Determination of Optimal Sampling Numbers to Investigate the Soil Organic Matter in a Typical County of the Yellow River Delta, China

Wenjing Wang 1,2, Mengqi Duan 1,2, Xiaoguang Zhang 1,2,* , Xiangyun Song 1,2, Xinwei Liu 1,2,* and Dejie Cui 1,2

1 College of Resources and Environment, Qingdao Agricultural University, Qingdao 266109, China; sxzzw@163.com (W.W.); duanmengqi@163.com (M.D.); xsong@qau.edu.cn (X.S.); cuidejie@163.com (D.C.)
2 Qingdao Agricultural Remote Sensing Application Engineering Research Center, Qingdao 266109, China
* Correspondence: zhangxg_666@163.com (X.Z.); sxdxw@163.com (X.L.); Tel.: +86-151-9200-3056 (X.Z.);
+86-158-6305-6712 (X.L.)

Abstract: Soil organic matter (SOM) plays a crucial role in promoting soil tillage, improving soil fertility and providing crop nutrients. Investigation and sampling are the premise and basis for understanding the spatial distribution of SOM. The number of sampling points will affect the accuracy of spatial variation of SOM. Therefore, it is important scientific work to determine a reasonable number of sampling points under the premise of ensuring accuracy. In this study, Kenli County, a typical area of the Yellow River Delta in China, was taken as an example to investigate the effect of different sampling points on spatial-variation expression of SOM. A total of 12 sample subsets (including 900 samples) were randomly sampled at equal intervals from the 900 sample points, using geographic information system (GIS) technology and geostatistical analyses to explore the optimal number of samples. The results showed that the SOM content in the study area had a lower-middle degree of variation. As the number of sample points decreased, the spatial distribution of SOM showed the gradual weakening of detail-characterization ability; and when the number of sample points was too small (<100), there was a wrong expression that was not consistent with the actual situation. The value of RMSE has no obvious regularity with the change of sample number. The values of both ME and ASE showed a significant inflection point when the number of samples was 150 and remained around 0 and 4 as the number of samples increased, respectively. Combined with the three indicators of ME, RMSE and ASE, collecting at least 150 samples can satisfy the spatial-variation expression of SOM, equivalent to 107 sample points within the area of 1000 km². The research results could provide important references for investigation of SOM content in areas with similar natural geographical conditions.

Keywords: soil attribute; GIS; ordinary Kriging; rational sampling numbers; spatial heterogeneity

1. Introduction

Soil organic matter (SOM) is one of the important components of soil [1]. It has the functions of providing crop nutrients, improving soil cultivability and promoting microbial activities, and it plays an important role in the quality of soil fertility. Due to the long-term influence of natural factors such as parent material, topography, climate and biology, as well as the intervention of human factors such as irrigation, fertilization and farming, the distribution of SOM within a certain area will show corresponding spatial differences [2,3].

For spatial-variation analysis of regional soil properties, spatial prediction and evaluation are often carried out through soil survey sampling and indoor laboratory assays. The accuracy of spatial prediction is related to the layout and density of sampling points. Theoretically, under the same spatial-distribution mode, the more sampling numbers and...
the greater the density, the better the spatial-prediction results and the higher the accuracies. Many studies have also confirmed this claim that the spatial-prediction accuracy of SOM exhibits a tendency to increase with the number of sample points [4–6]. Nevertheless, in practice, too high a sampling density can cause a waste of manpower, material resources and financial resources [7,8]. Moreover, some studies believe that blindly increasing the number of samples may not always improve the accuracy of spatial prediction [9]. Therefore, setting a reasonable number of sampling points within the region is of great significance for saving sampling costs, improving sampling efficiency and spatial-prediction accuracy and achieving efficient and sustainable utilization of soil resources.

In view of the above problems, some scientists have conducted a series of studies on the impact of soil-sampling number on the spatial variability of soil properties in different regions. When performing the spatial prediction of SOM in Guangdong Province, China, Lai et al. found that regression Kriging combined with 800 soil-sampling points could economically and accurately provide spatial-distribution information of SOM, equivalent to collecting 5 samples per 1000 km² [10]. Marques Jr et al. showed that during a spatial-variation study against micronutrients and aluminum in the state of Sao Paulo, Brazil, based on standardized variograms, the minimum number of sampling points available for this region is 400 (equivalent to 2 sample points collected per 1000 km²) [11]. In addition, they believed that the minimum sampling spacing varied across topographic units. Long et al. took Fujian Province with an area of 124,000 km² in Southeast China as the study area [6]. They found that a reasonable number of samples per 1000 square kilometer were 11,000, 10,000 and 9000, respectively, for SOM sampling studies in different terrain areas such as valley–basin, hill–mountain, and plain–platform. Moreover, this study also showed that the spatial interpolation accuracy of the SOM was more sensitive to the sampling density under relatively simple topographic conditions. Wang et al. in typical coal-mining areas at the Loess Plateau showed that 40 soil-sampling points could guarantee the accuracy of SOM and total nitrogen spatial expression in the region (equivalent to 90,900 sample points collected per 1000 km² [5]. Furthermore, in the other two studies, Pang et al. and Zhang et al. analyzed the effect of sampling density changes on spatial-interpolation accuracy, using soil copper and SOM, respectively [12,13]. The results of both suggested that increasing sampling density and selection of reasonable interpolation methods facilitated accurate estimation of spatial variation in soil properties. These studies indicate that the number of reasonable sampling points required in different regional geographical environment conditions were different. Therefore, seeking an accurate and economical optimal sampling number of SOM in specific natural geographical conditions is still an issue that we currently need to pay attention to.

