

## Article

# Influence of the Selection of Interpolation Method on Revealing Soil Organic Carbon Variability in the Red Soil Region, China

Zhongqi Zhang <sup>1,2</sup> , Dongsheng Yu <sup>2,3,\*</sup> , Xiyang Wang <sup>2,3</sup>, Yue Pan <sup>2,3</sup>, Guangxing Zhang <sup>2,3</sup> and Xuezheng Shi <sup>2,3</sup>

<sup>1</sup> School of Geography, Geomatics, and Planning, Jiangsu Normal University, Xuzhou 221116, China; zhangzq128@126.com

<sup>2</sup> State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; deemhunter88@163.com (X.W.); panyue611@126.com (Y.P.); gxz1984@gmail.com (G.Z.); xzshi@issas.ac.cn (X.S.)

<sup>3</sup> Graduate University of the Chinese Academy of Sciences, Beijing 100039, China

\* Correspondence: dshyu@issas.ac.cn; Tel.: +86-25-86881272

Received: 25 May 2018; Accepted: 29 June 2018; Published: 3 July 2018



**Abstract:** At present, various interpolation methods have been used for obtaining the spatial variability of soil organic carbon (SOC) in some regions. However, systematic research on accuracy comparison and the influence of method selection on revealing SOC variability is still lacking for the hilly red soil region of China. In this study, the spatial distributions of SOC were interpolated via three kriging methods (ordinary kriging (OK), kriging combined with soil type information (KSt), and with land-use information (KLu)) and three polygon-based methods (linking sample points data to the sampling grid (PGr), to soil type polygon (PSt), and to land-use polygon (PLu)) in Yujiang County, Jiangxi Province, China. To illustrate the influence of method selection on revealing SOC variability, the prediction efficiencies of these methods were compared with soil validation samples. The results showed that the KLu and PLu had high prediction accuracy, followed immediately by the KSt and PSt, while the OK and PGr had very low accuracies. Meanwhile, the SOC distribution contours obtained by KLu, PLu, KSt, and PSt could better reflect the real conditions of the study area, as they were consistent with the variation of land-use or soil types. However, the contours from the OK and PGr had poor performance with sketchy borderlines and regular homogeneous grid cells, respectively. Therefore, great differences exist in the prediction efficiency of different methods, and the interpolation method selection has an important impact on revealing SOC variability in the hilly red soil region of China. The results can provide a useful reference for the efficient SOC prediction and simulation in similar soil regions.

**Keywords:** soil organic carbon; interpolation method; hilly red soil region of China

## 1. Introduction

Soil carbon pool is the largest carbon reservoir in terrestrial ecosystems, and has played a key role in global climate change [1–4]. Generally, the soil organic carbon (SOC) has high spatial variability due to complicated soil-forming factors and increasing human activities [5,6]. As the main control factors are usually different, there are great differences in the SOC variability characteristics among different regions [7,8]. Accurately detecting the spatial variability of the regional SOC is a precondition for implementing proper measures for regional soil carbon sequestration and agricultural environment management.

In recent years, the study of the spatial variability of SOC has become an important field in soil research [6,9,10]. Based on soil sampling points, scholars developed several methods to realize point to surface expansion for obtaining the regional SOC distribution including the multivariate regression method [11], the trend-surface method [12], the inverse distance weighted method (IDW) [13], the traditional polygon-based method [14–16], and the kriging methods [17,18]. These methods have made certain contributions to SOC prediction and simulation in the corresponding soil regions. Among these methods, existing literature reports have shown that polygon-based and kriging methods have been applied more extensively than other methods for regional SOC prediction and simulation in recent years.

Before the wide use of the geostatistical interpolation method, the traditional polygon-based method, linking the property data of soil samples to the corresponding delineations according to the identity or similarity principle, was most commonly used to realize the expansion from the sampling point to the surface distribution of the soil property [15,19]. For instance, Davidson and Lefebvre [20] calculated and compared the total soil C stocks of the state of Maine using different soil map scales with this method. In recent years, this method has still been used by some scholars. Yu et al. studied the soil assessment unit scale effects quantifying the CH<sub>4</sub> emissions from rice fields in the Tai-Lake region of China [19]. With the development of geographic information system (GIS) technology and geostatistics, kriging has gradually become the dominant method for SOC spatial prediction and simulation. Based on the theory of regionalized variables, kriging can make an unbiased and optimal estimation for soil property in local regions. Using this method, the distribution of the soil properties and the corresponding predicted error can be expressed. For example, by using kriging and GIS approaches, Marchetti et al. investigated the spatial variability of the soil organic matter (SOM) in the Abruzzo region of central Italy under the Mediterranean climatic condition [21]. Darilek et al. examined the temporal variability of SOM and other fertility parameters on two main soil types using this method in Zhangjiagang County, China during three periods of 1983–1989, 1989–1999, and 1999–2004 [22]. Obviously, this method has received favorable remarks from soil scholars.

Confronting different soil environment conditions and increasing the prediction accuracy requirement, the polygon-based and kriging methods have made great progress and a series of methods have been developed for practical application in recent years. For the polygon-based method, soil property data can be linked to soil type, land-use pattern, climate zone, agricultural region, and even administrative region unit and regular grid [19,23–25]. Meanwhile, kriging has also evolved into the methods such as universal kriging (UK), co-kriging (CK), regression kriging (RK), and kriging combined with other auxiliary information (soil, land-use, vegetation type, and other related information) [26–29]. After being modified, the polygon-based and kriging methods are usually much more adaptive in the spatial prediction of SOC in some soil regions. Some scholars also have compared the performance of the different derivative methods in corresponding soil regions [15,27,30,31]. Zhao et al. evaluated the SOC prediction uncertainty of different prediction methods including linking soil profiles to polygons of a digital soil map, multiple linear regression (MLR), universal kriging (UK), and regression-kriging (RK) in Hebei Province, China [15]. He found that the prediction uncertainty of RK was less than the other methods. Chai et al. predicted the spatial distribution of soil organic matter in the Pinggu district of Beijing with several methods including regression-kriging and empirical best linear unbiased predictor, and evaluated their performance [27]. Kumar et al. compared ordinary kriging and co-kriging methods using different terrain attributes as auxiliary factors in a watershed of the Himalayan landscape in the western part of India, and pointed out that co-kriging with terrain attributes could greatly improve the prediction accuracy of the soil total carbon [31]. Although studies have been conducted in some soil regions, the obtained conclusions cannot be used directly in other soil regions as the SOC has its specific variability with the featured regional characteristics. Hence, the efficient comparison of interpolation methods still needs to be conducted in the regions lacking comparative research.

