Rational sampling numbers of soil pH for spatial variation: A case study from Yellow River Delta

Yingxin Zhang

Qingdao Agricultural University

Mengqi Duan

Qingdao Agricultural University

Xiaoguang Zhang (✉ zhangxg_66@sina.com)

Qingdao Agricultural University

Research Article

Keywords: Sampling number, Soil pH, Spatial variation, Yellow River Delta, Ordinary Kriging

Posted Date: March 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1402057/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Spatial variation of soil pH is important for the evaluation of environmental quality. A reasonable number of sampling points has an important meaning for accurate quantitative expression on spatial distribution of soil pH and resource savings. Based on the grid distribution point method, 797, 700, 594, 499, 398, 299, 200, 149, 100, 75 and 50 sampling points, which were randomly selected from 908 sampling points, constituted 12 sample sets. Semi-variance structure analysis was carried out for different point sets, and ordinary Kriging was used for spatial prediction and accuracy verification, and the influence of different sampling points on spatial variation of soil pH was discussed. The results showed that the pH value in Kenli County was generally between 7.8 and 8.1, and the soil was alkaline. Semi-variance models fitted by different point sets could well reflect the spatial structure characteristics of soil pH. With the decrease of the number of sampling points, the Sill value of sample set increased, and the spatial autocorrelation gradually weakened. Considering the prediction accuracy, spatial distribution and investigation cost, the number of sampling points greater than or equal to 150 could satisfy the spatial variation expression of soil pH at the county level in the Yellow River Delta. This was equivalent to taking at least 107 sampling points per 1 000 km². The results in this study are applicable to areas with similar environmental and soil conditions as the Yellow River Delta, and have certain reference significance for them.

Introduction

Soil pH is an important index for evaluation of land quality 1. Soils with acid and alkali over national standard are not conducive to the utilization of land resources 2. Therefore, it is very important to understand the spatial distribution of soil pH. At the same time, it is of great guiding significance to accurately grasp the spatial variation characteristics of soil pH for evaluating the salinization and acidification of soil environment, rational fertilization and efficient utilization of nutrients 3. However, sampling number affects the accuracy of soil properties and their spatial variation information and the degree of quantitative expression. The layout pattern and sampling number must be fully considered to ensure the accuracy of spatial interpolation in the study on spatial variation of soil pH 4~6. Generally speaking, the larger the sampling density, the smaller the sample error and the higher the accuracy of the research results. But the work cycle will be prolonged, and huge manpower, material resources and financial resources will be consumed 7. The sampling cost will also limit the sampling density to a large extent. If the number of sampling points is reduced, the interpolation accuracy of soil pH spatial variation will be difficult to guarantee, and local features may not be displayed 8~9. Therefore, it is of great significance to study the reasonable number of sampling points for accurate quantitative expression on spatial distribution of soil pH and resource savings.

Spatial geostatistics is one of the most accurate spatial prediction methods, which is often used to model the spatial variability for soil properties and evaluate their spatial uncertainty ¹⁰. At present, the method combining geostatistics and GIS is used to study the spatial distribution of regional soil properties from the perspective of spatial prediction. And it is also used to analyze the influence of sampling density on

the spatial variation of soil properties¹¹. In recent years, many scholars have carried out a large number of geostatistics researches on the influences of sampling number on spatial variability for different soil properties in different areas, such as soil organic matter^{12–14}, nitrogen¹⁵, exchangeable potassium, calcium and magnesium¹⁶ and heavy metals^{17,18}, salt, etc.¹⁹. The reasonable sampling quantity of different soil properties is basically different for their different characteristics. Even for the same specific soil index, the results of different researches are different. For example, for soil organic matter, the reasonable sampling points of spatial variation in typical areas of Yangtze River Delta were 91 per 1 000 km²²⁰, 547 per 1 000 km² in Fei County which is a typical county of North China Plain²¹. And for soil organic carbon, 908 sampling points were reasonable in typical gully areas of Loess Plateau, which means that 178 sampling points were needed per 1 000 km². These studies show that there were differences in the number of reasonable sampling points even with the same soil index because of the differences in natural geographical environment such as topography and geomorphology in different areas.