The Yellow River Delta (about 5400 km²) is a region of high soil salinity along the eastern coastal areas of China, whose unique natural geographical conditions determine the uniqueness of SOM changes [14,15]. Therefore, based on soil sampling data, this paper conducts a study in Kenli County located in the Yellow River Delta, to explore effects of different sampling numbers on the spatial variability expression of soil organic matter and to further clarify the optimal number of sampling points for predicting the spatial distribution of SOM in the estuarine plain.

2. Materials and Methods

2.1. Study Area

Kenli County is located in the Yellow River Delta region in the northeast of Shandong Province, China (118°15′–119°19′ E, 37°24′–38°10′ N), affiliated with Dongying City. It has jurisdiction over 7 townships with a land area of about 2331 km² (Figure 1). Due to the frequent swing of the tail section of the Yellow River in history, a typical delta landform has been formed, and the terrain is slightly inclined from southwest to northeast. The county is located in the lower reaches of the Yellow River and has low terrain (most areas between 6 m–8 m), coupled with high groundwater level, and is affected by seawater infiltration, the soil salinization within the county being more serious [16]. The main soil types in
Kenli County are fluvo-aquic soil and coastal saline soil, and the corresponding soil group names of WRB are Cambisols and Solonchaks, respectively. The soil texture is mainly sandy loam, and the texture of the plough layer is sandy loam, light loam, medium loam, heavy loam and clay, among which sandy loam and light loam are the most widely distributed, accounting for more than 70% of the area. The county has a temperate monsoon climate but is significantly affected by the continental monsoon. The northwest wind prevails in winter and the southeast wind prevails in summer. The local crops mainly include winter wheat, corn, rice and cotton [17–19].

Figure 1. Location and elevation of Kenli County in Yellow River Delta.

2.2. Soil Sampling and Laboratory Analysis

Firstly, a total of 1000 sample points were uniformly deployed according to the pattern of the grid distribution, and then samples of 0~20 cm tillage layer were collected. In the actual sampling process, the number of samples in each grid was determined according to the main crop types in the grid. When a crop type was in the grid, we sampled at the center of the plot. When there were multiple crop types in the grid, we selected the plot with the larger type for sampling. At the same time, some sample points were slightly adjusted by referencing the land use and road traffic situation. The central plot of the sampling site was sampled by “S shape”, and 5~10 points were collected from each site and fully mixed. 1 kg soil samples were retained by quartering method and loaded into the sample bag for soil
analysis. Global Positioning System (GPS) was used to precisely locate each sampling site. The eastern area was not sampled because of conservation policy from the Yellow River Delta Nature Reserve. Finally, a total of 926 soil samples were collected in fall 2009 before and after crop harvest, and the sampling area was approximately 1400 km².

The collected soil samples were taken back to the laboratory for air drying, grinding and passing through a 0.25 mm nylon sieve for analysis. The soil organic carbon (SOC) was determined by the K₂Cr₂O₇ oxidation-titration method [20]. The content of SOM was obtained by multiplying the content of SOC by the van Benmmelen coefficient (1.724).

2.3. Data Processing and Analysis

Some outliers were removed by adding and subtracting 3 times standard deviation and yielding 900 SOM samples after removing the outliers. For analyzing the effect of different sampling-point numbers on spatial prediction, 800, 700, 600, 500, 400, 300, 200, 150, 100, 75 and 50 samples were extracted from 900 samples according to the random-sampling method at equal intervals. Adding the original set of 900 sample points, 12 different set series of sample points were formed. These sets were used to study the spatial-distribution characteristics of SOM under different sampling numbers and to study the suitable number of sampling points within the study area.

The SPSS13.0 was used to conduct descriptive statistical analysis of SOM content in 12 sample sets, which obtained multiple statistical features including mean value, median value, standard deviation, skewness coefficient, kurtosis coefficient and coefficient of variation. K-S test was performed to verify whether the SOM data conformed to normal distribution.

In predicting the spatial distribution of SOM, the semivariance model is needed to infer its spatial variation structure. In this paper, the sampling point data were calculated to obtain the theoretical model of the semivariance function by GS+ 7.0 software, and further to reflect the proportion of the random component to the structural component in the spatial variation through the Nugget ratio value.

2.4. Spatial Prediction and Validation

Expansion from point to surface was performed using the ordinary Kriging interpolation method in ArcGIS10.0, and the spatial distribution of SOM was obtained in different numbers of point.

