); eight vegetation variables (NDVI and DFI of four quarters); three topographic factors (elevation, slope, aspect) and land-cover type (LC). The data sources of all variables were directly used online on the GEE platform, on which all related calculations were also completed. The unified projection was WGS 84 and the spatial resolution was 30 m. Through the reproject function {ee.Image().reproject[{crs:'EPSG:4326', scale: 30}]} of the GEE platform. The method of resampling is nearest neighbor resampling, the unified projection was WGS 84 and the spatial resolution was 30 m.

In the selection of vegetation index, we considered both photosynthetic vegetation information and non-photosynthetic vegetation information, so we chose NDVI and DFI two vegetation indexes. DFI is the band index, which has good potential for estimating dead fuel (DF) coverage in steppe areas (Cao et al., 2010). DF is also called non-photosynthetic vegetation (NPV) or plant litter, and includes dry/cured/dead grasses, dead foliage and twigs, dead wood, slash wood, litter and duff, as well as crop residue (Cao et al., 2010). In addition, we calculated the vegetation index of the four quarters within a year for comparative experiments to study whether the multi-temporal vegetation index information can improve the estimation accuracy of soil carbon content.

The selected remote sensing image source was Sentinel-2 data, and the spatial resolution of the bands differed. The highest spatial resolution was 10 m and the revisit period was 5 days. The image was obtained from the Sentinel-2 surface reflectance dataset (GEE ID: COPERNICUS/S2_SR) from the GEE. The period from 1 January to 31 December 2019 was used to ensure data quality by incorporating more images into the analysis. After determining the time and range, the MASK function provided by the GEE platform was first used to filter

TABLE 2 | Factors used in the model.

Theme	Name	Source	Resolution	Reference/Data source
Vegetation indices	NDVI, DFI	Sentinel-2	10 m	https://sentinel.esa.int/web/sentinel/missions/sentinel-2
Climate factors	Mean annual temperature (°C)	ERA5 Daily	0.25°	Copernicus Climate Change Service (2017)
	Mean annual precipitation (mm)		0.25°	
Topographic factors	Elevation	SRTM90_V4	90 m	Jarvis et al. (2008)
	Slope			
	Aspect			
Land cover information	Land cover	MCD12Q1 V6	500 m	Friedl and Sulla-Menashe (2019)

the images with less than 20% cloud cover. We calculated the vegetation indexes on the cloudless images, established a time series according to the quarters, and took the max value of the vegetation index of the image set in each quarter as a covariate. That means a year's Sentinel-2 data is divided into four sets. Every 3 months is a set (January to March, April to June...), and four quarters' NDVI and DFI were calculated.

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$DFI = 100 * \frac{1 - \frac{SWIR2}{SWIR1}}{\frac{Red}{NIR}} \quad (2)$$

Here, NIR, Red, SWIR2, and SWIR1 are the reflectance values of the Sentinel-2 image in the near-infrared, red, and two short wave infrared bands, respectively, and correspond to bands 8, 4, 11, and 12 of Sentinel-2 data, respectively.

In this study, MAT and MAP were selected as climate factors. The MAT and MAP sources were based on the ERA5 Daily aggregates dataset (GEE ID: ECMWF/ERA5/DAILY), obtained by the mean value function [return.mean()] for the GEE. The data was collected from 2010 to 2019, the spatial resolution was 0.25° arc degrees, and the temporal resolution was 1 day.

The altitude factor was derived from the 90-m resolution SRTM90 (GEE ID: CGIAR/SRTM90_V4) topographic data in the GEE dataset. Slope and aspect were obtained by running a terrain analysis [ee.Algorithms.Terrain()] on the GEE platform.

The land-cover data source was the MCD12Q1 V6 product (GEE ID: MODIS/006/MCD12Q1), which provides global land-cover types at yearly intervals (2001–2016) derived from six different classification schemes. The type of classification scheme we selected was the Land-Cover Type 1: Annual International Geosphere-Biosphere Programme (IGBP) classification.

The high-resolution variation of SOM was acquired through the utilization of Sentinel-2 based VIs, namely NDVI and DFI. Except these high-resolution predictors, we do utilize some coarse resolution predictors, mainly due to the dataset availability in the GEE platform. However, these coarse resolution variables are meaningful for estimating the SOM level in the coarse-scale, and when combined with high-spatial resolution predictors, could generate a more detailed SOM map.

Modelling Methodology

The relationship between SOM and predictor variables was analyzed using correlation analysis. To explore the potential of

quarterly green and dry vegetation information, we divided all predictor variables into three groups: 1) NDVI (July–September, corresponding to maximum green vegetation coverage), topography, climate, and land-cover types (covariate set A); 2) NDVI (four quarters), topography, climate, and land-cover types (covariate set B); and 3) the DFI (four quarters) was added for covariate set B (covariate set C).

Three modeling techniques were selected for SOM prediction: Multiple linear regression (MLR), RF, and SVM. Two machine learning methods (i.e., RF and SVM) can model the complex nonlinear relationship between variables and organic matter and are considered adept at predicting SOM in various regions (Were et al., 2015; Zhou et al., 2020). Moreover, RF can also sort the importance of variables by the change in the error outside the bag (Zeraatpisheh et al., 2019; Zhou et al., 2020). As a simple linear model, MLR was employed as an evaluation benchmark to compare the performances of machine learning and traditional statistical models.

RF is an ensemble learning method based on decision trees (Breiman, 2001). Re-sampling by the bootstrap method is used to randomly select observations in the dataset, using the principle of minimum error to select the optimal segmentation feature and the optimal segmentation point. The segmentation point (value) divides the data into two parts, determines the corresponding output value, and finally generates a regression decision tree after multiple divisions (Wang et al., 2018). The combination of trees generated by multiple choices constitutes an RF, and the final value of the predictor variable is the linear average of the prediction results of multiple trees (Ballabio, 2009; Were et al., 2015; Zhi et al., 2018). Based on the stability of the RF model, the excellent model performance, and the ability to evaluate the importance of variables, RFs have been widely used in various classification and regression problems (Wang et al., 2018; Keskin et al., 2019; Yang et al., 2020). On the GEE platform RF, only two parameters need to be set to generate a predictive model: 1) the number of trees (n tree); and 2) the maximum number of leaf nodes in each tree (maximum nodes). To realize these two model parameters, we determined the final value of the parameters by setting the step size to traverse the parameter combination within a suitable range in the GEE platform. For the RF model, we set the number of trees (n tree) in GEE within the range of 10–10,000, and the maximum tree depth 5–50 to conduct the experiment and select the most suitable model parameters. The final model parameters are shown in **Table 3**:

SVM is a machine learning method developed by Cortes and Vapnik (1995) based on the principle of structural risk

TABLE 3 | Parameters of RF.

Parameters	Covariate set A	Covariate set B	Covariate set C
<i>n</i> tree	1,000	1,000	1,000
max_depth	10	10	30

TABLE 4 | Parameters of SVM.

Parameters	Covariate set A	Covariate set B	Covariate set C
C	32	8	1,024
P	0.25	0.0625	0.05
G	0.5	0.03125	0.0625

minimization. It is specifically proposed for small-sample learning problems. The use of kernel functions solves the problem of dimensionality, which is suitable for dealing with nonlinear problems. In addition, the structural risk minimization principle enables the SVM to have a very good generalization ability (Were et al., 2015; Zhi et al., 2018). The hyperplane is used to optimally divide all data into different classes, which results in acceptable learning ability and smaller prediction error (Ballabio, 2009). In this study, the realization of an SVM algorithm was completed in the GEE platform. For the optimal parameter setting of the SVM models, the kernel function of SVM used the widely recognized radial-basis kernel (RBF) (Wang et al., 2018), the cost factor C, the minimum loss function P, and the gamma coefficient G were determined using the LIBSVM toolbox (Chang and Lin, 2011). In the process of parameter setting, the root mean square error (RMSE) of different parameter combinations was measured by 5-fold cross-validation. Finally, the parameters are determined as shown in **Table 4**:

The MLR model was used to establish a linear equation between a dependent variable Y and multiple independent variables (X1, X2... Xn), and the least-squares method was used to estimate the parameters of each independent variable under the condition of minimizing the sum of the square of errors. It is a commonly used early modeling method for regression problems. Because the method is simple and provides accurate estimations, it has been used in many studies (Meersmans et al., 2008; Besalatpour et al., 2013; Zhang et al., 2017). Although MLR is relatively simple, in many studies, it is well fitted to the relationship between variables. Furthermore, when evaluating the superiority of a machine learning model, a linear model is often used as a benchmark. Therefore, we selected MLR to estimate the SOM content due to its robust and widespread applications (Chagas et al., 2016; Lee et al., 2019).