In addition, the survey scale has a certain influence on reasonable sampling number. For example, in the relatively consistent geomorphic units, such as Fujian Province and its typical counties, the reasonable sampling points of spatial distribution for soil organic matter were 10,000 and 11 000 for every 1 000 km², respectively¹². The existing research methods and conclusions still need to be tested because of the different evaluation indexes, the different natural geographical conditions and the different influence of human activities in different study areas.

To sum up, although many scholars have carried out relevant researches on the reasonable sampling number, there are few reports on the related research for soil pH at present. Especially, the research at county-level areas is lacking in the influence of different sampling points on the spatial variation of soil pH. And also, because of the fragile and salinized soil environment in the Yellow River Delta region and great attention from the government, it is of great significance to monitor and master the spatial distribution status of soil pH for green development in agriculture on the premise of clarifying the reasonable sampling numbers. Therefore, Kenli County in the Yellow River Delta was selected as the study area, and the spatial distribution of soil pH with 12 different sampling sets was predicted using geostatistical Kriging interpolation. The overall objectives of this research were 1) to clarify the influence of different sampling numbers on the prediction accuracy of spatial distribution for soil pH; 2) to determine the reasonable sampling densities to realize the spatial variation of soil pH at county scale.

Results

Descriptive statistics characteristics of soil pH for different sets of sample points

Descriptive statistics analysis was made on 12 sample sets, and the results were shown in Table 1. The soil pH values of 908 sampling points in the study area were 7.00, 8.80, and 7.85, respectively. The soil was weak alkaline. The coefficient of variation was 5.35%, and the variability was weak, indicating that the alkalization degree was very concentrated. The skewness coefficient was -0.31, and the kurtosis

coefficient was 2.12. The results of normal test on skewness coefficient and kurtosis coefficient showed that the soil pH values of different sampling points were in accordance with normal distribution.

The minimum value of soil pH for 100 and 50 sampling points was 7.10g/kg, and the minimum value of other sampling points was 7.00g/kg. The maximum values of 908, 797, 700, 594, 499, 398, 299, 200 and 100 sample points were all 8.80g/kg, and the maximum values of only 149, 75 and 50 sample points were 8.60g/kg. But they were still very similar. Among the 11 sub-samples, the average value and standard deviation of soil pH also fluctuated around the average value and standard deviation of the complete set, which indicated that although the number of sampling points decreased, the 11 samples can still represent the complete set. The coefficient of variation of soil pH ranged from 5.06% to 5.36% among the 11 sub-samples, and the variability was weak. To sum up, the analysis of each index of each subset showed that the selected subsets were all representative.

Table 1 Descriptive statistics characteristics of soil pH for different sets of sample points

Sampling point number	Min (g/kg)	Max (g/kg)	Average (g/kg)	Standard deviation (g/kg)	Skewness	Kurtosis	Median (g/kg)	Variation (%)
908	7.00	8.80	7.85	0.42	-0.31	2.12	7.90	5.35
797	7.00	8.80	7.84	0.42	-0.32	2.15	7.90	5.36
700	7.00	8.80	7.85	0.41	-0.33	2.18	7.90	5.22
594	7.00	8.80	7.85	0.42	-0.35	2.18	7.90	5.35
499	7.00	8.80	7.86	0.41	-0.37	2.19	7.90	5.22
398	7.00	8.80	7.88	0.41	-0.37	2.18	7.90	5.20
299	7.00	8.80	7.88	0.40	-0.38	2.30	7.90	5.08
200	7.00	8.80	7.89	0.42	-0.43	2.33	7.90	5.32
149	7.00	8.60	7.90	0.42	-0.59	2.46	8.00	5.32
100	7.10	8.80	7.90	0.40	-0.35	2.40	7.95	5.06
75	7.00	8.60	7.88	0.40	-0.55	2.43	8.00	5.08
50	7.10	8.60	7.85	0.41	-0.15	2.07	7.90	5.22