Covering a large area of approximately  $1.18 \times 10^6 \text{ km}^2$ , the hilly red soil region is a critical agricultural zone with a dense population, and plays an important role in the food security of China [32]. Meanwhile, this area plays an important role in China as added soil carbon sequestration due to superior heat and moisture conditions. However, the SOC has high spatial variability in this area due to the complex terrain, broken geomorphic units, diversified soil types and land-use patterns [33]. As a result, it is difficult to predict the SOC spatial variability, and the prediction accuracy consequently has not always been satisfied. Facing this problem, the aforementioned modified methods perhaps can improve the prediction accuracies given the consideration of specific influence factors of SOC. However, the performance of existing modified methods on improving the prediction accuracy is not clear. Which method is the priority selection still needs further discussion in the hilly red soil region. Although some studies have been reported in other regions [15,27], the results may not be suitable for the red soil region due to the complex terrain and fast changing soil types and land-use patterns. A comprehensive comparative study for different interpolation methods in this area is still necessary and has important practical significance. For this reason, this study was initiated in the hilly red soil region of Yujiang County in Jiangxi Province, South China, and various polygon-based and kriging-based methods were compared to obtain the spatial variability characteristic of the SOC content. The primary objectives of this research were to: (1) quantify and compare the SOC prediction accuracies of various polygon-based and kriging-based methods with the same soil samples; and (2) analyze the influence of the interpolation method selection on SOC variability in the hilly red soil region in Southern China.

## 2. Materials and Methods

### 2.1. Study Area

The two study areas are located in the east-central part of Yujiang County ( $116^\circ 41' - 117^\circ 09' \text{ E}$ ,  $28^\circ 04' - 28^\circ 37' \text{ N}$ ), and dominated by hills and low mountains [34]. The areas lie in a subtropical zone with a warm climate, long frost-free period, abundant heat and sunshine, and plentiful rainfall. The mean annual precipitation of Yujiang County is 1758 mm, and the mean annual temperature is  $17.6^\circ \text{C}$  (1945–2004, Yujiang meteorological station, Jiangxi Province). Paddy soil (anthrosols, WRB 2006) [35,36] and red soil (acrisols, WRB 2006) are the two dominant soil types (Soil Survey Office of Yujiang County, 1986), and fluvo-aquic soil (cambisols, WRB 2006) is distributed sporadically across the area [37]. The main soil parent materials are red sandstone, quaternary red clay, shale, and alluviums. Paddy fields, crop upland, and forestland are the three major land-use patterns from the records of the Chinese Land Use Database in the two study areas, and the main cultivated crops include rice, peanut, sweet potato, and sesame.

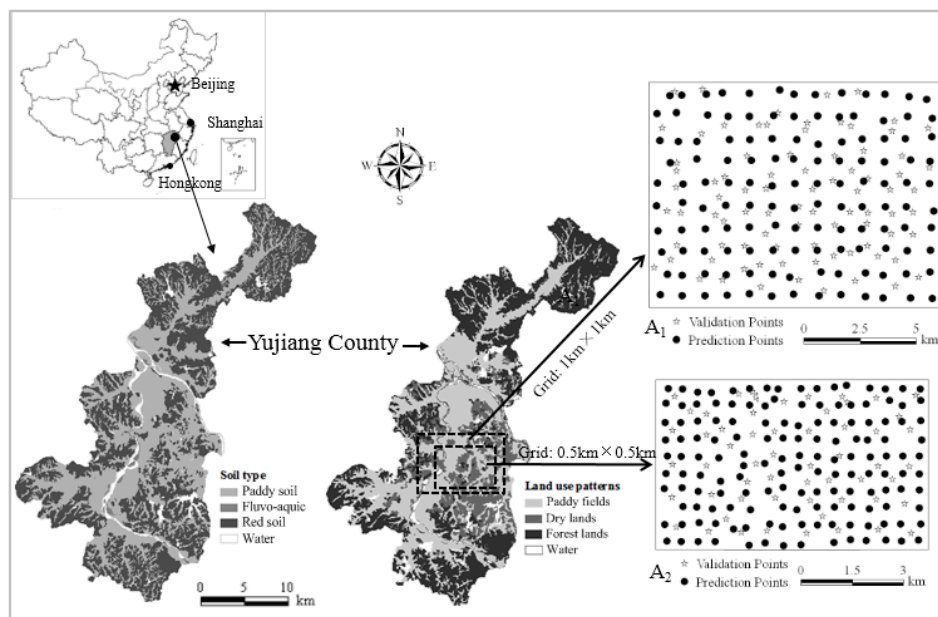
### 2.2. Soil Sampling Design and Analysis

First, 129 soil samples were collected in a  $1 \text{ km} \times 1 \text{ km}$  grid over a  $140 \text{ km}^2$  area ( $A_1$  in Figure 1). Among these samples, 83, 5, and 41 soil samples were taken from three soil types of paddy soil, fluvo-aquic soil, and red soil, which were about 59%, 5%, and 36% of the total area of  $A_1$ , respectively. Furthermore, 68, 45, and 16 samples were taken from three land-use patterns of paddy fields, uplands, and forestlands, which occupied about 56%, 33%, and 11% of the total area of  $A_1$ , respectively (from China Land Use Database of Chinese Academy of Sciences, 1:50,000, 2005). Soil samples were not collected in a few grids due the grids being occupied by water or mostly bare rock. These 129 samples were used for SOC prediction via various prediction methods, and the other 80 soil samples were collected with a homogenous distribution to evaluate the prediction efficiencies of various methods.

Second, to clarify whether the comparison of the prediction efficiencies of various methods coincided in different coverage areas, another independent sample collecting scheme was implemented in a  $40 \text{ km}^2$  area ( $A_2$  in Figure 1). In this scheme, 149 samples were collected with a  $0.5 \text{ km} \times 0.5 \text{ km}$  grid for the SOC spatial prediction. Among these samples, 88, 6, and 55 samples were taken from the

three soil types of paddy soil, fluvo-aquic soil, and red soil, which occupied about 61%, 6%, and 33% of  $A_2$ , respectively. Additionally, 75, 54, and 20 samples were taken from three land-use patterns of paddy fields, uplands, and forestlands, which occupy about 55%, 31%, and 14% of  $A_2$ , respectively. Another 55 samples were collected with a homogenous distribution for evaluating the prediction efficiencies of various methods; these samples contained all of the main soil types and land-use patterns present in the prediction samples.