To measure the accuracy of the SOM spatial-interpolation results at different numbers of point, validation of the spatial-interpolation results was needed. The common verification methods include cross-validation and independent validation. Because cross-validation is simple and fast, some scholars chose this method to validate the result, namely excluding a sample point and then using remaining points to predict the value on this position [21–23]. However, cross-validation could not accurately describe the prediction error of spatial interpolation in many cases, so the independent validation method was selected in this study. After extracting 800 samples from the entire 900 sample-point sets, the remaining 100 samples were taken as the validation dataset for verifying the spatial-interpolation results of the 12 sample-point sets, respectively. The mean error (ME), root-mean-square error (RMSE) and average standard error (ASE) were selected to measure the accuracy of spatial prediction. The closer the absolute value of the ME is to 0, the smaller RMSE; and the closer the RMSE to the ASE, the higher the accuracy of spatial prediction. The calculation formulas for each verification index are as follows:

\[
ME = \frac{\sum_{i=1}^{n} [z(x_i) - z^*(x_i)]}{n} \tag{1}
\]

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n} [z(x_i) - z^*(x_i)]^2}{n}} \tag{2}
\]
\[ ASE = \sqrt{\frac{\sum_{i=1}^{n} \sigma(x_i)}{n}} \]  
(3)

where \(z(x_i)\) = SOM observations, \(z^*(x_i)\) = SOM predictions, \(\sigma(x_i)\) = prediction standard error at the point \(x_i\) and \(n\) = the number of sampling points in the validation dataset (in this paper, \(n = 100\)).

3. Results

3.1. Descriptive Statistics of SOM

From the descriptive statistics of the SOM content at all sample points (Table 1), the minimum and maximum SOM content in the study area were 5.00 g/kg and 27.30 g/kg, respectively, and the average content was 10.96 g/kg, which belonged to the lower-middle level of soil fertility. The coefficient of variation (CV) was applied to clarify the total heterogeneity of the variables. According to the criteria proposed by Nielsen and Bouma, the coefficient of variation can be divided into low variability (CV < 0.1), moderate variability (CV 0.1–1) and high variability (CV > 1) [24]. The coefficient of variation was 35.80%, so it had a lower-middle degree of variation, indicating the small fluctuation and dispersion degree of the SOM content at the sampling site. This phenomenon further revealed that there is no particularly high variation of SOM in this study area, which may be related to the overall location of Kenli County in the Yellow River Delta and little difference in topography. Moreover, the parent material of the Yellow River Delta was transported from the Loess Plateau by the Yellow River. Under the both influence of the Yellow River water and sea water, similar fluvo-aquic soil and saline soil were formed. Thus, the same parent material type and similar soil type may also be responsible for the low degree of variability.

Table 1. Descriptive statistics of SOM content in different sampling number.

| Number of Samples | Minimum (g/kg) | Maximum (g/kg) | Mean (g/kg) | Standard Deviation (g/kg) | Skewness | Kurtosis | Median (g/kg) | Coefficient of Variation(%) |
|-------------------|---------------|----------------|-------------|---------------------------|----------|----------|---------------|----------------------------|
| 900               | 5.00          | 27.30          | 10.96       | 3.93                      | 0.98     | 4.13     | 10.30         | 35.80                      |
| 800               | 5.00          | 27.30          | 10.89       | 3.99                      | 1.05     | 4.30     | 10.20         | 36.60                      |
| 700               | 5.00          | 27.30          | 10.93       | 4.05                      | 1.06     | 4.33     | 10.25         | 37.05                      |
| 600               | 5.00          | 27.30          | 10.97       | 3.99                      | 0.96     | 4.09     | 10.35         | 36.40                      |
| 500               | 5.00          | 25.50          | 10.99       | 3.87                      | 0.86     | 3.76     | 10.40         | 35.18                      |
| 400               | 5.00          | 24.60          | 11.10       | 3.80                      | 0.67     | 3.25     | 10.60         | 34.27                      |
| 300               | 5.00          | 24.60          | 11.22       | 3.81                      | 0.70     | 3.41     | 10.80         | 33.99                      |
| 200               | 5.00          | 24.60          | 11.23       | 3.89                      | 0.82     | 3.65     | 10.80         | 34.63                      |
| 150               | 5.00          | 22.20          | 11.03       | 3.79                      | 0.70     | 3.20     | 10.60         | 34.38                      |
| 100               | 5.20          | 24.60          | 11.49       | 4.18                      | 0.88     | 3.47     | 10.60         | 36.42                      |
| 75                | 5.20          | 24.60          | 11.61       | 4.33                      | 0.90     | 3.45     | 10.80         | 37.29                      |
| 50                | 5.80          | 24.60          | 11.56       | 3.99                      | 0.99     | 3.93     | 10.75         | 34.49                      |

From the extracted 11 subsets of sample points (800, 700, 600, 500, 400, 300, 200, 150, 100, 75 and 50), the maximum value of SOM at 150 samples was 23.20 g/kg, slightly smaller than the maximum at other samples, and the minimum value of SOM at 50 sample points was 5.80 g/kg, slightly greater than the minimum value of other sample points. Furthermore, for each subset, the minimum, maximum, mean, standard deviation, coefficient of variation and other statistical indicators of SOM content were closely similar to the results at 900 sample points. This indicated that the sample subsets selected in this study can still be highly representative of the whole and could essentially satisfy the requirements for the overall descriptive estimation of SOM content in Kenli County.