The influence of different sampling points on the semi-variance structure of soil pH

Semi-variance analysis of soil pH was carried out by geostatistical method. Table 2 shows the semi-variance function values of soil pH under different sampling points. The spatial variation structure of soil pH at other sampling densities conformed to the exponential model, except for 75 sampling points. The level of decision coefficient represented the effect of fitting variogram by the model. The higher the

decision coefficient, the better the effect of fitting variogram by the model ²². The determination coefficients of different sampling points were between 0.39 and 0.69, indicating that the model can better reflect the spatial structure characteristics of soil pH.

The ratio of Nugget to base Sill (C_0+C) can reflect the degree of spatial autocorrelation of variables. It was considered a strong spatial autocorrelation when the ratio was less than 25%—it was considered moderate spatial autocorrelation when the ratio was between 25% and 75%—it was considered weak spatial autocorrelation when the ratio was greater than 75%. Nugget/Sill of the total sample set (908 sample points) in this study was less than 25%, showing a strong spatial autocorrelation. The Nugget/Sill of 200-800 samples in the other 11 sample subsets was less than 25%, which indicated that these sample sets had strong spatial autocorrelation. However, when the number of samples was less than 200, the Nugget/Sill ranged from 25% to 75% (with moderate spatial autocorrelation). Especially when the number of samples was less than 150, the Nugget/Sill reached 40%, which indicated that the spatial autocorrelation of these samples was weakened.

The variable range represented the autocorrelation range of the variogram ²³, it can reflect the size of the autocorrelation range in the variable space. In this paper, the fitting range of soil pH under different number of points was more than 4 km, indicating that the spatial autocorrelation distance was relatively large. Among them, when the number of sampling points was reduced to 100, the range increased obviously, reaching 26.52, which was about 5 times that of 200 sampling points.

Table 2 Parameters of semi-variance function of soil pH under different sampling points

Sampling point number	Model	Nugget (C0)	Sill (C0+C)	C0/Sill (%)	Range (km)	Determination	Residual
926	exponential model	0.0350	0.1786	19.60	4.80	0.68	0.00
801	exponential model	0.0360	0.1778	20.25	5.13	0.72	0.00
704	exponential model	0.0360	0.1752	20.55	4.77	0.68	0.00
598	exponential model	0.0360	0.1780	20.22	4.75	0.69	0.00
502	exponential model	0.0330	0.1742	18.94	4.82	0.69	0.00
401	exponential model	0.0350	0.1750	20.00	4.70	0.60	0.00
300	exponential model	0.0350	0.1658	21.11	4.81	0.60	0.00
201	exponential model	0.0295	0.1810	16.30	4.87	0.44	0.00
150	exponential model	0.0442	0.1736	25.44	6.54	0.39	0.00
100	exponential model	0.0952	0.1914	49.74	26.52	0.64	0.00
75	spherical model	0.0869	0.1748	49.71	13.77	0.49	0.00
50	exponential model	0.1247	0.3174	39.29	160.59	0.61	0.00

The influence of different sampling points on the spatial prediction accuracy of soil pH

For each sample set, Kriging interpolation was used to carry out the spatial prediction of soil pH. Root mean square error (RMSE), mean error (ME) and average standard error (ASE) were used to measure the prediction accuracy of soil pH under different sampling points.

The closer the ME was to 0, the higher the accuracy of spatial prediction is. It can be seen from Fig. 1 that the ME of 150, 200 and 300 sampling points was greater than 0, and that of other sampling points was less than 0. With the decreased of sample size, the distance of ME deviating from X axis first increased, then decreased and then slightly increased. But it basically surrounded the X axis except for 75 and 100 sample points. Therefore, when the number of sample points was too small (75, 100), the ME of predicted values would become larger.