Model Performance Assessment

Here we split the data into five groups, and use each of them in turn to evaluate the model fit on the other 4/5 of the data. The determination coefficient (R^2), RMSE, ratio of prediction deviation (RPD), and the ratio of performance to interquartile distance (RPIQ) were selected as indicators to measure the model accuracy.

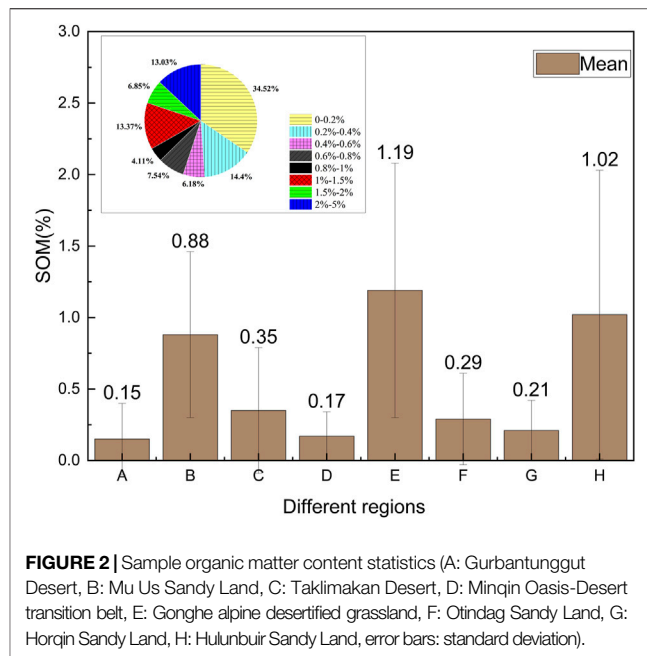


FIGURE 2 | Sample organic matter content statistics (A: Gurbantunggut Desert, B: Mu Us Sandy Land, C: Taklimakan Desert, D: Minqin Oasis-Desert transition belt, E: Gonghe alpine desertified grassland, F: Otindag Sandy Land, G: Horqin Sandy Land, H: Hulunbuir Sandy Land, error bars: standard deviation).

$$R^2 = \left(\frac{\sum (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \right)^2 \tag{3}$$

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

$$RPD = \frac{SD}{RMSE} \tag{5}$$

$$RPIQ = \frac{IQ}{RMSE} \tag{6}$$

In Eqs. 3–6, P_i and O_i are the predicted and observed SOM content, respectively; n is the number of samples; \bar{P} and \bar{O} are the means of the predicted and observed SOM content, respectively; SD is the standard deviation of measured values; IQ is the interquartile range ($IQ = Q3 - Q1$) of the measured SOM in the validation dataset; and $Q1$ and $Q3$ are the first and third quarters of the SOM content in the validation dataset, respectively. In general, a good model prediction should correspond to high R^2 , RPD, and RPIQ values, and low RMSE values. In particular, the model classification criteria adopted in this study were based on RPD values, which were divided into five model classes: Very good ($RPD = 2.5 - 2.0$), good ($RPD = 2.0 - 1.8$), fair ($RPD = 1.8 - 1.4$), poor ($RPD = 1.4 - 1.0$), and very poor ($RPD < 1.0$) (Viscarra Rossel et al., 2006; Wan et al., 2019).

RESULTS

SOM Content Distribution

Statistical analysis of the soil sample points showed that the maximum SOM content in the study area was 3.8%, the

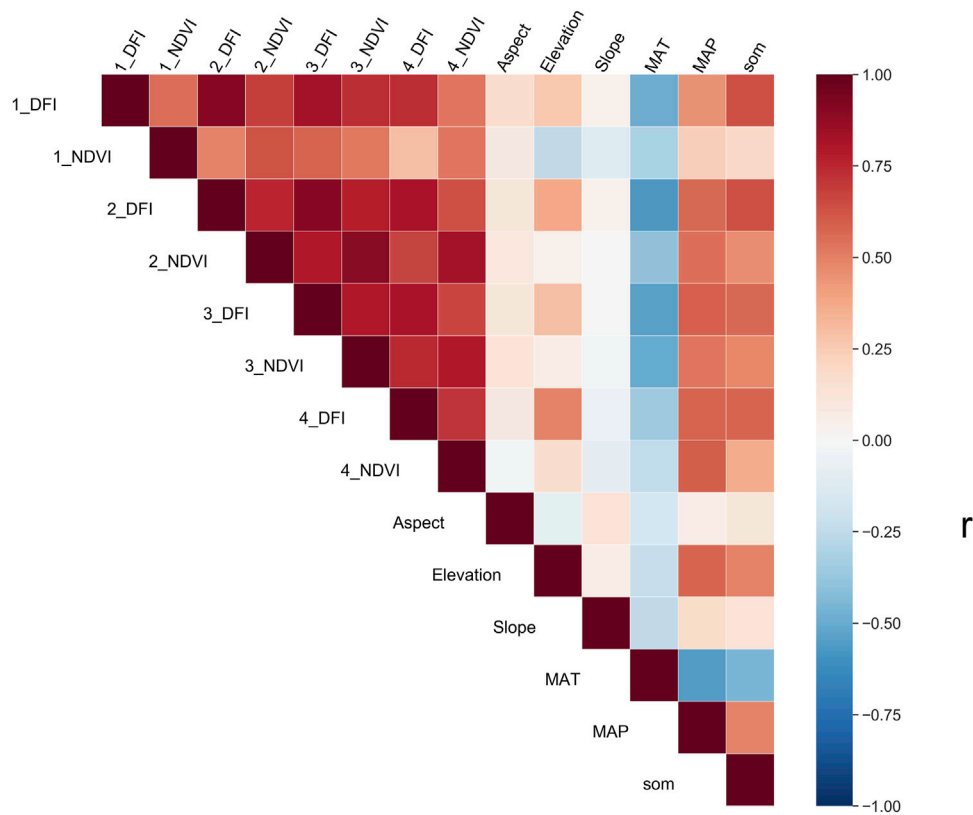


FIGURE 3 | Correlations between variables, indicating that the variables are significantly correlated at the 0.01 level. MAT: Mean annual temperature; MAP: Mean annual precipitation; 1_NdVI: Normalized difference vegetation index (NDVI) in the first quarter; 2_NdVI: NDVI in the second quarter; 3_NdVI: NDVI in the third quarter; 4_NdVI: NDVI in the fourth quarter; 1_DFI: Dead fuel index (DFI) in the first quarter; 2_DFI: DFI in the second quarter; 3_DFI: DFI in the third quarter; 4_DFI: DFI in the fourth quarter; SOM: Soil organic matter content.

minimum was 0.001%, the mean was 0.559%, the standard deviation was 0.749%, and the coefficient of variation was 0.334. From the distribution map (Figure 2), it can be seen that the overall SOM content (0–20 cm) was low (i.e., <1% in most areas), which is consistent with the overall characteristics of the SOM in desertified land. Among the 243 samples, 34.36% had a SOM content <0.2%, whereas only 13.03% of samples had a SOM content >2%. According to the different sampling regions, the SOM content in the Gonghe alpine desertified grassland, Mu Us Sandy Land, and Hulunbuir Sandy Land were significantly higher than in other areas.