The mean values of all sample sets were distributed between 10.89 g/kg–11.61 g/kg, and the median values were between 10.20 g/kg–10.80 g/kg. The difference between the mean and the median value of each sample set was not large. The skewness coefficients of each sample-point set were generally distributed between 0.67–1.06, and the K-S test showed that the SOM content data under different numbers of sample points were in a moderate-skew distribution state. After logarithmic transformation, it obeyed the normal
distribution (Table 2). The histogram of the normal distribution for 900 sample points is listed (Figure 2).

| Number of Samples | Minimum / (g/kg) | Maximum / (g/kg) | Mean / (g/kg) | Standard Deviation / (g/kg) | Skewness | Kurtosis | Median / (g/kg) | Coefficient of Variation(%) |
|-------------------|------------------|------------------|---------------|-----------------------------|----------|----------|-----------------|---------------------------|
| 900               | 1.61             | 3.31             | 2.33          | 0.35                        | 0.10     | 2.53     | 2.33            | 14.85                     |
| 800               | 1.61             | 3.31             | 2.33          | 0.35                        | 0.15     | 2.55     | 2.32            | 15.09                     |
| 700               | 1.61             | 3.31             | 2.33          | 0.35                        | 0.16     | 2.54     | 2.33            | 15.23                     |
| 600               | 1.61             | 3.31             | 2.33          | 0.35                        | 0.07     | 2.48     | 2.34            | 15.19                     |
| 500               | 1.61             | 3.24             | 2.34          | 0.35                        | 0.01     | 2.49     | 2.34            | 14.81                     |
| 400               | 1.61             | 3.20             | 2.35          | 0.34                        | −0.10    | 2.42     | 2.36            | 14.68                     |
| 300               | 1.61             | 3.20             | 2.36          | 0.34                        | −0.13    | 2.53     | 2.38            | 14.51                     |
| 200               | 1.61             | 3.20             | 2.36          | 0.34                        | −0.03    | 2.56     | 2.38            | 14.52                     |
| 150               | 1.61             | 3.14             | 2.34          | 0.35                        | 0.10     | 2.43     | 2.36            | 15.12                     |
| 100               | 1.65             | 3.20             | 2.38          | 0.35                        | 0.10     | 2.43     | 2.36            | 14.89                     |
| 75                | 1.65             | 3.20             | 2.39          | 0.36                        | 0.11     | 2.45     | 2.38            | 15.17                     |
| 50                | 1.76             | 3.20             | 2.39          | 0.33                        | 0.23     | 2.47     | 2.37            | 13.76                     |

Figure 2. Histogram of normal distribution of SOM at 900 sample points after logarithmic transformation.

3.2. Characteristics of the Spatial-Variation Structure of SOM

The spatial-variability structure of the SOM can be fitted using the semivariance function. The ratio of $C_0$ to $(C_0 + C)$ is nugget coefficients, which indicate the proportion of the random components in the spatial-variation structure. The higher the nugget coefficient, the weaker the structural components and the stronger the random components [25]. By fitting the semivariance function of SOM under different sampling points, it was found that when the number of samples was large, the system could also greatly satisfy the spatial autocorrelation. Taking a subset of 800 sample points as an example (Table 3), the determining coefficient ($R^2$) of the semivariance function was 0.22. The nugget value was 9.09 and the sill value was 13.02. The nugget coefficient was reached at 69.81%, indicating that the system also had a moderate degree of autocorrelation. However, when the number of sample points decreased, the overall fitting accuracy of the semivariance function of each subset was low; the $R^2$ was mostly below 0.22. The value of nugget coefficient was overall higher, mostly close to 80%, and the value of nugget coefficient did not change much with
the decrease in sample numbers. Therefore, we are not going to discuss the impact of the number of points on the spatial variation structure in SOM.

Table 3. The semivariance model of SOM content at 800 sample points and its fitting parameters.

| Number of Samples | Model Type | Nugget ($C_0$) | Sill ($C_0 + C$) | Nugget Ratio ($C_0$/sill) | Range/(m) | $R^2$ | RS |
|-------------------|------------|-----------------|-----------------|-------------------------|----------|-------|----|
| 800               | Spherical  | 9.09            | 13.02           | 69.81%                  | 2780.00  | 0.22  | 14.60 |

3.3. Effect of Different Sampling-Point Numbers on the Prediction Accuracy of SOM Content

The closer the ME approaches 0, the higher the spatial prediction accuracy of SOM. As can be seen from Figure 3, when the number of sampling points was greater than 150, the value of ME were almost all close to 0 and varied little. The ME of the predicted value reached a minimum of $-0.12$ when the number of samples was 150. The value of the ME increased significantly when the number of samples was less than 150, and the ME of the predicted values at both 100 and 75 was approximately close to 1. This showed that when the number of sample points was greater than or equal to 150, Kriging interpolation method could better preserve the accuracy on the spatial prediction.

Figure 3. The spatial prediction error of SOM in different number of sampling points.

The smaller the value of RMSE, the higher the spatial prediction accuracy of the SOM. There was no obvious change trend in RMSE with the increase in sampling numbers, and all values fluctuated between 3.60 and 4.65. The value of RMSE reached the maximum when the number of sample points was 50; when the number of sample points changed from 75 to 400, the value of RMSE was relatively stable and essentially maintained at around 3.7; but when the number of sample points was greater than 400, the value of RMSE showed a certain degree of volatility. However, it did not directly explain the relationship between the
increase or decrease in sample points and the spatial-prediction error of SOM. This showed that the single verification index of RMSE did not indicate the result of spatial-prediction accuracy.

ASE is an indicator used to evaluate the variability of a predicted value. If the value of ASE is equal to the value of RMSE, it means that the Kriging interpolation correctly estimates the spatial variation of SOM. The variability of spatial prediction is overestimated if ASE is greater than RMSE, and underestimated if ASE is less than RMSE. As can be seen from Figure 3, when the number of sampling points was less than 150, the values of ASE were mostly greater than 4, which were much higher than the value of RMSE. The large difference between the ASE and RMSE indicated that the variability of SOM spatial prediction was overestimated at this time, and the spatial prediction results had great uncertainty. When the number of sampling points was between 150 and 800, the value of ASE dropped to around 4.0 and remained relatively stable, and the difference between ASE and RMSE was relatively small in general. It reflected that the spatial prediction of the Kriging interpolation method was relatively stable and the prediction results were relatively accurate when the number of samples was more than or equal to 150.