The RMSE was slightly larger with 50 samples. There was no obvious change trend with the decreased of sample size, and it could basically remain at a certain value. It showed that the RMSE had no obvious difference in the spatial prediction accuracy of different sample subsets, that is, the accuracy of Kriging interpolation had no significant difference with the decreased of sample size.

The uncertainty of Kriging prediction was measured by the ASE. The closer the ASE was to the RMSE, the more accurate the prediction of attribute value was. In this study, when there were 150, 200 and 900 samples, ASE and RMSE were basically equal, indicating that the spatial variability of the earth was properly estimated. When the number of sampling points was 300-800, the ASE was smaller than the RMSE, and the difference between them was very small, indicating that these sampling points overestimated the spatial variability. However, when the number of sampling points was less than 150, the ASE was obviously less than the RMSE, so these sample sets cannot reasonably predict the variability.

Considering the ME, RMSE and ASE, 107 sampling points per 1 000 km² could meet the needs of spatial variation expression of soil pH in the Yellow River Delta. According to the above analysis, we can also know that it was not enough to evaluate Kriging prediction accuracy only by a single evaluation index. Therefore, it was necessary to use a variety of evaluation indexes and combine them with practice to accurately evaluate the prediction results.

Effects of different sampling points on spatial distribution of soil pH

To more intuitively show the influence of different sampling point sets on the spatial distribution of soil pH, Kriging method was used to carry out spatial interpolation operation (Fig 2). The eastern of Kenli County was not sampled because the area is the Yellow River Delta National Nature Reserve. However, for the beauty and completeness of pictures, Kriging interpolation was extended to the whole Kenli County. When analyzing the spatial variation characteristics of soil pH, the Yellow River Delta National Nature Reserve in the east of Kenli County did not focus on analysis.

In the study area, the yellow part had a large area (Fig. 2), that is, the soil pH in Kenli County of the Yellow River Delta was generally between 7.8 and 8.1, which proved that the soil in this area was alkaline. The areas with high soil pH value (orange-red part) were distributed in the middle and southwest of study area. With the decreased of the number of sample points, the ability of describing details was gradually weakened. When the number of sampling points was reduced to 100, the details of soil pH in the middle and north of the study area were no longer detailed, and only the general distribution trend could be seen. Therefore, considering the precision and research funds, it is suggested that the reasonable sampling number in the Yellow River Delta should not be less than 150.

Conclusion

In this paper, Kenli County in the Yellow River Delta was selected as the research area. 12 sample sets consisting of 908, 797, 700, 594, 499, 398, 299, 200, 149, 100, 75 and 50 sampling points were selected to

study the influence of the sampling number on the spatial variation of soil pH. And the reasonable sampling number was determined, which provided for the collection of soil samples. Soil samples getting with minimum human, material and financial resources provided convenience for the research of the soil pH.

According to the semi-variance structure analysis of soil pH with different sampling sets, the determination coefficient of different sampling sets was between 0.39 and 0.69, and the model could better reflect the spatial structure characteristics of soil pH. With the decreased of the number of sampling points, the C0/Sill value of the sample set increased, and the spatial autocorrelation decreased gradually. In this paper, the variation range of soil pH fitting by different number of points was greater than 4 km, and the spatial autocorrelation distance was relatively large. Among them, when the number of sampling points was reduced to 100, the range increased obviously, reaching 26.52, which was about 5 times that of 200 sampling points.

According to the analysis of spatial prediction accuracy of soil pH with different sampling sets, the ME of predicted values would gradually increase when the number of sampling points was too small. The RMSE could basically be maintained at a certain value with the decrease of sampling numbers except for being slightly larger at 50 sample points. Comprehensive analysis of ME, RMSE and ASE showed that when the number of sampling points was 150, the prediction accuracy was the highest, which can satisfy the spatial variation expression of soil pH in the Yellow River Delta region.