All samples were taken from the surface layer (0–20 cm) after the crop (rice, peanut, sweet potato, etc.) harvest from October to November 2007. After all of the samples had been air-dried, ground, and passed through a 0.147 mm sieve, the SOC content was determined via the  $K_2Cr_2O_7$  oxidation–titration method, following Nelson and Sommers [38].



**Figure 1.** Study area locations and soil sampling designs in Yujiang County of south China.

### 2.3. Spatial Prediction Methods

Six methods were used for performing SOC spatial prediction in the present study, including three kriging methods and three polygon-based methods. The three kriging methods were ordinary kriging (OK), kriging combined with soil-type information (KSt), and kriging combined with land-use information (KLu). OK has been widely used, and its theoretical principle has already been described in detail in the literature [31,39]. As KSt and KLu are conceptually the same, KSt was taken as an example here to describe the principle of both methods. The SOC content value  $z(x_{kj})$  of every sample was divided into two parts in the KSt method, including the mean value  $\mu(t_k)$  of the corresponding soil type (land use type for KLu method) and the residual  $r(x_{kj})$  after removing the mean value (Equation (1)).

$$z(x_{kj}) = \mu(t_k) + r(x_{kj}), \quad (1)$$

where  $z(x_{kj})$  is the SOC content of the sample at  $x_{kj}$  location, and  $t_k$  is the soil type associated with  $x_{kj}$ . In this method, the residual  $r(x_{kj})$  was treated as a new regionalized variable to execute the OK interpolation. The variogram and the corresponding prediction formula of the residuals are stated as Equations (2) and (3), respectively. The final predicted SOC value of the unknown point was the

sum of the predicted residual value  $r^*(x_{kj})$  and the SOC mean value  $\mu(t_k)$  associated with the soil type (or land use type) to which it belongs (Equation (4)) [40].

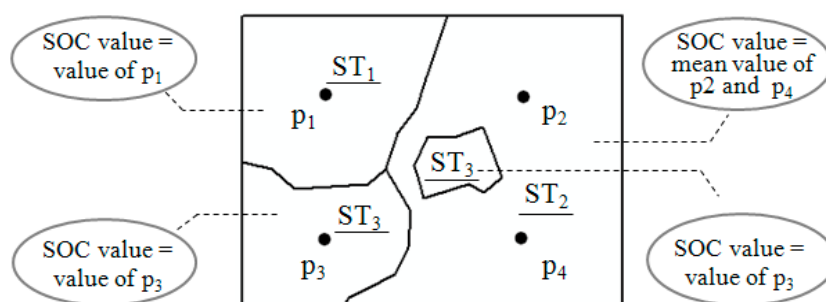
$$\gamma_r(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [r(x_{kj}) - r(x_{kj} + \mathbf{h})], \quad (2)$$

$$r^*(x_{kj}) = \sum_{k=1}^{n(j)} \sum_{j=1}^m \lambda_{kj} r(x_{kj}), \quad (3)$$

$$z^*(x_{kj}) = \mu(t_k) + r^*(x_{kj}), \quad (4)$$

where  $N(\mathbf{h})$  is the number of data pairs separated by distance  $\mathbf{h}$ , and  $r(x_{kj})$  and  $r(x_{kj} + \mathbf{h})$  are the residuals of the measured SOC values at the positions of  $x_{kj}$  and  $x_{kj} + \mathbf{h}$ , respectively.

Polygon-based methods, called the PKB method by some scholars [15,19], link the SOC values of the sampling point to the corresponding polygon, including the sampling grid (1 km  $\times$  1 km in  $A_1$  and 0.5 km  $\times$  0.5 km in  $A_2$ ) (PGr), the soil type polygon (PSt), and the land-use delineations (PLu). According to the figures of  $A_1$  and  $A_2$ , there were 214 and 159 soil polygons, and 192 and 121 land-use polygons in two areas, respectively. Note that for PGr, the few grids without samples were linked with the mean value of all of the samples around them. For PSt and PLu, if only one sample was collected from the corresponding polygons, then only the SOC value was linked to the polygon directly; if two or more samples were collected from one polygon, then the calculated mean SOC value from these samples was linked to it; and if no sample was collected in any polygon, then the SOC value of the sample with the closest distance and same type was linked to this polygon according to the close similarity principle (Figure 2).



**Figure 2.** Schematic diagram of the polygon-based methods (taking PSt as an example:  $P_1$ – $P_4$  represent sampling points and  $ST_1$ – $ST_3$  represent different soil types).

#### 2.4. Uncertainty Evaluation of Various Prediction Methods

To evaluate the prediction accuracies of the various methods, the Pearson's correlation coefficient ( $r$ ) between the measured and predicted SOC values of the validation samples in  $A_1$  ( $n = 80$ ) and  $A_2$  ( $n = 55$ ) were compared for various methods. In addition, the root mean square error (RMSE) and mean absolute error (MAE) of the validation samples were chosen as evaluation indicators (Equations (5) and (6)). Larger  $r$  and smaller RMSE and MAE indicate a higher prediction accuracy.

$$RMSE = \sqrt{\frac{1}{N} \sum (x_{oi} - x_{pi})^2}, \quad (5)$$

$$MAE = \frac{1}{N} \sum ABS(x_{oi} - x_{pi}), \quad (6)$$

where  $N$  is the number of validation points, and  $x_{oi}$  and  $x_{pi}$  are the measured and predicted values of the validation points, respectively. The letter "ABS" means the absolute value.



The classical statistical analyses were conducted using SPSS 16.0 for Windows (SPSS, Chicago, IL, USA). The soil map of the study areas was from the Soil Survey Office of Yujiang County, and the land-use map was obtained from the China Land Use Database compiled by the Chinese Academy of Sciences. The geo-statistical analysis of OK, KSt, and KLu were completed using the software package of Gamma Design Software GS+ (version 5.3, 1994), and the SOC distribution maps from all methods were compiled with ArcGIS 9.3 (ESRI, Redlands, CA, USA).