Based on the above three verification indexes of spatial-prediction accuracy, it can be inferred that collecting at least 150 soil samples could meet the demand of spatial-variation expression for SOM within the Kenli County of the Yellow River Delta. It was equivalent to at least 107 sample points to be set on every 1000 km$^2$ area.

3.4. Effect of Sampling-Point Numbers on the Expression of Spatial Distribution of SOM Content

In order to further understand the effect of sampling-point numbers on the spatial-distribution characteristics of SOM content in the study area, an ordinary Kriging interpolation method was conducted for the 12 sample-point sets to obtain the spatial-distribution maps of SOM at different sampling-point densities (Figure 4). Since the sampling site did not cover the Yellow River Delta National Nature Reserve in the east of the study area, we did not analyze the eastern region when studying the spatial variability of the SOM.

According to Figure 4, when the numbers of samples were between 400 and 900, the SOM in the study area showed a similar spatial distribution pattern: soils with a medium content of SOM (9 g/kg~12 g/kg) occupied most of the study area; soils with relatively low content of SOM (<9 g/kg) were distributed in the southern part of the study area with the form of spots; while soils with relatively high SOM content (12 g/kg~14 g/kg) were mainly distributed in the central and northern parts with the form of discontinuous patches. Because of the smooth effect of the Kriging interpolation method, soils with SOM content greater than 14 g/kg formed narrow band regions with high value in the northeast when the number of sample points were 500 and 600. When the number of sample points dropped to 300~150, the ability to depict details was slightly weakened. Furthermore, when the blocky low-value area in the southern part of the study area was no longer obvious, the overall trend of SOM spatial variation could still be kept largely unchanged. When the sampling point number was 100, the distribution proportion of soils with relatively high SOM content (12 g/kg to 14 g/kg) was increased significantly within the study area. A large area of soils with relatively high SOM content appeared in the southern part where the SOM content was actually low, and the distribution continuity of the patches also increased. When the number of sample points dropped to 75, the SOM spatial-distribution information with the highest or lowest value could not displayed in the map. The expression information of SOM spatial distribution was relatively simple, and many details were difficult to be described. When only 50 sample points were left, there was a marked local overestimation in the southern area. Combined with the case of 100 and 75 sampling-point sets, it can be shown that when the number of sample points was less than 150, the Kriging interpolation method overestimated the variability of SOM spatial prediction, especially in the truly low-value region. These characteristics of the spatial expressions were consistent with the performance of spatial-prediction errors.
Based on the above three verification indexes of spatial prediction accuracy, it can be inferred that collecting at least 150 soil samples could meet the demand of spatial variation expression for SOM within the Kenli County of the Yellow River Delta. It was equivalent to at least 107 sample points to be set on every 1000 km² area.

3.4. Effect of Sampling Point Numbers on the Expression of Spatial Distribution of SOM Content

In order to further understand the effect of sampling-point numbers on the spatial-distribution characteristics of SOM content in the study area, an ordinary Kriging interpolation method was conducted for the 12 sample-point sets to obtain the spatial-distribution maps of SOM at different sampling-point densities (Figure 4). Since the sampling site did not cover the Yellow River Delta National Nature Reserve in the east of the study area, we did not analyze the eastern region when studying the spatial variability of the SOM.

Overall, with the decrease in sample numbers, the spatial distribution of SOM showed the gradually weakened ability of detail characterization. Moreover, with the further reduction of the number of sample points, the predicted spatial distribution of SOM by Kriging interpolation might also have appeared to be a false characterization that was inconsistent with the actual situation. Combined with the independent verification index of SOM spatial prediction, when predicting the spatial distribution of SOM in Kenli County, it essentially required at least 150 sample points to ensure the accuracy of spatial prediction and the correctness of SOM spatial-distribution trend.

4. Discussion

4.1. Comparison of the Rational Sampling Numbers in Different Regions

In the spatial-prediction process of soil attribute, the number of sampling points or sampling-point density were important factors related to prediction accuracy and prediction...
cost [13]. In the area where the terrain was relatively flat and the soil salinization was more severe, such as Kenli County, it was believed that the minimum number of sample points was 150. It was equivalent to sampling more than 107 points on the area of 1000 km$^2$ for satisfying the prediction accuracy of SOM. This differs from the results found by other scholars at different regions and scales. For example, in a small karst-basin scale with an area of 75 km$^2$ in Guizhou Province, southern China, 357 sample points could better express the spatial variation of 0–20 cm soil organic carbon, equivalent to 4760 sample points collected on an area of 1000 km$^2$ [26]. In a small undulating hilly area with an area of 184 hm$^2$ in São Paulo state of southwestern Brazil, collecting one sample point per 3.75 hm$^2$ and one sample point per 7.2 hm$^2$ could satisfy the expression of the spatial variability of SOM and clay, which was equivalent to 26,667 samples and 13,889 samples collected on 1000 km$^2$, respectively [27]. Gascuel-Odoux et al. found that sampling 200 points could better describe spatial distribution of soil salinity within a 288 hm$^2$ scale in the Senegal River valley, correspond to sampling 69,444 soil samples per 1000 km$^2$ [28]. These studies all showed that a high number of samples were needed to express spatial variability of soil properties. It may be due to the fact that these study areas were located with undulating terrain and complex topography (e.g., hills, valleys, mountains, etc.), which resulted in high spatial variability of soil properties. Therefore, more sampling points were needed for spatial-distribution prediction.