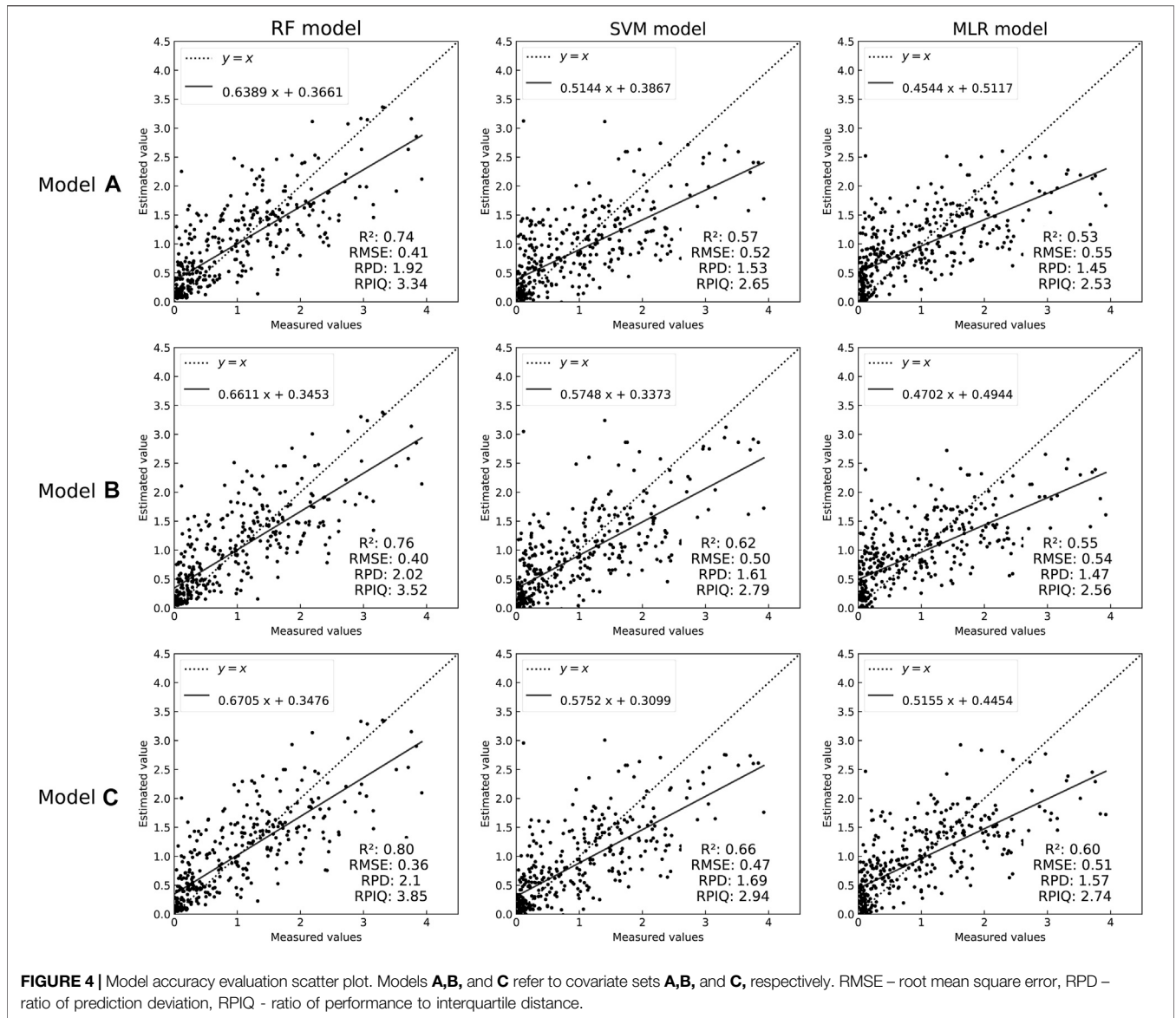
Correlations Between Single Input Factors and SOM Content

The results of the correlation analysis between the single input factors and the SOM content are shown in Figure 3. SOM content was positively correlated with mean annual precipitation ($r = 0.48$) and elevation ($r = 0.48$), and negatively correlated with mean annual temperature ($r = -0.45$). Perhaps more worthy of attention was the very high correlation between SOM and the quarterly vegetation index. Not unexpectedly, NDVI was positively correlated with SOM, and the correlation of NDVI in the third quarter ($r = 0.47$) was

higher than that in other quarters. The performance of DFI was yet unexpected. The DFI in the four quarters showed a good correlation with SOM ($r > 0.5$), especially in the first quarter ($r = 0.64$).

In the study area, water is the major limiting factor of plant production, thus the increase of precipitation could contribute to the accumulation of SOM (Plaza et al., 2018). An increase in temperature will lead to accelerated decomposition of organic matter in the soil and thus decrease the SOM reserves. Additionally, the positive relationship between elevation and SOM is mainly due to the influence of precipitation and temperature. As the elevation increases, precipitation increases and temperature decreases (McBratney et al., 2003; Plaza et al., 2018; Lamichhane et al., 2019).

Topography will affect the water temperature conditions and the distribution of soil-forming materials (Jenny, 1941). Among all topography variables, the correlation between elevation and SOM reached 0.48, slope reached 0.15, and aspect correlation reached 0.12. Because elevation and slope, slope direction often affects the redistribution of surface water and heat conditions, the water accumulation, and the decomposition rate of surface litter, as well as reflects the soil erosion and deposition process, which indirectly affects the change of soil organic matter content (Plaza et al., 2018).



Model Validation and Comparison

Figure 4 and Table 5 show the performance of the three modeling techniques (RF, SVM, and MLR) under different combinations of covariates. According to the RPD value of the model, the RF-B and RF-C models (RPD = 2.0, RPD = 2.1) can be evaluated as being very good models. The three models of the SVM method can be considered fair, and the three linear models are considered very poor.

Comparing the changes in the evaluation index values of each model, we found that the accuracy of the model has a close effect on the selection of the modeling method and the parameter combination. With the addition of the quarterly NDVI and introduction of the quarterly DFI index, the model effect gradually improved. After adding the quarterly information of NDVI, the R² in the RF-B model increased by 0.2, the RMSE decreased by 0.01, the R² in the SVM-B model increased by 0.5, and the RMSE decreased by 0.02. After simultaneously

combining the DFI and NDVI in the four seasons quarterly as variables, the model accuracy was more obviously improved. The RF model R² increased by 0.06, the RMSE decreased by 0.05, RPD increased by 0.2, and RPIQ increased by 0.5. The SVM model R² increased by 0.09, the RMSE decreased by 0.05, RPD increased by 0.16, RPIQ increased by 0.29. The MLR model R² increased by 0.07, RMSE decreased by 0.04, RPD increased by 0.12, and RPIQ increased by 0.21. The results further show the importance of multi-temporal data and dry vegetation on the surface of arid areas and demonstrate that the combination of all predictors (covariate set C) results in the best predictive performance.

Compared with the performance of the three models, the two machine learning methods showed better estimation ability than the linear regression model. This is similar to the complex nonlinear relationship between soil attributes and environmental factors proved by many soil attribute estimation studies, which are more suitable for fitting using a

TABLE 5 | Performance results of RF, SVM, and MLR in predicting SOM based on different combinations of covariates. The bold values are the precision of the optimal model.

Modelling methodology	Covariate set	R ²	RMSE	RPD	RPIQ
RF	A	0.74	0.41	1.92	3.34
	B	0.76	0.40	2.02	3.35
	C	0.80	0.36	2.1	3.85
SVM	A	0.57	0.52	1.53	2.65
	B	0.62	0.50	1.61	2.79
	C	0.66	0.47	1.69	2.94
MLR	A	0.53	0.55	1.45	2.53
	B	0.55	0.54	1.47	2.56
	C	0.60	0.51	1.57	2.74

machine learning method (Besalatpour et al., 2013). For the two machine learning methods, the RF model performed better than the SVM model under different parameter combinations. The three modeling methods consistently showed the best effect of model C, which also shows that the addition of the DFI index and multi-temporal vegetation index can effectively improve the model effect.

Variable Importance

For the SOM mapping using the model RF, the ranking of the predictors sorted by relative importance is shown in Figure 5 (note: Importance was converted to a percentage). The importance of the variables in models A, B, and C differed slightly. In covariate set A, the importance of MAT reached 39%, elevation reached 26%, and NDVI in the third quarter (Q3) reached 13%; in covariate set B, MAT and elevation were the two most important variables, followed by the NDVI in the first quarter (Q1); in covariate set C, DFIs in Q1 and the second quarter (Q2) (importance of 23 and 23%, respectively) were important variables for predicting the spatial distribution of SOC, followed by elevation, MAP, and MAT. This may be due to the mutual influence between variables, and models with different combinations of variables had shown different importance.