### 3. Results

#### 3.1. Descriptive Statistics of the SOC Contents in Study Areas

Descriptive statistics of the SOC contents in the two areas are presented in Table 1. The mean SOC values were 12.1 and 11.8 g kg<sup>-1</sup>, and their variation ranged from 2.6 to 26.2 g kg<sup>-1</sup> and 2.4 to 24.2 g kg<sup>-1</sup> in A<sub>1</sub> and A<sub>2</sub>, respectively. The values were lower than those of black soil in Northeast China and slightly higher than the soil in Eastern China [41]. The coefficients of variation (CV) of the SOC contents were 0.47 and 0.48 for A<sub>1</sub> and A<sub>2</sub>, respectively, which were similar and both belonged to medium degree of the variation. Among the three soil types in two areas, the SOC contents of paddy soil were both the highest (14.9 and 15.7 g kg<sup>-1</sup> in A<sub>1</sub> and A<sub>2</sub>), while the red soil had the lowest (6.7 and 5.9 g kg<sup>-1</sup> in A<sub>1</sub> and A<sub>2</sub>) values, i.e., only approximately half the value of paddy soil. The SOC contents of fluvo-aquic soil were both in between in two areas. Among the three land-use patterns, the paddy fields had the highest SOC contents (15.7 and 16.1 g kg<sup>-1</sup> in A<sub>1</sub> and A<sub>2</sub>), while the contents of the uplands were the lowest (6.5 and 5.2 g kg<sup>-1</sup>), only 1/3 to 1/2 of the paddy fields. For the CVs of the SOC contents, the red soils had the highest values, whereas the paddy soils had the lowest among three soil types, and the uplands had the highest values, whereas the paddy fields had the lowest values among the land-use patterns. Judging from the statistical analysis, the data of the SOC contents in the two areas all fit to a normal distribution (Table 1).

**Table 1.** Descriptive statistics of the SOC contents in the two study areas.

Area	Sampling Grids	Land-Use	Soil Type	N <sup>a</sup>	Min. <sup>b</sup>	Max. <sup>c</sup>	Mean	Std.	CV. <sup>d</sup>	Skew. <sup>e</sup>	Kurt. <sup>f</sup>
					g kg <sup>-1</sup>						
A1	1 km × 1 km	Paddy fields		68	4.40	26.15	15.69	4.21	0.27		
		Uplands		45	2.58	13.35	6.50	2.61	0.40		
		Forest lands		16	3.47	18.91	12.20	4.62	0.38		
		Paddy soil		83	3.13	26.15	14.88	4.78	0.32		
		Fluvo-aquic		5	4.58	13.35	8.84	3.25	0.37		
		Red soil		41	2.58	14.36	6.72	2.72	0.41		
		Total		129	2.58	26.15	12.05	5.66	0.47	0.12	−0.91
A2	0.5 km × 0.5 km	Paddy fields		75	5.45	24.15	16.12	4.50	0.28		
		Uplands		54	2.41	11.23	5.24	1.79	0.34		
		Forest lands		20	5.20	21.62	13.38	4.34	0.32		
		Paddy soil		88	3.87	24.15	15.74	4.77	0.30		
		Fluvo-aquic		6	5.89	12.62	8.44	2.77	0.33		
		Red soil		55	2.41	14.66	5.89	2.46	0.43		
		Total		149	2.41	24.15	11.81	5.68	0.48	0.17	−1.36

<sup>a</sup>: sample size; <sup>b</sup>: minimum value; <sup>c</sup>: maximum value; <sup>d</sup>: coefficient of variation; <sup>e</sup>: skewness; <sup>f</sup>: kurtosis.

The results of the analysis of variance (ANOVA) showed that soil type and land-use pattern both had a significant impact on the SOC contents in the two areas (Table 2). The SOC contents were significantly different among the soil types and land-use patterns. This result indicated that the effect of soil type and land-use pattern on the SOC contents should not be ignored. In comparison, the variances of SOC between land-use patterns were greater than those between soil types. This indicated that the impact on SOC content from land-use patterns was greater than the one from soil types.

**Table 2.** Analysis of variance (ANOVA) of the SOC contents with different soil types and land use patterns in two study areas.

Area	Sampling Grids	Types	Source of Variance	df	Sum of Squares	Mean Square	F Value
A <sub>1</sub>	1 km × 1 km	soil types	Between types	1948.1	2	974.1	57.2 **
			Within types	2146.6	126	17.0	
			Total	4094.7	128		
		land-use patterns	Between patterns	2499.1	2	1249.6	98.7 **
			Within patterns	1595.5	126	12.7	
			Total	4094.7	128		
A <sub>2</sub>	0.5 km × 0.5 km	soil types	Between types	3578.9	2	1789.5	117.4 **
			Within types	2225.7	146	15.2	
			Total	5804.6	148		
		land-use patterns	Between patterns	3849.9	2	1925.0	143.8 **
			Within patterns	1954.7	146	13.4	
			Total	5804.6	148		

\*\*  $p < 0.01$ .

### 3.2. Geostatistics Analysis of SOC Content for the Kriging Methods

The semi-variogram models and their parameters for the three kriging methods are shown in Table 3. The semi-variogram of OK, KSt, and KLu in A<sub>1</sub> were well fit by exponential, exponential, and spherical models, and fit by spherical, exponential, and spherical models in A<sub>2</sub>, respectively. In the two areas, the C/Sill ratios of OK, KSt, and KLu were in the range of 0.554 to 0.709 and had a medium degree of spatial autocorrelation [42]. Among the three Kriging methods, the Sills of KSt and KLu are all lower than OK in the two areas because the SOC residuals data have great stability after subtracting the corresponding mean values (Equation (1)). The low volatility of interpolating data is beneficial to spatial prediction. Meanwhile, the C<sub>0</sub> values of KSt and KLu increased, while the auto-correlation distances (Ranges) decreased relative to OK, which means that the semi-variogram models of KSt and KLu could describe the detailed characteristics of SOC variability better than OK.

**Table 3.** Parameters of the best-fit semi-variogram models for the original SOC contents and the corresponding residual data in the two study areas.