The results in the Yangtze River Delta region showed that the reasonable sampling numbers of SOM were about 170 sample points per 1000 km$^2$ [29,30]. Although the topography of the Yangtze River Delta was similar to the Yellow River Delta, it had developed agriculture, high farming intensity and high geographical complexity, so the minimum sampling number was slightly higher than that of the Yellow River Delta region. In a typical agricultural planting area of Huang-Huai-Hai Plain in North China (Yucheng County), Sun et al. concluded that 82 samples points per 1000 km$^2$ area could satisfy the expression of the spatial variability of soil organic carbon [31]. Yucheng County belongs to a temperate monsoon climate zone and is located in the alluvial plains of the middle and lower reaches of the Yellow River. The county has similar climate and topographic characteristics as Kenli County, the study area of this paper, so the rational number of sampling points from Yucheng County was similar to that of this paper. At the same time, Yucheng County is located in the hinterland of the Huang-Huai-Hai plain, and the salinity degree was light at present, while Kenli County in this paper was affected by seawater infiltration and hydrogeology, and the soil salinity degree was relatively high. Thus, the environmental conditions of Kenli County were more complex than Yucheng, and furthermore, the minimum sampling number of SOM spatial variation was slightly more than that of Yucheng.

Therefore, the minimum sampling number of 107 sample points per 1000 km$^2$ determined in this paper was only suitable for areas with similar geographical environment to the Kenli County. Under other different natural geographical conditions and different agricultural production models, how many sampling points required need to be adjusted according to local actual conditions when carrying out the spatial prediction of soil properties.

### 4.2. Number of Sampling Points and Spatial-Prediction Error

The error of the spatial prediction was related to the number of sampling points. Some studies had shown that the spatial prediction errors of soil properties decreased as the samples number increases [13,29,32]. In this paper, the prediction accuracy of SOM was evaluated by three indicators: mean error (ME), root-mean-square error (RMSE) and mean standard error (MSE). Although the spatial-prediction error value of the SOM was missing at 900 sampling points, the trend of the predicted error changing from 50 to 800 sampling points still showed that the sample point numbers had some effect on the prediction accuracy of SOM.
The values of ME achieved the optimal prediction accuracy when the number of sample points was greater than or equal to 150, essentially keeping about 0, and there was no obvious trend with the changing of the sample point numbers. The values of RMSE did not have a particularly obvious trend with the increase in the number of samples. It had a certain degree of volatility when the sample numbers were greater than 400; in particular, the values of RMSE at 500 and 600 samples were significantly higher than those of in other sample sets. This showed that the single RMSE index did not well-indicate the prediction accuracy of SOM in this paper. The value of ASE remained around 4 when the sample-point numbers were greater than or equal to 150, and there was less impact on its trend with the increase in sampling-point numbers. At this time, the differences between ASE and RMSE were small, and the accuracies of spatial prediction were improved. It indicated that the accuracies of spatial prediction were stable at an acceptable level when the number of samples increased to a certain value, and the accuracies of spatial prediction were not significantly improved by the further increase in the number of samples. According to the study of Wu et al., this situation may be due to the fact that when the number of sample points was small, the sample points could not be ensured to be distributed in “key areas” that could effectively reflect the spatial-distribution characteristics of soil properties [33]. When the sample points reached a certain number, the integrity and rationality of sample points in spatial distribution would be improved, and the spatial prediction accuracy of soil properties may have improved to a certain extent and maintained at a reasonable value.

4.3. Influencing Factors of Spatial Prediction of SOM

SOM is a very sensitive to time and space. This paper mainly studies the relationship between the number of soil samples and the spatial-prediction accuracy of SOM. However, factors such as topography [6,21], land-use patterns [34,35], vegetation types [36] and human management [2] can all affect the spatial variability of SOC or SOM. Therefore, in the future research, we should further explore the factors that affect the spatial variation of SOM in Kenli County from the perspective of different influencing factors.

5. Conclusions

Spatial distribution of SOM showed a moderate variability in Kenli County, which is located in the Yellow River Delta, and both the spatial-prediction accuracy and spatial-distribution characteristics of SOM were influenced by the number of sampling points. The mean value of the SOM content in the study area was 10.96 g/kg, with a lower-middle fertility level. In terms of spatial variability, soils with SOM content of 9 g/kg~12 g/kg occupied most of the study area; soils with SOM content <9 g/kg were distributed in the southern part of Kenli County with the shape of spots, and soils with SOM content of 12 g/kg~14 g/kg were mainly distributed in the middle and northern parts of the study area. With the decrease in sampling-point numbers, the spatial variability characteristics of SOM showed the phenomenon of gradually weakened detail characterization, information loss and wrong description. From the independent verification results of the predicted values of SOM under different sampling numbers, the RMSE value did not have obvious regularity with the changes in the sampling numbers. Both the ME and ASE showed an obvious inflection point when the number of samples was 150, and remained at about 0 and 4, respectively, as the sampling numbers increased.