Spatial Prediction

According to different model methods (RF, SVM, and MLR) and three sets of parameter combinations (covariate sets A, B, and C), the SOM maps for the study area were produced with a 30 m resolution based on the GEE platform (Figure 6). The results showed a distribution trend of high SOM in the east and low SOM

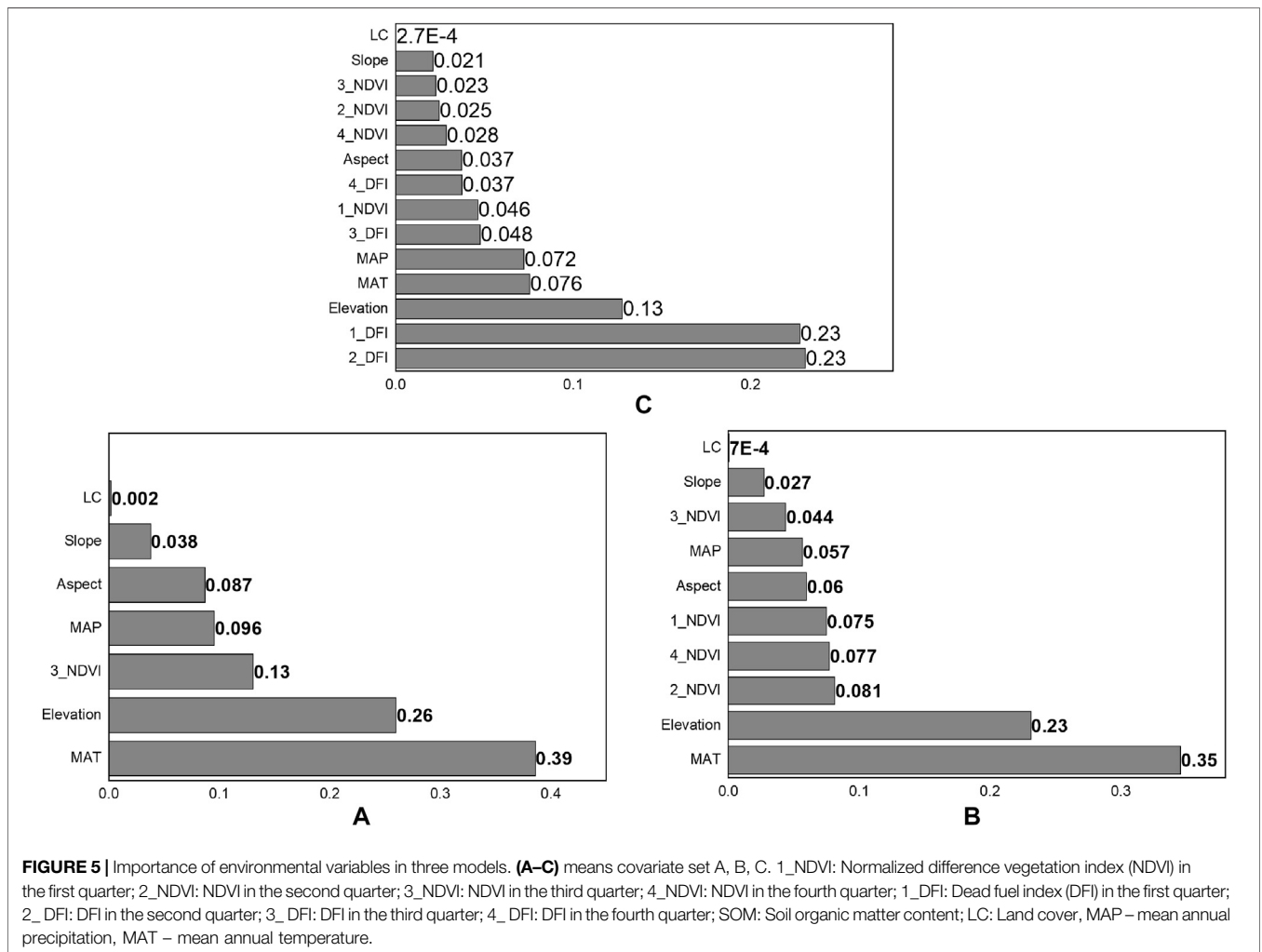


FIGURE 5 | Importance of environmental variables in three models. (A–C) means covariate set A, B, C. 1_NDVI: Normalized difference vegetation index (NDVI) in the first quarter; 2_NDVI: NDVI in the second quarter; 3_NDVI: NDVI in the third quarter; 4_NDVI: NDVI in the fourth quarter; 1_DFI: Dead fuel index (DFI) in the first quarter; 2_DFI: DFI in the second quarter; 3_DFI: DFI in the third quarter; 4_DFI: DFI in the fourth quarter; SOM: Soil organic matter content; LC: Land cover, MAP – mean annual precipitation, MAT – mean annual temperature.

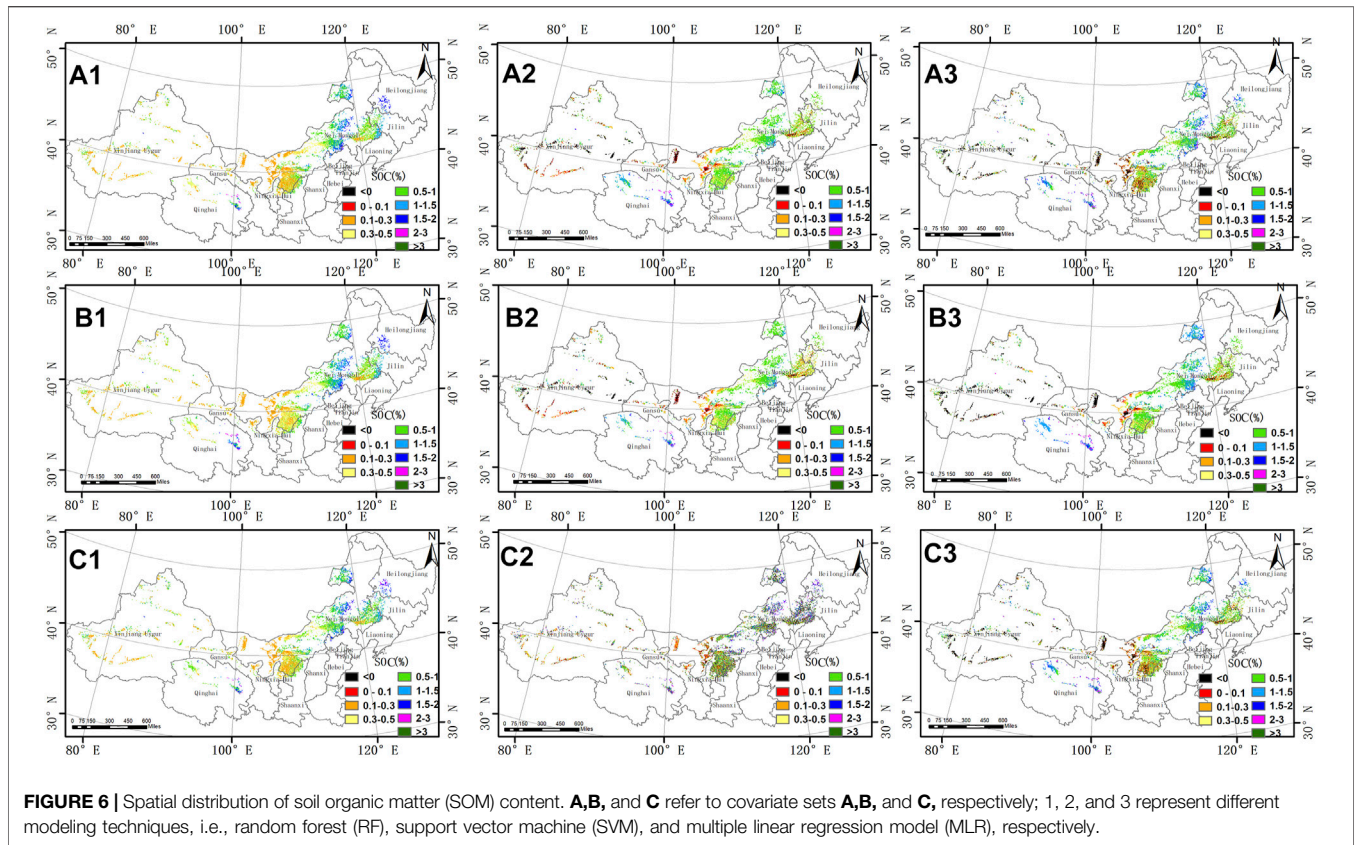


FIGURE 6 | Spatial distribution of soil organic matter (SOM) content. **A,B,** and **C** refer to covariate sets **A,B,** and **C,** respectively; 1, 2, and 3 represent different modeling techniques, i.e., random forest (RF), support vector machine (SVM), and multiple linear regression model (MLR), respectively.

in the west, and high-value areas were mainly distributed in the farming-pastoral transition zone near the Hunshandake Sandy Land, Hulunbuir Sandy Land, and Qinghai Gonghe Alpine. The areas with lower values were located near the Mu Us Sandy Land and the marginal and central part of the Gurbantunggut Desert in the northwest. This can be partly explained by the land-cover type and vegetation coverage.