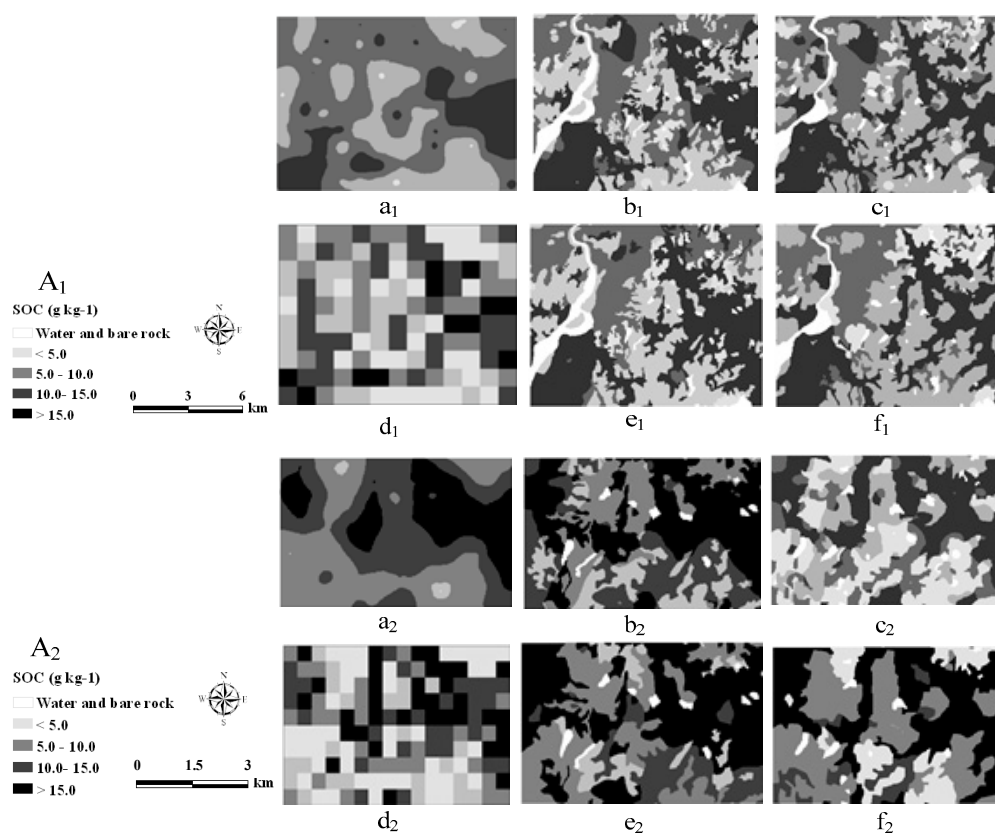
Regions	Sampling Grid	Method	Distribution	Model	C <sub>0</sub>	Sill	C/Sill	Range (m)	R <sup>2</sup>
A <sub>1</sub>	1 km × 1 km	OK	Normal	Exponential	0.345	1.029	0.665	2520	0.871
		KST	Normal	Exponential	0.29	0.995	0.709	1680	0.696
		KLU	Normal	Spherical	0.309	0.962	0.679	1490	0.745
A <sub>2</sub>	0.5 km × 0.5 km	OK	Normal	Spherical	0.471	1.055	0.554	1930	0.939
		KST	Normal	Exponential	0.378	1.033	0.634	1890	0.814
		KLU	Normal	Spherical	0.444	1.039	0.573	1780	0.911

C<sub>0</sub> is the nugget variance; C is the auto-correlated variance; Sill, the sum of C<sub>0</sub> and C, is the maximum variance of the property; C/Sill ratio is the relative proportion of the C from the Sill; Range is the auto-correlation distances at which the sill is obtained, and the R<sup>2</sup> is determinant coefficient.

### 3.3. Spatial Interpolation Contours of the SOC Derived from the Various Methods

The spatial interpolation contours of the SOC contents from various methods are shown in Figure 3. All of the SOC distribution contours from the kriging methods and polygon-based methods had similar patterns in A<sub>1</sub>. There were higher SOC contents in the southeast and southwest regions, whereas lower contents were found in the southern region of A<sub>1</sub>. The regions with high SOC contents were mainly occupied by paddy fields or paddy soils with long-term water flooding condition, and the regions with low SOC contents were mainly occupied by uplands or red soils due to the lower SOC input and the more rapid SOC decomposition [43,44]. Similarly, higher SOC contents existed in the east-central and northwest regions, whereas lower values exist in the southern region in A<sub>2</sub>. The high SOC contents were also mainly occupied by paddy fields or paddy soils, and low SOC contents mainly occupied by uplands or red soils.

Furthermore, there were evident differences with the local details of the contours. The contour maps generated from OK all exhibited smooth boundaries and large polygons with small SOC variation ranges, and those from PGr exhibited jagged interlacing and regular square grids with different SOC contents. Comparably, the contours from KSt and KLu were more sophisticated, with smaller polygons and wider SOC variations; the contours from KSt and KLu varied with the soil type and the land-use pattern, respectively. It was easy to see that the polygons of paddy fields and paddy soils all had higher SOC contents, while the polygons of uplands and red soils had lower SOC contents. As the expansion principles from points to surface were different, the contours from the PSt and PLu had some differences in the local details (Figure 3). Considering the diversified soil types and land-use patterns with complicated terrain in the hilly red soil region of China, it was not difficult to determine that the SOC contours from the KSt, KLu, PSt, and PLu could all well reflect the actual SOC distribution characteristics with soil and land-use variation. Relatively, the contours from the OK and PGr were manifestly inconsistent with the actual situation in the hilly red soil region of China.



**Figure 3.** Contours of the SOC content from various prediction methods in: A<sub>1</sub> (a<sub>1</sub>–f<sub>1</sub>); and A<sub>2</sub> (a<sub>2</sub>–f<sub>2</sub>), from (a<sub>1</sub>,a<sub>2</sub>) OK; (b<sub>1</sub>,b<sub>2</sub>) KSt; (c<sub>1</sub>,c<sub>2</sub>) KLu; (d<sub>1</sub>,d<sub>2</sub>) PGr; (e<sub>1</sub>,e<sub>2</sub>) PSt; and (f<sub>1</sub>,f<sub>2</sub>) PLu. Grid resolutions of d<sub>1</sub> and d<sub>2</sub> are 1 km and 2 km, respectively; and the resolutions of the other maps are all 30 m.