According to the analysis of spatial distribution and expression of soil pH with different sampling sets, the pH value in Kenli County was generally between 7.8 and 8.1, and the soil was alkaline. The areas with high soil pH value were distributed in the middle and southwest of the study area. With the decreased of sampling points, the detailed characteristics of spatial variation of soil pH will gradually disappear. When the number of sampling points was 150, it can not only describe the spatial distribution of soil pH in detail, but also accurately describe the spatial distribution pattern of soil pH.

Therefore, considering the prediction accuracy, spatial distribution and research funds, it is suggested that the reasonable sampling number should be no less than 150 in the Yellow River Delta region, which is equivalent to at least 107 sampling points for 1 000 km². The results of this study are applicable to areas with similar environmental and soil conditions in the Yellow River Delta, and have certain reference significance for them. However, the number of sampling points will be different to reasonably express the spatial distribution of soil pH in other areas, which needs to be analyzed in combination with local environmental conditions.

Discussion

Spatial variation of soil properties is controlled by various structural and random factors. The larger the ratio of Nugget to Sill, the more obvious the influence of human activities, such as irrigation, fertilization and cultivation. On the contrary, structural factors such as soil parent material, climate, biology,

topography and other natural factors play a major role^{24,25}. The ratio of Nugget to Sill of the whole sample set (908 sampling points) in this study was less than 25% and showed strong spatial autocorrelation, indicating that the spatial variation of soil pH in the study area was mainly affected by structural components such as topography, climate and soil parent material. When the sampling point was less than 150, the ratio of Nugget to Sill reaches 40%, indicating that the spatial autocorrelation of these sample sets was weakened, and was influenced by structural components and random factors. With the decreased of the number of sampling points, the small-scale structural factors and random factors gradually increased, while the influence of large-scale structural factors such as parent material, topography and soil type on soil pH gradually weakened, which caused the soil pH to change strongly with the small number of sampling points.

Interpolation prediction error and interpolation distribution map have a certain synergistic relationship with the change of sample numbers. When there were less than 150 samples, the difference between ASE and RMSE was large. Correspondingly, the spatial distribution maps of soil pH with less than 150 sampling points were too smooth to show the spatial distribution of soil pH accurately. However, when the number of samples was increased to 150 or more, the details of the spatial variation expression diagram of soil pH were more obvious, which can show the spatial variation expression of soil pH more accurately. In addition, at 150 or more sampling points, the prediction error maps of RMSE, ME and ASE of soil pH and the expression maps of spatial variation of soil pH in different point sets can better predict the spatial variation of soil pH. So it can be shown that the prediction error of soil pH was consistent with the spatial distribution of soil pH.

In this paper, the influence of the number of different sampling points on the spatial distribution of soil pH in the Yellow River Delta region was discussed by using the ordinary Kriging method. It was concluded that the number of sampling points most suitable for Kenli County in the Yellow River Delta region should be no less than 150, that is, at least 107 sampling points should be taken every 1 000 km². This paper was al-so compared with the existing literature investigating reasonable sampling numbers of soil pH. The only literature²⁶ showed that the rational sampling number at the county scale was about 4900 samples per 1 000 km² which was inconsistent with our study. It may be due to that study areas of the former were located in a more undulating mountainous and hilly region and resulting a reduced spatial auto-correlation of soil pH. While this result was the same as the most reasonable number of sampling points needed for soil salinity in Kenli County of the Yellow River Delta studied by Zhang et al.¹⁹. It revealed that the number of sampling points for spatial variation expression of different soil properties in areas with similar environmental conditions may get closer.

Methods

Study area

Kenli County in the Yellow River Delta region was selected as the study area (Fig. 3). This county is located at the mouth of the Yellow River in the Yellow River Delta of northern Shandong Province,

between 37°24' - 38°10' N and 118°15' - 119°19' E ²⁷, with a total area of 2 331 km². It belongs to temperate monsoon climate, with high annual temperature and uneven spatial-temporal distribution of precipitation. The terrain is fan-shaped and slightly inclined from southwest to northeast ²⁸. The altitude ranges from 2 m to 11.61 m. Kenli County has rich in soil resources, and is one of the most abundant land reserve resources in the coastal areas of eastern China, with great potential for agricultural development.