However, the difference between the three modeling methods was also obvious (Figure 6). RF produced more high-value regions (SOM >1.5%) and provided a better simulation of the low-value regions on the edge of the Gurbantunggut Desert. The spatial differences in the organic matter maps brought about by the three different combinations of variables were reflected in the Mu Us Sandy Land and the desertified area in northwest Qinghai. For example, in the desertified land northwest of Qinghai, the SOM contents determined using the RF model were 0.1–0.3% for group A, 0.3–0.5% for group B, and 0.5–1% for group C.

Statistics of the SOM content (i.e., results of RF model C) were analyzed according to different desertification degrees, and the results are shown in Figure 7. There was a significant negative correlation between the degree of land desertification and the SOM content. The maximum and minimum SOM contents corresponding to different degrees of desertification exhibited no obvious regularity, mainly because desertified land is generally a mixture of vegetation and bare soil in different proportions, while the SOM content prediction is based on the pixel scale.

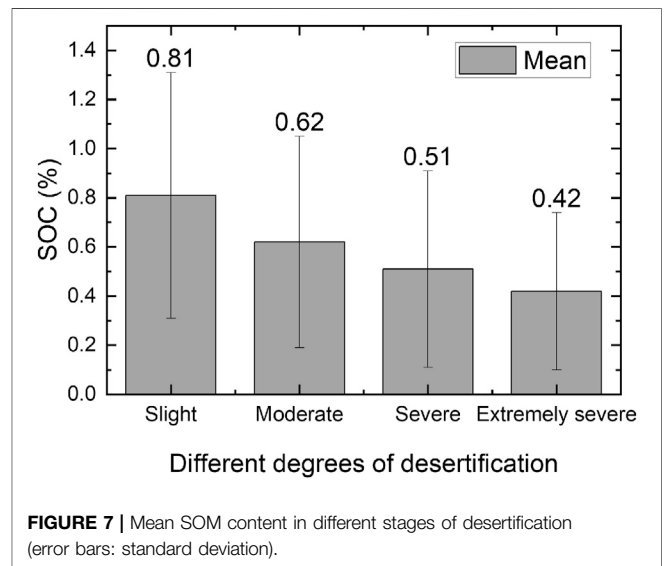
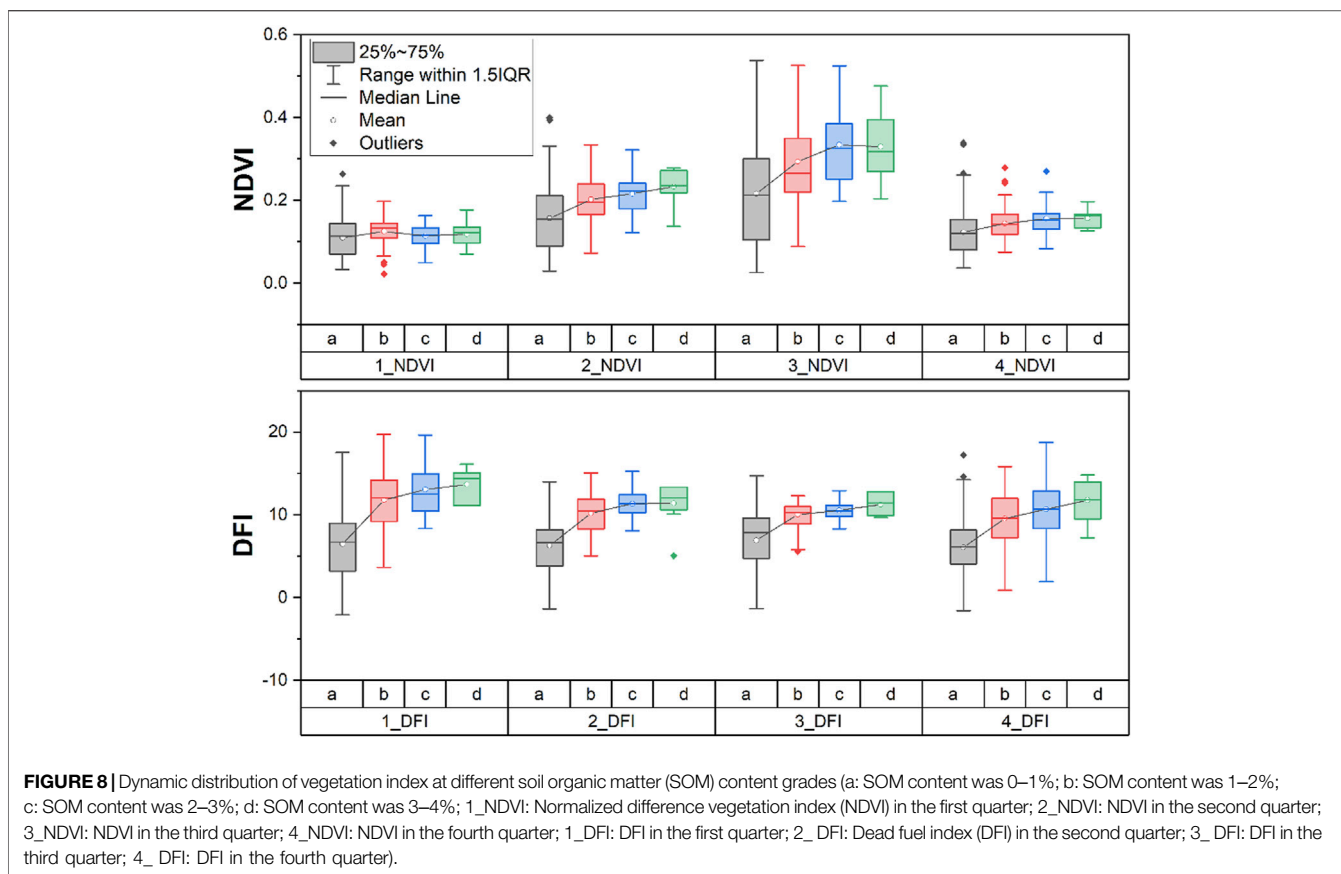


FIGURE 7 | Mean SOM content in different stages of desertification (error bars: standard deviation).

DISCUSSION

Variable Covariate Importance

The NDVI has been regarded as the most important and commonly used alternative index and plays an important role in SOM and SOC emissions (McBratney et al., 2003; Yang et al.,



2020; Zhao et al., 2014). However, in desertified areas with sparse and non-photosynthetic vegetation coverage, NDVI does not provide sufficient vegetation information to predict SOM (Wang et al., 2018). As shown in **Figure 5**, the contribution of NDVI was lower than that of MAT and elevation factors, and the inclusion of quarterly NDVI information resulted in little improvement, whereas DFI had important effects on SOM estimation in the desertified lands. DFIs in Q1 and Q2 were the most important variables (**Figure 5C**), and the relative importance reached 46%, followed by that of elevation and MAT, while the importance of the best performance of the seasonal NDVI only reached 4%. This finding is consistent with other studies (Maynard and Levi, 2017), who found that when bare land reaches or exceeds 20%, NDVI (as a representative variable of vegetation characteristics) no longer has statistical significance. At the same time, Wang et al. (2018) also proved that the indicator function of non-green vegetation on soil carbon content cannot be ignored.