### 3.4. Prediction Accuracy Assessments of Various Interpolation Methods

The correlation between the observed and predicted SOC values of the validation samples for various methods in A<sub>1</sub> are shown in Table 4. The  $r$  value from PGr reached as significant level ( $p < 0.05$ ), and the other  $r$  values were all reached very significant level ( $p < 0.01$ ). Among the  $r$  values, the ones from KLu and PLu were 0.87 and 0.85, respectively. The  $r$  values from KSt ( $r = 0.85$ ) and PSt ( $r = 0.81$ ) are both slight lower than KLu and PLu. The  $r$  values from OK ( $r = 0.41$ ) and PGr ( $r = 0.29$ ) were both substantially lower than all of the above methods. This result showed that KLu and PLu had the highest prediction accuracy, closely followed by KSt and PSt. OK had very low accuracy, and PGr had



the lowest accuracy. In A<sub>2</sub>, the corresponding *r* values from various methods had consistent regularity with A<sub>1</sub> although all the *r* values reached very significant level.

The RMSE and MAE values from the validation samples in the two areas are also shown in Table 4. The RMSEs and MAEs from various methods in A<sub>1</sub> were both in the sequence: KLu < PLu < KSt < PSt < OK < PGr. Simultaneously, the same results appeared in A<sub>2</sub> without exception. The predicted errors from OK and PGr were much higher than other four methods, especially the PGr method.

**Table 4.** Predicted RMSE and MAE from various methods in the two study areas.

Areas	Errors (g kg <sup>−1</sup> )	Kriging Methods			Polygon-Based Methods		
		OK	KSt	KLu	PGr	PSt	PLu
A1	<i>r</i> values	0.41 **	0.85 **	0.87 **	0.29 *	0.81 **	0.85 **
	RMSE	6.09	3.78	3.15	7.12	3.91	3.29
	MAE	5.26	3.36	2.71	6.12	3.63	2.85
A2	<i>r</i> values	0.52 **	0.85 **	0.88 **	0.48 **	0.83 **	0.87 **
	RMSE	5.37	3.44	3.01	6.56	3.59	3.13
	MAE	5.04	2.80	2.33	5.83	2.99	2.83

*r* values are the correlation coefficients between the observed and predicted SOC values of the validation samples. \* means a significant level at the 95% confidence level, and \*\* means a very significant level at the 99% confidence level.

## 4. Discussion

### 4.1. Comparison of the Prediction Efficiency for Various Methods

Three kriging and three polygon-based methods were used for obtaining the SOC distribution contours in the red soil hilly region of Southern China. The results showed that their prediction accuracies were different. Among the three kriging methods, the OK method had the largest prediction errors with the strongest smoothing effect due to the lack of considering the influence of land-use and soil types, which have great impact on SOC content (Tables 1 and 2). This smoothening effect will reduce the high SOC content values and increase the low values for the unknown points. Lee et al. also found that using the SOC residual data by removing the mean value of the corresponding soil types to predict SOC distribution could greatly improve the interpolation accuracy in Taiwan's Choshui River area [42], and the results agree with this study. Comparatively, the KLu and KSt recognized the influence and eliminated the smoothing effect to some degree by removing the mean values of the soil types and the land-use patterns. As a result, the parameters of sills in the semi-variogram of the KLu and KSt were lower than OK (Table 3). This indicated that the residual data had better stability and were more conducive to SOC prediction. Meanwhile, the KLu and KSt had a more detailed description for SOC's local change than OK as their parameters of ranges are lower than OK (Figure 3 and Table 4). Consequently, the prediction RMSEs of the KLu and KSt in A<sub>1</sub> increased by 48.3% and 37.9% relative to the OK. Consequently, the prediction RMSEs of the KLu and KSt in A<sub>2</sub> were 43.9% and 35.9% higher than that from the OK, respectively. In comparison, the KLu had higher prediction accuracy than KSt in the two areas.

Regarding the three polygon-based methods, the PGr had the greatest prediction errors in two areas as the sample point in every grid usually poorly represented the variety of soil types and land-use patterns. For example, one grid is mainly occupied by paddy fields, but its central position is occupied by upland or forestry land, so if the sampling point located in the central position was used for representing the whole grid, then the SOC content will be underestimated, and the prediction result will have great uncertainty to reflect the actual SOC distribution. Comparatively, the PSt and PLu considered the difference among the soil types and land-use patterns, respectively, and the SOC content values were linked to the same soil type or land-use polygons. In other words, the SOC value of the sample collected from the red soil was linked to the corresponding red soil polygon to which it belonged [15], and the value of the sample from dryland was linked to the matched dryland polygon. Undoubtedly, this approach can decrease the prediction uncertainty and improve the prediction accuracy (Table 4). As a result, the RMSE values from the PSt and PLu were 53.8% and 45.1% lower

than the PGr in  $A_1$ , and the values from the PSt and PLu in  $A_2$  were 52.3% and 45.3% lower than that of the PGr, respectively. Compared to the PSt, the PLu was more efficient in obtaining the spatial variability information of the SOC in the red soil region.

#### 4.2. Impacts of the Method Selection on SOC Prediction

The above comparative analysis shows that the prediction accuracies of different interpolation methods are quite different. The selection of prediction methods directly affects the revealing of regional SOC variation. For example, Huang et al. used the OK for obtaining the SOC distribution in three periods, and calculated the SOC changing rates in the three periods in Rugao County of Jiangsu Province, China [45]. However, the results of this study show that the SOC variability information obtained by the OK has large prediction uncertainty in each period, and the uncertainty of the SOC evolution will be amplified by the interpolation error accumulating in these periods. Thus, the result's reliability of the derived SOC temporal evolution must be reconsidered. Another example, Yu et al. used grids with various cell resolutions ranging from  $0.5 \text{ km} \times 0.5 \text{ km}$  to  $64 \text{ km} \times 64 \text{ km}$  to estimate  $\text{CH}_4$  emissions in the Tai-Lake region of China by the DNDC model [19]. However, the results of this study pointed out that the PGr method was not a compatible selection as the soil sample usually has weak representation to the sampling grid. This view has also been pointed out by other scholars [27,33]. Comparatively, the kriging and polygon-based methods combining with soil type and land-use information can improve the prediction accuracies greatly (Table 4). Furthermore, the comparative analysis showed that the kriging and polygon methods combining with land-use information had better performance than those combining with soil type information. This is because that land-use has a larger influence on the SOC variability than soil type (Table 3). The above analysis indicates that the selection of reasonable interpolation method is very important for SOC prediction and simulation in hilly red soil region. It is necessary to pay more attention to this issue.