Combined with the independent verification index and spatial-distribution characteristics, when predicting the spatial distribution of SOM in Kenli County, the rational sample points to ensure the spatial prediction accuracy of SOM numbered 150. It was equivalent to 107 sample points per 1000 km².

Author Contributions: Conceptualization, X.Z.; methodology, X.Z. and W.W.; software, M.D. and W.W.; validation, M.D.; formal analysis, X.Z., W.W. and M.D.; investigation, X.Z., X.S., X.L. and D.C.; resources, X.Z.; data curation, X.Z., W.W. and M.D.; writing—original draft preparation, W.W.; writing—review and editing, X.Z., M.D., W.W. and X.S.; visualization, M.D.; supervision, X.Z. and
Funding: This research was funded by the National Key Research and Development Program of China [grant number 2021YFD1900900]; the Key Research and Development Program of Shandong Province, China [grant number 2021CXGC010801, 2021CXGC010804]; Shandong Province Modern Agricultural Industry Technology System Cotton Post Innovation Team [grant number SDAIT-03-06]; the Talent Fund of Qingdao Agricultural University, China [grant number 1114344].

Institutional Review Board Statement: Not applicable.

Acknowledgments: The authors acknowledge all the funds that provided funding during the writing process of the paper.

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

References
1. Medina-Méndez, J.; Velke-Haller, V.; Cortés-Flores, J.; Galvis-Spinola, A.; Santiago-Cruz, M.D.J. Soil Organic Matter and Grain Yield of Rainfed Maize in Luvisols of Campeche, Mexico. Agric. Sci. 2019, 10, 1602–1613. [CrossRef]
2. Huang, B.; Sun, W.X.; Zhao, Y.C.; Zhu, J.; Yang, R.Q.; Zou, Z.; Ding, F.; Su, J.P. 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]
3. Baskan, O.; Dengiz, O.; Gunturk, A. Effects of toposequence and land use-land cover on the spatial distribution of soil properties. Environ. Earth Sci. 2016, 75, 448. [CrossRef]
4. Yu, D.S.; Zhang, Z.Q.; Yang, H.; Shi, S.Z.; Tan, M.Z.; Sun, W.S.; Wang, H.J. Effect of Soil Sampling Density on Detected Spatial Variability of Soil Organic Carbon in a Red Soil Region of China. Pedosphere 2011, 21, 207–213. [CrossRef]
5. Wang, J.M.; Yang, R.X.; Bai, Z.K. Spatial variability and sampling optimization of soil organic carbon and total nitrogen for Minesolos of the Loess Plateau using geostatistics. Ecol. Eng. 2015, 82, 159–164. [CrossRef]
6. Long, J.; Liu, Y.L.; Xing, S.H.; Qiu, L.X.; Huang, Q.; Zhou, B.Q.; Shen, J.Q.; Zhang, L.M. Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography. Ecol. Indic. 2018, 93, 562–571. [CrossRef]
7. Lütticken, R.E. Automation and standardisation of site specific soil sampling. Precis. Agric. 2000, 2, 179–188. [CrossRef]
8. 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]
9. Li, Y. Can the spatial prediction of soil organic matter contents at various sampling scales be improved by using regression kriging with auxiliary information? Geoderma 2010, 159, 63–75. [CrossRef]
10. Lai, Y.Q.; Wang, H.L.; Sun, X.L. A comparison of importance of modelling method and sample size for mapping soil organic matter in guangdong, china. Ecol. Indic. 2021, 126, 107618. [CrossRef]
11. Marques, J.J.; Alleoni, L.R.F.; Teixeira, D.D.B.; Siqueira, D.S.; Pereira, G.T. Sampling planning of micronutrients and aluminium of the soils of São Paulo, Brazil. Geoderma Regul. 2015, 4, 91–99. [CrossRef]
12. Pang, S.; Li, T.X.; Wang, Y.D.; Yu, Y.H.; Li, X. Spatial interpolation and sample size optimization for soil copper(cu) investigation in cropland at county scale using cokriging. Agric. Sci. China 2009, 8, 1369–1377. [CrossRef]
13. Zhang, Z.Q.; Yu, D.S.; Shi, X.Z.; Wang, N.; Zhang, G.X. Priority selection rating of sampling density and interpolation method for detecting the spatial variability of soil organic carbon in China. Environ. Earth Sci. 2015, 73, 2287–2297. [CrossRef]
14. Lv, Z.Z.; Liu, G.M.; Yang, J.S.; Zhang, M.M.; He, L.D.; Shao, H.B.; Yu, S.P. Spatial Variability of Soil Salinity in Bohai Sea Coastal Wetlands, China: Partition into Four Management Zones. Plant Biosyst. Int. J. Deal. Asp. Plant Biosyst. 2013, 147, 1201–1210. [CrossRef]
15. Zhang, X.G.; Wang, Z.G.; Song, X.Y.; Liu, P.R.; Li, S.M.; Yang, X. Effect of sampling on spatial variability in soil salinity in the Yellow River Delta Area. Resour. Sci. 2016, 38, 2375–2382. (In Chinese) [CrossRef]