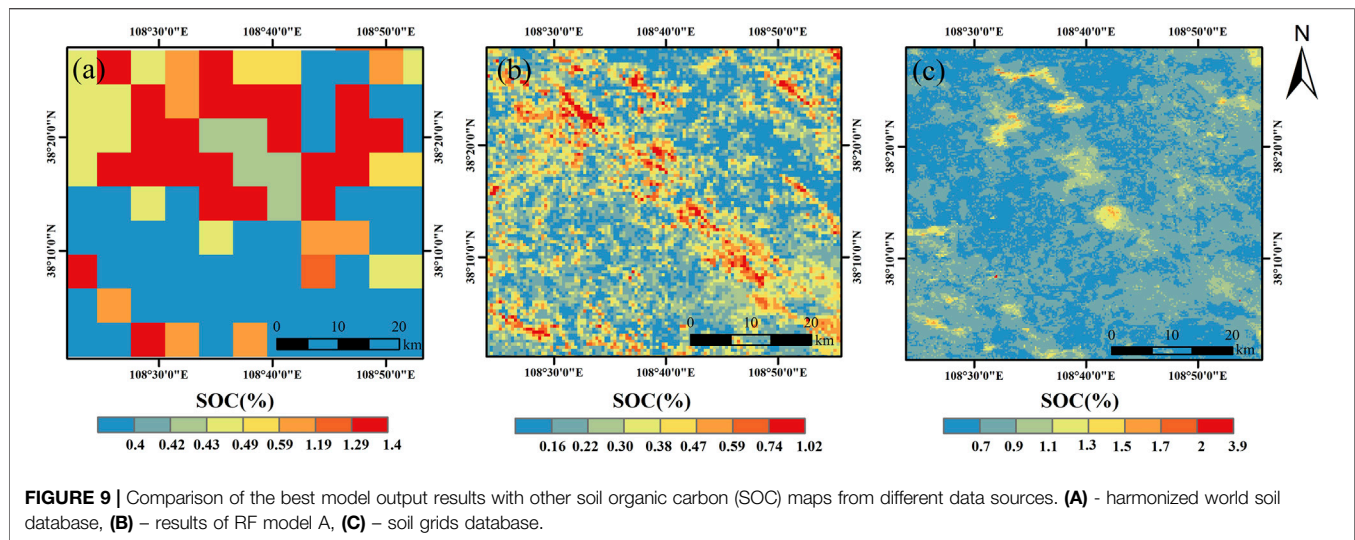
The dynamic distribution of quarterly NDVI and DFI was plotted against four grades of SOM content from low to high, namely 0–1, 1–2, 2–3, and 3–4%, respectively (**Figure 8**). The box plot shows that NDVI and DFI had different quarterly trends. At the same level of SOM content, NDVI in Q2 was higher than NDVI in other quarters, while DFI reached its maximum value in Q1. Additionally, as the level of SOM increased, the values of quarterly NDVI and DFI also increased. This is also consistent with the results of previous studies (Wang et al., 2018, Wang et al.

2017) because vegetation controls the input of surface organic matter, and soil fertility will also determine the growth of ground vegetation (Plaza et al., 2018). The importance of DFI in Q1 and Q2 can be explained, since the separability in the distribution of DFI values for the 0–1% SOM degradation in Q1 and Q2 is most significant, which is completely different from the upper limit, lower limit, and quantile of other SOM content level, compared to the distributions of other vegetation indices.

Comparison of Soil Organic Carbon Map to Other Existing Datasets

Currently, research on SOM or SOC in China focuses on small areas in a province or a small ecosystem (Li et al., 2018; Sun et al., 2019). Very few large-scale studies have been conducted and have a low resolution (1 km). Furthermore, the sample data for the estimation of SOM in China is from the 1980 soil census data, time is old, and errors exist (Feng et al., 2002; Liang et al., 2019b). Moreover, this study shows that the estimation of SOM content in northern China is highly uncertain because of the uneven distribution of historical samples and the low SOM content (Liang et al., 2019a). Our work addresses these knowledge gaps to some extent and has the potential to be extended to other areas.

To demonstrate the advantages of our products, comparisons with the harmonized world soil database (WIEDER, 2014) and soil grids (Hengl et al., 2017) [i.e., two of the most widely used global soil organic matter products and with a higher resolution (250 m)] were



conducted. The results of the RF model were converted into soil organic carbon using a coefficient of 0.58, and we took an area of Mu Us Sandy Land as an example to make small-scale comparisons of different data products (Figure 9). Clearly, our 30-m resolution product shows more spatial details. In the desertification area of northern China, the ground situation is relatively complex, and the main vegetation types in the sandy land are sparsely distributed shrubs with strong spatial heterogeneity. Estimation of SOM content with a low resolution would involve a large number of mixed pixels. The shrub-covered area with relatively high SOM content could not be distinguished from bare soil or sand with extremely low SOM content, and the spatial variation of the SOM content cannot be expressed.

Effects of Desertification on SOM Content

According to our research, the SOM content decreased as the degree of desertification increased. The SOM content changed from 0.81 for slight desertification to 0.62 (g/100 g), 0.51, and 0.42 as the degree of desertification deepened, and the SOM content decreased by 48% from areas with slight to extremely severe desertification. The majority of the SOM is lost in the early stages of desertification, and less loss occurs in the subsequent stage. In a study of soil nutrients in different degrees of desertification on the Qinghai-Tibet Plateau, Li et al. (2006) also found that the deepening of the degree of land desertification will lead to a reduction of SOM and that although there was a significant difference between the mild and heavy stages of desertification ($p < 0.01$), there was no significant difference between the heavy and severe stages of desertification, which also means that most of the organic matter in the topsoil had been lost in the heavy stage of desertification.

Although different studies have reached the same conclusion on the impact of land desertification on SOM (Li et al., 2006; Zhao et al., 2009; Tang et al., 2015), the differences in SOM between different degrees of desertification were inconsistent. In our study, the SOM content of slight to extremely severe desertification decreased by 48%, while Zhao et al. reported a 29% reduction (Zhao et al., 2009). In short, the amount of SOM

loss caused by land desertification both depends on the initial level of SOM in the study area and the type of desertification.

Given the large area of sandy land in northern China (37.59 M Ha) and obviously SOM content in different desertification degrees, the large scale restoration of deserted land under way, such as the Beijing and Tianjin Sandstorm Source Controlling Program (Wu et al., 2013), the Grain for Green Project (Stokes et al., 2010), and the Three-North Shelterbelt Project (Ma, 2004), would have an important effect on SOM gains in the deserted land. At the same time, it is easier to obtain an increase in SOM by first restoring the deserted lands not seriously degraded.

CONCLUSION

In this study, we produced a high spatial resolution (30 m) SOM content map for the deserted land in northern China based on the online remote sensing dataset and machine learning techniques of the GEE platform. Our study initially used quarterly NDVI and DFI as remote sensing indicators to predict SOM contents using machine learning and linear regression models. The results show that the quarterly information of the green and non-photosynthetic vegetation was identified as a key covariate for SOM content estimation in deserted land, and an obvious improvement was observed after simultaneously combining the DFI and NDVI of the four quarters (i.e., R^2 increased by 0.06, the RMSE decreased by 0.05, RPD increased by 0.2, and RPIQ increased by 0.5). The importance of DFI in Q1 and Q2 for low SOM content estimation was illustrated in this study. The performance of machine learning was better than that of the linear regression model, of which RF ($R^2 = 0.80$; RMSE = 0.36) showed some advantages compared to SVM ($R^2 = 0.66$; RMSE = 0.47). This is promising for SOM content mapping for deserted land with low SOM content and sparse vegetation in arid regions, especially with the assistance of the GEE platform. This study also reveals that the high spatial resolution SOM content map (30 m) is more appropriate for heterogeneous deserted land than other existing coarse spatial resolution global SOM products. In

addition, our study found that the majority of SOM is lost in the early stage of desertification and that restoring desertified land has the potential to increase SOC storage when considering the vast and wide distribution of desertified land. In the future, SOM content mapping of desertified land could be improved by incorporating more representative predictor variables and improving machine learning training. However, the products developed in this study could be utilized for desertification and soil carbon studies in northern China.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Agriculture Chemistry Speciality Council (1983). *Agriculture Chemistry Speciality Council, Soil Science Society of China*. Beijing: Science Press.
- Allington, G. R. H., and Valone, T. J. (2010). Reversal of Desertification: The Role of Physical and Chemical Soil Properties. *J. Arid Environments* 74, 973–977. doi:10.1016/J.JARIDENV.2009.12.005
- An, H., Li, Q.-L., Yan, X., Wu, X.-Z., Liu, R.-t., and Fang, Y. (2019). Desertification Control on Soil Inorganic and Organic Carbon Accumulation in the Topsoil of Desert Grassland in Ningxia, Northwest China. *Ecol. Eng.* 127, 348–355. doi:10.1016/j.ecoleng.2018.12.014
- Ballabio, C., Fava, F., and 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. doi:10.1016/j.geoderma.2012.04.002
- Ballabio, C. (2009). Spatial Prediction of Soil Properties in Temperate Mountain Regions Using Support Vector Regression. *Geoderma* 151, 338–350. doi:10.1016/j.geoderma.2009.04.022
- Beaton, A., Whaley, R., Corston, K., and Kenny, F. (2019). Identifying Historic River Ice Breakup Timing Using MODIS and Google Earth Engine in Support of Operational Flood Monitoring in Northern Ontario. *Remote Sensing Environ.* 224, 352–364. doi:10.1016/j.rse.2019.02.011
- Besalatpour, A. A., Ayoubi, S., Hajabbasi, M. A., Mosaddeghi, M. R., and Schulin, R. (2013). Estimating Wet Soil Aggregate Stability from Easily Available Properties in a Highly Mountainous Watershed. *CATENA* 111, 72–79. doi:10.1016/j.catena.2013.07.001
- Breiman, L. (2001). Random Forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., and Edwards, T. C. (2015). Machine Learning for Predicting Soil Classes in Three Semi-arid Landscapes. *Geoderma* 239–240, 68–83. doi:10.1016/j.geoderma.2014.09.019
- Cao, X., Chen, J., Matsushita, B., and Imura, H. (2010). Developing a MODIS-Based index to Discriminate Dead Fuel from Photosynthetic Vegetation and Soil Background in the Asian Steppe Area. *Int. J. Remote Sensing* 31, 1589–1604. doi:10.1080/01431160903475274
- Castaldi, F., Hueni, A., Chabrilat, S., Ward, K., Buttafuoco, G., Bomans, B., et al. (2019). Evaluating the Capability of the Sentinel 2 Data for Soil Organic Carbon Prediction in Croplands. *ISPRS J. Photogrammetry Remote Sensing* 147, 267–282. doi:10.1016/j.isprsjprs.2018.11.026
- Chagas, C. d. S., de Carvalho Junior, W., Bhering, S. B., and Calderano Filho, B. (2016). Spatial Prediction of Soil Surface Texture in a Semi-arid Region Using Random forest and Multiple Linear Regressions. *CATENA* 139, 232–240. doi:10.1016/j.catena.2016.01.001
- Chang, C.-C., and Lin, C.-J. (2011). Libsvm. *Acm Trans. Intell. Syst. Technol.* 2, 1–27. doi:10.1145/1961189.1961199