### 5. Conclusions

The prediction efficiencies of three kriging (OK, KSt, and KLu) and three polygon-based (PGr, PSt, and PLu) methods were compared in two coverage areas in Yujiang County in Jiangxi Province, China. The results of the two areas were consistent and showed that the OK and PGr methods were both not reasonable selections for SOC prediction and simulation in the hilly red soil region of China due to their large uncertainties. In contrast, the KLu, KSt, PLu, and PSt all had relatively high prediction accuracies when considering the influence of soil type or land-use on the SOC variability. Furthermore, the KLu and PLu methods had the highest accuracies and were slightly higher than KSt and PSt, respectively. The research results indicate that, if the SOC distribution needs to be predicted and simulated in the red soil hilly region of China, the selection of interpolation method has great impact on the predicted and simulated results. This issue should arouse the attention of scholars. The findings of this research are applicable to the hilly red soil region in China, even those with similar soil regions.

**Author Contributions:** Z.Z. devised the research questions and analytical strategy. D.Y., X.W., Y.P., G.Z. and X.S. collected the questionnaire data. All authors collaborated on the interpretation of the results and on writing and revising the paper.

**Funding:** This research was funded by the National Natural Science Foundation of China (Nos. 41201213 and 41571206), and the State Key Laboratory of Soil and Sustainable Agriculture (No. Y20160008).

**Conflicts of Interest:** The authors declare no conflict of interest.

### References

1. Fang, J.; Chen, A.; Peng, C.; Zhao, S.; Ci, L. Changes in forest biomass carbon storage in China between 1949 and 1998. *Science* **2001**, *292*, 2320–2322. [[CrossRef](#)] [[PubMed](#)]
2. Lal, R. Soil carbon sequestration to mitigate climate change. *Geoderma* **2004**, *123*, 1–22. [[CrossRef](#)]

3. Ding, X.; Han, X.; Liang, Y.; Qiao, Y.; Li, L.; Li, N. Changes in soil organic carbon pools after 10 years of continuous manuring combined with chemical fertilizer in a Mollisol in China. *Soil Tillage Res.* **2012**, *122*, 36–41. [[CrossRef](#)]
4. Iwasaki, S.; Endo, Y.; Hatano, R. The effect of organic matter application on carbon sequestration and soil fertility in upland fields of different types of Andosols. *Soil Sci. Plant Nutr.* **2017**, *63*, 200–220. [[CrossRef](#)]
5. Curtin, D.; Beare, M.H.; Hernandez-Ramirez, G. Temperature and moisture effects on microbial biomass and soil organic matter mineralization. *Soil Sci. Soc. Am. J.* **2012**, *76*, 2055–2067. [[CrossRef](#)]
6. Wiaux, F.; Vanclooster, M.; Cornelis, J.T.; Van Oost, K. Factors controlling soil organic carbon persistence along an eroding hillslope on the loess belt. *Soil Biol. Biochem.* **2014**, *77*, 187–196. [[CrossRef](#)]
7. Álvaro-Fuentes, J.; Easter, M.; Paustian, K. Climate change effects on organic carbon storage in agricultural soils of northeastern Spain. *Agric. Ecosyst. Environ.* **2012**, *155*, 87–94. [[CrossRef](#)]
8. Cheng, W.; Padre, A.T.; Sato, C.; Shiono, H.; Hattori, S.; Kajihara, A.; Aoyama, M.; Tawarayama, K.; Kumagai, K. Changes in the soil C and N contents, C decomposition and N mineralization potentials in a rice paddy after long-term application of inorganic fertilizers and organic matter. *Soil Sci. Plant Nutr.* **2016**, *62*, 212–219. [[CrossRef](#)]
9. Tsui, C.; Guo, H.; Chen, Z. Estimation of soil carbon stock in Taiwan arable soils by using legacy database and digital soil mapping. In *Soil Processes and Current Trends in Quality Assessment*; Soriano, M.C.H., Ed.; Intech: Rijeka, Croatia, 2013. [[CrossRef](#)]
10. Hoffmann, U.; Hoffmann, T.; Jurasinski, G.; Glatzel, S.; Kuhn, N.J. Assessing the spatial variability of soil organic carbon stocks in an alpine setting (Grindelwald, Swiss Alps). *Geoderma* **2014**, *232*, 270–283. [[CrossRef](#)]
11. Luis, R.L.; Antonio, M.C. Modelling and mapping organic carbon content of topsoils in an Atlantic area of southwestern Europe (Galicia, NW-Spain). *Geoderma* **2015**, *245*, 65–73. [[CrossRef](#)]
12. Sun, L. Climatology, trend analysis and prediction of sandstorms and their associated dustfall in China. *Water Air Soil Pollut. Focus* **2003**, *3*, 41–50. [[CrossRef](#)]
13. Xia, X.; Yang, Z.; Liao, Y. Temporal variation of soil carbon stock and its controlling factors over the last two decades on the southern Song-nen Plain, Heilongjiang Province. *Geosci. Front.* **2010**, *1*, 125–132. [[CrossRef](#)]
14. Eswaran, H.; Vandenberg, E.; Reich, P. Organic-carbon in soils of the world. *Soil Sci. Soc. Am. J.* **1993**, *57*, 192–194. [[CrossRef](#)]
15. Zhao, Y.; Shi, X.; Weindorf, D.; Yu, D.; Sun, W.; Wang, H. Map scale effects on soil organic carbon stock estimation in North China. *Soil Sci. Soc. Am. J.* **2006**, *70*, 1377–1386. [[CrossRef](#)]
16. Zhang, Y.; Zhao, Y.; Shi, X.; Lu, X.; Yu, D.; Wang, H.; Sun, W.; Darilek, J. Variation of soil organic carbon estimates in mountain regions: A case study from Southwest China. *Geoderma* **2008**, *146*, 449–456. [[CrossRef](#)]
17. Mishra, U.; Lal, R.; Slater, B.; Calhoun, F.; Liu, D.S.; Van Meirvenne, M. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Sci. Soc. Am. J.* **2009**, *73*, 614–621. [[CrossRef](#)]
18. Kerry, R.; Goovaerts, P.; Rawlins, B.G.; Marchant, B.P. Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma* **2012**, *170*, 347–358. [[CrossRef](#)] [[PubMed](#)]
19. Yu, D.; Zhang, L.; Shi, X.; Warner, E.D.; Zhang, Z.; Zhao, Q. Soil assessment unit scale affects quantifying CH<sub>4</sub> emissions from rice fields. *Soil Sci. Soc. Am. J.* **2013**, *77*, 664–672. [[CrossRef](#)]
20. Davidson, E.A.; Lefebvre, P.A. Estimating regional carbon stocks and spatially covarying edaphic factors using soil maps at three scales. *Biogeochemistry* **1993**, *22*, 101–131. [[CrossRef](#)]
21. Marchetti, A.; Piccini, C.; Francaviglia, R.; Mabit, L. Spatial distribution of soil organic matter using geostatistics: a key indicator to assess soil degradation status in central Italy. *Pedosphere* **2012**, *22*, 230–242. [[CrossRef](#)]
22. Darilek, J.L.; Huang, B.; Wang, Z.; Qi, Y.; Zhao, Y.; Sun, W.; Gu, Z.; Shi, X. Changes in soil fertility parameters and the environmental effects in a rapidly developing region of China. *Agric. Ecosyst. Environ.* **2009**, *129*, 286–292. [[CrossRef](#)]
23. Post, W.M.; Emanuel, W.R.; Zinke, P. Soil carbon pools and world life zone. *Nature* **1982**, *298*, 156–159. [[CrossRef](#)]
24. Batjes, N.H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil. Sci.* **1996**, *47*, 151–163. [[CrossRef](#)]
25. Ni, J. Carbon storage in terrestrial ecosystems of China: Estimates at different spatial resolutions and their responses to climate change. *Clim. Chang.* **2001**, *49*, 339–358. [[CrossRef](#)]