16. Zhao, G.X.; Yan, M. Remote sensing image based information extraction for land salinized degradation and its evolution—A case study in Kenli County of the Yellow River Delta. In Proceedings of the Sixth International Conference on Natural Computation, ICNC 2010, Yantai, China, 10–12 August 2010.
17. Jiang, M.; Xu, L.; Chen, X.; Zhu, H.; Fan, H. Soil Quality Assessment Based on a Minimum Data Set: A Case Study of a County in the Typical River Delta Wetlands. Sustainability 2012, 12, 9033. [CrossRef]
18. Liu, J.; Zhang, L.; Dong, T.; Wang, J.; Fan, Y.; Wu, H.; Geng, Q.; Yang, Q.; Zhang, Z. The Applicability of Remote Sensing Models of Soil Salinization Based on Feature Space. Sustainability 2021, 13, 13711. [CrossRef]
19. Chang, C.; Lin, F.; Zhou, X.; Zhao, G. Hyper-spectral response and estimation model of soil degradation in Kenli County, the Yellow River Delta. PLoS ONE 2020, 15, e0227594. [CrossRef]
20. Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon and organic matter. In Methods of Soil Analysis, Part 2, 2nd ed.; Page, A.L., Miller, R.H., Keeney, D.R., Eds.; ASA and SSSA: Madison, WI, USA, 1982; Agronomy Monograph; Volume 9, pp. 534–580.
21. Hu, K.L.; Wang, S.Y.; Li, H.; Huang, E.; Li, B.G. Spatial scaling effects on variability of soil organic matter and total nitrogen in suburban Beijing. *Geoderma* **2014**, *226–227*, 54–63. [CrossRef]
22. Goovaerts, P. Geostatistical modelling of uncertainty in soil science. *Geoderma* **2001**, *103*, 3–26. [CrossRef]
23. Cerri, C.E.P.; Bernoux, M.; Chaplot, V.; Volkoff, B.; Victoria, R.L.; Melillo, J.M.; Paustian, K.; Cerri, C.C. Assessment of soil property spatial variation in an Amazon pasture: Basis for selecting an agronomic experimental area. *Geoderma* **2004**, *123*, 51–68. [CrossRef]
24. Nielsen, D.R.; Bouma, J. Soil spatial variability. In Proceedings of the Workshop of the International Society of Soil Science and Soil Science Society of America, Las Vegas, NV, USA, 30 November–1 December 1984; Pudoc: Wageningen, The Netherlands, 1985; p. 243.
25. Goovaerts, P. Geostatistics in soil science: State-of-the-art and perspectives. *Geoderma* **1999**, *89*, 1–45. [CrossRef]
26. Zhang, Z.M.; Zhou, Y.C.; Huang, X.F. Exploring the optimal sampling density to characterize spatial heterogeneity of soil carbon stocks in a Karst Region. *Agron. J.* **2020**, *113*, 99–110. [CrossRef]
27. Nanni, M.R.; Povh, F.P.; Dematte, J.A.M.; Oliveira, R.B.D.; Chicati, M.L.; Cezar, E. Optimum size in grid soil sampling for variable rate application in site-specific management. *Sci. Agric.* **2011**, *68*, 386. [CrossRef]
28. Gascuel-Odoux, C.; Boivin, P. Variability of variograms and spatial estimates due to soil sampling: A case study. *Geoderma* **1994**, *62*, 165–182. [CrossRef]
29. Hai, N.; Zhao, Y.C.; Tian, K.; Huang, B.; Sun, W.X.; Shi, X.Z. Effect of number of sampling sites on characterization of spatial variability of soil organic matter. *Acta Pedol. Sin.* **2015**, *52*, 783–791. (In Chinese) [CrossRef]
30. Wang, Z.G.; Zhao, Y.C.; Huang, B.; Darilek, J.L.; Sun, W.X. Effects of sample size on spatial characterization of soil fertility properties in an agricultural area of the Yangtze River delta region, China. *Soils* **2010**, *42*, 421–428. (In Chinese) [CrossRef]
31. Sun, W.X.; Zhao, Y.C.; Huang, B.; Shi, X.Z.; Darilek, J.L.; Yang, J.S.; Wang, Z.G.; Zhang, B.E. Effect of sampling density on regional soil organic carbon estimation for cultivated soils. *J. Plant Nutr. Soil Sci.* **2012**, *175*, 671–680. [CrossRef]
32. Heim, A.; Wehrli, L.; Eugster, W.; Schmidt, M. Effects of sampling design on the probability to detect soil carbon stock changes at the swiss carboeurope site laegeren. *Geoderma* **2009**, *149*, 347–354. [CrossRef]
33. Wu, Z.F.; Zhao, Y.F.; Cheng, D.Q.; Chen, J. Influences of sample size and spatial distribution on accuracy of predictive soil mapping on a county scale. *Acta Pedol. Sin.* **2019**, *56*, 1321–1335. (In Chinese) [CrossRef]
34. Chuai, X.; Huang, X.; Wang, W.; Zhang, M.; Lai, L.; Liao, Q. Spatial Variability of Soil Organic Carbon and Related Factors in Jiangsu Province, China. *Pedosphere* **2012**, *22*, 404–414. [CrossRef]
35. Zhang, Z.Q.; Yu, D.S.; Shi, X.Z.; Weindorf, D.C.; Wang, X.X.; Tan, M.Z. Effect of sampling classification patterns on SOC variability in the red soil region, china. *Soil Tillage Res.* **2010**, *110*, 1–27. [CrossRef]
36. Wang, Y.Q.; Zhang, X.C.; Zhang, J.L.; Li, S.J. Spatial variability of soil organic carbon in a watershed on the Loess Plateau. *Pedosphere* **2009**, *19*, 486–495. [CrossRef]