AUTHOR CONTRIBUTIONS

XL designed the experiment; JY analyzed the data and completed the draft; BS performed the field survey; CY provided the map of sandy land distribution; BW and ZG provided guidance on experiment design; JW revised the draft. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the National Key Research and Development Program (No. 2016YFC0500806), National Natural Science Foundation of China (No. 41571421), and the Dragon 5 Programme (No. 59313).

- Clark, M. L. (2017). Comparison of Simulated Hyperspectral HypsIRI and Multispectral Landsat 8 and Sentinel-2 Imagery for Multi-Seasonal, Regional Land-Cover Mapping. *Remote Sensing Environ.* 200, 311–325. doi:10.1016/j.rse.2017.08.028
- Copernicus Climate Change Service (2017). *ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate*. Copernicus Climate Change Service Climate Data Store (CDS). Available at: <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Cortes, C., and Vapnik, V. (1995). Support-vector Networks. *Mach. Learn.* 20, 273–297. doi:10.1007/bf00994018
- Dong, J., Xiao, X., Menarguez, M. A., Zhang, G., Qin, Y., Thau, D., et al. (2016). Mapping Paddy rice Planting Area in Northeastern Asia with Landsat 8 Images, Phenology-Based Algorithm and Google Earth Engine. *Remote Sensing Environ.* 185, 142–154. doi:10.1016/j.rse.2016.02.016
- Feng, Q., Endo, K. N., and Guodong, C. (2002). Soil Carbon in Desertified Land in Relation to Site Characteristics. *Geoderma* 106, 21–43. doi:10.1016/S0016-7061(01)00099-4
- Forkuor, G., Hounkpatin, O. K. L., Welp, G., and 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. doi:10.1371/journal.pone.0170478
- Friedl, M., and Sulla-Menashe, D. (2019). *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006*. Sioux Falls: NASA EOSDIS Land Processes DAAC. doi:10.5067/MODIS/MCD12Q1.006
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing Environ.* 202, 18–27. doi:10.1016/j.rse.2017.06.031
- Guo, L., Fu, P., Shi, T., Chen, Y., Zhang, H., Meng, R., et al. (2020). Mapping Field-Scale Soil Organic Carbon with Unmanned Aircraft System-Acquired Time Series Multispectral Images. *Soil Tillage Res.* 196, 104477. doi:10.1016/j.still.2019.104477
- 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. doi:10.1016/j.catena.2019.104141
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLoS One* 12, e0169748. doi:10.1371/journal.pone.0169748
- Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, Available from the CGIAR-CSI SRTM 90m Database. Available at: <https://srtm.csi.cgiar.org>.
- Jenny, H. (1941). *Factors of Soil Formation*. New York, NY: McGraw-Hill.
- Keskin, H., Grunwald, S., and Harris, W. G. (2019). Digital Mapping of Soil Carbon Fractions with Machine Learning. *Geoderma* 339, 40–58. doi:10.1016/j.geoderma.2018.12.037