26. Minasny, B.; Mcbratney, A.B. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma* **2007**, *140*, 324–336. [[CrossRef](#)]
27. Chai, X.; Shen, C.; Yuan, X.; Huang, Y. Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. *Geoderma* **2008**, *148*, 159–166. [[CrossRef](#)]
28. Sumfleth, K.; Duttman, R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecol. Indic.* **2008**, *8*, 485–501. [[CrossRef](#)]
29. Zhang, Z.; Yu, D.; Shi, X.; Warner, E.; Ren, H.; Sun, W.; Tan, M.; Wang, H. Effect of sampling classification patterns on SOC variability in the red soil region, China. *Soil Tillage Res.* **2010**, *110*, 2–7. [[CrossRef](#)]
30. Addis, H.K.; Klik, A.; Strohmeier, S. Performance of frequently used interpolation methods to predict spatial distribution of selected soil properties in an agricultural watershed in Ethiopia. *Appl. Eng. Agric.* **2016**, *32*, 617–626. [[CrossRef](#)]
31. Kumar, S.; Singh, R.P. Spatial distribution of soil nutrients in a watershed of Himalayan landscape using terrain attributes and geostatistical methods. *Environ. Earth Sci.* **2016**, *75*, 473. [[CrossRef](#)]
32. Cao, Z.; Zhou, J. *Soil Quality of China*; Science Press of China: Beijing, China, 2008.
33. Yu, D.; Zhang, Z.; Yang, H.; Shi, X.; Tan, M.; Sun, W.; Wang, H. Effect of sampling density on detecting the temporal evolution of soil organic carbon in hilly red soil region of South China. *Pedosphere* **2011**, *21*, 207–213. [[CrossRef](#)]
34. Li, Z.; Zhang, T.; Chen, B. Changes of organic carbon and nutrient contents in highly productive paddy soils in Yujiang County of Jiangxi Province. *Sci. Agric. Sin.* **2006**, *39*, 324–330.
35. IUSS Working Group WRB. *World Reference Base for Soil Resources 2006*; World Soil Resources Reports No. 103; Food and Agriculture Organization of United Nations: Rome, Italy, 2006; pp. 71–72.
36. Shi, X.; Yu, D.; Xu, S.; Warner, E.D.; Wang, H.; Sun, W.; Zhao, Y.; Gong, Z. Cross-reference for relating Genetic Soil Classification of China with WRB at different scales. *Geoderma* **2010**, *155*, 344–350. [[CrossRef](#)]
37. Soil Survey Office of Yujiang County. *Soil Species of Yujiang County*; Science and Technology Press: Beijing, China, 1986.
38. Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon, and organic matter. In *Methods of Soil Analysis, Part 2-Chemical and Microbiological Properties*; Page, A.L., Miller, R.H., Keeney, D.R., Eds.; ASA-SSSA: Madison, WI, USA, 1982; pp. 539–594.
39. Kerry, R.; Oliver, M.A. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma* **2007**, *140*, 383–396. [[CrossRef](#)]
40. Zhang, Z.; Shi, X.; Yu, D.; Weindorf, D.; Sun, W.; Wang, H.; Zhao, Y. Effect of prediction methods on detecting the temporal evolution of soil organic carbon in hilly red soil region of South China. *Environ. Earth Sci.* **2011**, *64*, 319–328. [[CrossRef](#)]
41. Qiu, J.; Wang, L.; Li, H.; Tang, H.; Li, C.; Van Ranst, E. Modeling the Impacts of soil organic carbon content of croplands on crop yields in China. *Agric. Sci. China* **2009**, *8*, 464–471. [[CrossRef](#)]
42. Liu, T.; Juang, K.; Lee, D. Interpolating soil properties using kriging combined with categorical information of soil maps. *Soil Sci. Soc. Am. J.* **2006**, *70*, 1200–1209. [[CrossRef](#)]
43. Lal, R.; Kimble, J.M. Conservation tillage for carbon sequestration. *Nutr. Cycl. Agroecosyst.* **1997**, *49*, 243–253. [[CrossRef](#)]
44. Liu, D.; Wang, Z.; Zhang, B.; Song, K.; Li, X.; Li, J.; Li, F.; Duan, H. Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. *Agric. Ecosyst. Environ.* **2006**, *113*, 73–81. [[CrossRef](#)]
45. Huang, B.; Sun, W.; Zhao, Y.; Zhu, J.; Yang, R.; Zou, Z.; Ding, F.; Su, J. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. *Geoderma* **2007**, *139*, 336–345. [[CrossRef](#)]