- 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. doi:10.1016/j.geoderma.2019.05.031
- Lee, Y., Jung, C., and Kim, S. (2019). Spatial Distribution of Soil Moisture Estimates Using a Multiple Linear Regression Model and Korean Geostationary Satellite (COMS) Data. *Agric. Water Manag.* 213, 580–593. doi:10.1016/j.agwat.2018.09.004
- Li, Q.-q., Yue, T.-x., Wang, C.-q., Zhang, W.-j., Yu, Y., Li, B., et al. (2013). Spatially Distributed Modeling of Soil Organic Matter across China: An Application of Artificial Neural Network Approach. *CATENA* 104, 210–218. doi:10.1016/J.CATENA.2012.11.012
- Li, X. R., Jia, X. H., and Dong, G. R. (2006). Influence of Desertification on Vegetation Pattern Variations in the Cold Semi-arid Grasslands of Qinghai-Tibet Plateau, North-west China. *J. Arid Environments* 64, 505–522. doi:10.1016/j.jaridenv.2005.06.011
- Li, Y., Wang, X., Niu, Y., Lian, J., Luo, Y., Chen, Y., et al. (2018). Spatial Distribution of Soil Organic Carbon in the Ecologically Fragile Horqin Grassland of Northeastern China. *Geoderma* 325, 102–109. doi:10.1016/j.geoderma.2018.03.032
- Liang, Z., Chen, S., Yang, Y., Zhao, R., Shi, Z., and Viscarra Rossel, R. A. (2019a). National Digital Soil Map of Organic Matter in Topsoil and its Associated Uncertainty in 1980's China. *Geoderma* 335, 47–56. doi:10.1016/J.GEODERMA.2018.08.011
- Liang, Z., Chen, S., Yang, Y., Zhou, Y., and Shi, Z. (2019b). High-resolution Three-Dimensional Mapping of Soil Organic Carbon in China: Effects of SoilGrids Products on National Modeling. *Sci. Total Environ.* 685, 480–489. doi:10.1016/J.SCITOTENV.2019.05.332
- Ma, Q. (2004). Appraisal of Tree Planting Options to Control Desertification: Experiences from the Three-North Shelterbelt Programme. *Int. Forest. Rev.* 6, 327–334. doi:10.1505/ifer.6.3.327.59972
- Mahmoudabadi, E., Karimi, A., Haghnia, G. H., and Sepehr, A. (2017). Digital Soil Mapping Using Remote Sensing Indices, Terrain Attributes, and Vegetation Features in the Rangelands of Northeastern Iran. *Environ. Monit. Assess.* 189, 500–520. doi:10.1007/s10661-017-6197-7
- Maynard, J. J., and Levi, M. R. (2017). Hyper-temporal Remote Sensing for Digital Soil Mapping: Characterizing Soil-Vegetation Response to Climatic Variability. *Geoderma* 285, 94–109. doi:10.1016/j.geoderma.2016.09.024
- McBratney, A. B., Mendonça Santos, M. L., and Minasny, B. (2003). On Digital Soil Mapping. *Geoderma* 117, 3–52. doi:10.1016/S0016-7061(03)00223-4
- MEA – Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well Being: Desertification Synthesis*. Washington, DC, USA: Island Press.
- 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. doi:10.1016/j.geoderma.2007.08.025
- O'Kelly, B. C. (2004). Accurate Determination of Moisture Content of Organic Soils Using the Oven Drying Method. *Drying Tech.* 22, 1767–1776. doi:10.1081/DRT-200025642
- Plaza, C., Gascó, G., Méndez, A. M., Zaccone, C., and Maestre, F. T. (2018). "Soil Organic Matter in Dryland Ecosystems," in *Chapter 2 - Soil Organic Matter in Dryland Ecosystems*. Editors C. Garcia, P. Nannipieri, and T.B.T Hernandez (Academic Press), 39–70. doi:10.1016/B978-0-12-811687-6.00002-X
- Qi, F., Liu, W., Liu, Y., Yanwu, Z., and Yonghong, S. (2004). Impact of Desertification and Global Warming on Soil Carbon in Northern China. *J. Geophys. Res.* 109. doi:10.1029/2003jd003599
- Sayão, V. M., Demattê, J. A. M., Bedin, L. G., Nanni, M. R., and Rizzo, R. (2018). Satellite Land Surface Temperature and Reflectance Related with Soil Attributes. *Geoderma* 325, 125–140. doi:10.1016/j.geoderma.2018.03.026
- Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., et al. (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.* 601–602, 821–832. doi:10.1016/j.scitotenv.2017.05.239
- Sims, N. C., England, J. R., Newnham, G. J., Alexander, S., Green, C., Minelli, S., et al. (2019). Developing Good Practice Guidance for Estimating Land Degradation in the Context of the United Nations Sustainable Development Goals. *Environ. Sci. Pol.* 92, 349–355. doi:10.1016/j.envsci.2018.10.014
- State Forestry Administration of China (2015). A Bulletin of Status of Desertification and Sandification in China. Available online: <http://www.forestry.gov.cn/main/69/content-831684.html> (Accessed December 29, 2015).
- Stokes, A., Sotir, R., Chen, W., and Ghestem, M. (2010). Soil Bio- and Eco-Engineering in China: Past Experience and Future Priorities. *Ecol. Eng.* 36, 247–257. doi:10.1016/j.ecoleng.2009.07.008
- Sun, B., Wang, Y., Li, Z., Gao, W., Wu, J., Li, C., et al. (2019). Estimating Soil Organic Carbon Density in the Otindag Sandy Land, Inner Mongolia, China, for Modelling Spatiotemporal Variations and Evaluating the Influences of Human Activities. *CATENA* 179, 85–97. doi:10.1016/J.CATENA.2019.03.034
- Tang, Z.-S., An, H., and Shangguan, Z.-P. (2015). The Impact of Desertification on Carbon and Nitrogen Storage in the Desert Steppe Ecosystem. *Ecol. Eng.* 84, 92–99. doi:10.1016/J.ECOLENG.2015.07.023
- Tao, W. (2014). Aeolian Desertification and its Control in Northern China. *Int. Soil Water Conservation Res.* 2, 34–41. doi:10.1016/S2095-6339(15)30056-3
- Tongway, D. J., and Ludwig, J. A. (2007). "Heterogeneity in Arid and Semiarid Lands," in *Ecosystem Function in Heterogeneous Landscapes* (New York: Springer), 189–205. doi:10.1007/0-387-24091-8_10
- Tucker, C. J. (1979). Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing Environ.* 8, 127–150. doi:10.1016/0034-4257(79)90013-0
- UNCCD (1994). "United Nations Convention to Combat Desertification in Those Countries Experiencing Serious Drought And/or Desertification," in *Particularly in Africa* (Geneva: UNEP).
- Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., and Skjemstad, J. O. (2006). Visible, Near Infrared, Mid Infrared or Combined Diffuse Reflectance Spectroscopy for Simultaneous Assessment of Various Soil Properties. *Geoderma* 131, 59–75. doi:10.1016/j.geoderma.2005.03.007
- Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., et al. (2019). Estimation of Soil pH Using PXRF Spectrometry and Vis-NIR Spectroscopy for Rapid Environmental Risk Assessment of Soil Heavy Metals. *Process Saf. Environ. Prot.* 132, 73–81. doi:10.1016/j.psep.2019.09.025
- Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., et al. (2018). High Resolution Mapping of Soil Organic Carbon Stocks Using Remote Sensing Variables in the Semi-arid Rangelands of Eastern Australia. *Sci. Total Environ.* 630, 367–378. doi:10.1016/J.SCITOTENV.2018.02.204
- Wang, D., Li, X., Zou, D., Wu, T., Xu, H., Hu, G., et al. (2020). Modeling Soil Organic Carbon Spatial Distribution for a Complex Terrain Based on Geographically Weighted Regression in the Eastern Qinghai-Tibetan Plateau. *CATENA* 187, 104399. doi:10.1016/j.catena.2019.104399
- 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, 250–263. doi:10.1016/J.GEODERMA.2017.05.048
- Wang, T., Xiang, S., Chang-zhen, Y., Sen, L., and Jia-li, X. (2011). Remote Sensing Analysis of Aeolian Desertification Trend in Northern China during 1975–2010. *J. Desert Res.* 31, 1351–1356. doi:10.1080/00405000.2010.522047
- Wang, X., Li, Y., Gong, X., Niu, Y., Chen, Y., Shi, X., et al. (2019). Storage, Pattern and Driving Factors of Soil Organic Carbon in an Ecologically Fragile Zone of Northern China. *Geoderma* 343, 155–165. doi:10.1016/j.geoderma.2019.02.030
- Were, K., Bui, D. T., Dick, Ø. B., and 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. *Ecol. Indicators* 52, 394–403. doi:10.1016/j.ecolind.2014.12.028
- WIEDER, W. (2014). *Regridded Harmonized World Soil Database v1.2*. doi:10.3334/ORNLDAAAC/1247
- Wiesmeier, M., Urbanski, L., Hobbey, E., Lang, B., von Lützw, M., Marin-Spiotta, E., et al. (2019). Soil Organic Carbon Storage as a Key Function of Soils - A Review of Drivers and Indicators at Various Scales. *Geoderma* 333, 149–162. doi:10.1016/j.geoderma.2018.07.026
- Wu, W., Zucca, C., Karam, F., and Liu, G. (2016). Enhancing the Performance of Regional Land Cover Mapping. *Int. J. Appl. Earth Obs. Geoinform.* 52, 422–432. doi:10.1016/j.jag.2016.07.014
- Wu, W., Zucca, C., Muhaimeed, A. S., Al-Shafie, W. M., Fadhil Al-Quraishi, A. M., Nangia, V., et al. (2018). Soil Salinity Prediction and Mapping by Machine

- Learning Regression in Central Mesopotamia, Iraq. *Land Degrad. Develop.* 29 (11), 4005–4014. doi:10.1002/ldr.3148
- Wu, Z., Wu, J., Liu, J., He, B., Lei, T., and Wang, Q. (2013). Increasing Terrestrial Vegetation Activity of Ecological Restoration Program in the Beijing-Tianjin Sand Source Region of China. *Ecol. Eng.* 52, 37–50. doi:10.1016/j.ecoleng.2012.12.040
- Xu, D., You, X., and Xia, C. (2019). Assessing the Spatial-Temporal Pattern and Evolution of Areas Sensitive to Land Desertification in North China. *Ecol. Indicators* 97, 150–158. doi:10.1016/j.ecolind.2018.10.005
- Yang, L., He, X., Shen, F., Zhou, C., Zhu, A.-X., Gao, B., et al. (2020). Improving Prediction of Soil Organic Carbon Content in Croplands Using Phenological Parameters Extracted from NDVI Time Series Data. *Soil Tillage Res.* 196, 104465. doi:10.1016/j.still.2019.104465
- 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, 445–452. doi:10.1016/j.geoderma.2018.09.006
- Zhang, H., Wu, P., Yin, A., Yang, X., Zhang, M., and Gao, C. (2017). Prediction of Soil Organic Carbon in an Intensively Managed Reclamation Zone of Eastern China: A Comparison of Multiple Linear Regressions and the Random forest Model. *Sci. Total Environ.* 592, 704–713. doi:10.1016/j.scitotenv.2017.02.146
- Zhang, Z., and Huisingsh, D. (2018). Combating Desertification in China: Monitoring, Control, Management and Revegetation. *J. Clean. Prod.* 182, 765–775. doi:10.1016/j.jclepro.2018.01.233
- Zhao, H.-L., He, Y.-H., Zhou, R.-L., Su, Y.-Z., Li, Y.-Q., and Drake, S. (2009). Effects of Desertification on Soil Organic C and N Content in sandy farmland and Grassland of Inner Mongolia. *CATENA* 77, 187–191. doi:10.1016/j.catena.2008.12.007
- 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. Indicators* 39, 120–133. doi:10.1016/j.ecolind.2013.12.015
- Zhi, J., Zhang, G., Yang, R., Yang, F., Jin, C., Liu, F., et al. (2018). An Insight into Machine Learning Algorithms to Map the Occurrence of the Soil Mattic Horizon in the Northeastern Qinghai-Tibetan Plateau. *Pedosphere* 28, 739–750. doi:10.1016/S1002-0160(17)60481-8
- 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. doi:10.1016/j.scitotenv.2020.138244

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Junting, Xiaosong, Bo, Junjun, Bin, Changzhen and Zhihai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.