Collection and processing of sample data

Based on the grid distribution point method, more than 1 100 sampling points were distributed in this study. The outliers were removed after collecting the samples. Finally, 908 samples were obtained. 797, 700, 594, 499, 398, 299, 200, 149, 100, 75 and 50 sampling points were randomly selected from 908 sampling points. Each sampling point was extracted from the last sampling point set to compare the characteristics of different sample sets. For example, 700 sample set was extracted from 797 sample set, and 299 sample set was extracted from 398 sample set. Combining all the sample data, a total of 12 sample sets were formed. The pH of each sample was measured in 1:2.5 mixtures of soil and deionized water with a pH meter ²⁹.

Spatial prediction and verification method

Geostatistical methods are widely used to predict the spatial distribution of soil properties ³⁰⁻³². In this paper, we chose the Ordinary Kriging (OK) method to predict the spatial distribution of soil pH ³³. The OK satisfies the intrinsic hypothesis, and the average value of the regionalized variables is an unknown constant ³⁴. OK is a linear estimation of regionalized variables, which is similar to weighted moving average in the process of interpolation research. However, the weights of weighted moving average are determined from different sources. The weighted sliding average weight values come from known spatial functions, while the weights of ordinary kriging come from spatial data analysis ³⁴. It is necessary to verify the prediction results after spatial distribution prediction. In this paper, independent verification method was adopted.

The OK interpolation results are verified. The verification method of independent data set extracts some samples from all samples as independent data sets, take the remaining samples as simulation data sets without repetition, and take each sample in independent data sets as inspection points ³⁵. In this paper, after 797 samples were extracted from 908 samples, the remaining 111 samples were taken as independent verification sets. The spatial prediction results of 12 sample sets consisting of 908, 797, 700, 594, 499, 398, 299, 200, 149, 100, 75 and 50 samples were verified, respectively. In verification method of independent data set, the most representative evaluation indexes are root mean square error (RMSE), mean error (ME) and average standard error (ASE). Two verification indexes, root mean square error and average error, were chosen to evaluate the accuracy of prediction. The smaller the RMSE is, the closer the ME is to zero, indicating that the accuracy of spatial prediction is higher. The average standard error was used to measure the uncertainty of Kriging prediction value.

Declarations

Acknowledgements

This work was supported by the National Key Research and Development Program of China [grant number 2021YFD190090206, 2021YFD190090308], the Key Research and Development Program of Shandong Province, China [grant number 2021CXGC010801, 2021CXGC010804], the Open Fund of the State Key Laboratory of Soil and Sustainable Agriculture, China [grant number Y20160007], and the Postgraduate Innovation Program of Qingdao Agricultural University [grant number QNYCX20080].

Author Contributions statement

Y.Z. and M.D. wrote the main manuscript text. M.D. contributed to software, data curation and methodology. X.Z. contributed to investigation, conceptualization, project administration, writing - review & editing and supervision. All authors reviewed the manuscript.

Data Availability Statement

All data generated or analysed during this study are included in this published article.

Competing interests

The author(s) declare no competing interests.

References

1. Xu, H. & Zhang, C. Investigating spatially varying relationships between total organic carbon contents and ph values in european agricultural soil using geographically weighted regression. *Science of The Total Environment*. **752**, 141977. <https://doi.org/10.1016/j.scitotenv.2020.141977>. (2020)
2. Kang, E. et al. Soil pH and nutrients shape the vertical distribution of microbial communities in an alpine wetland. *Science of The Total Environment*. **774**, 145780. <https://doi.org/10.1016/j.scitotenv.2021.145780>. (2021)
3. Vanderlinden, K., Polo, M. J., Ordóñez, R., & Giráldez, J. Spatiotemporal evolution of soil ph and zinc after the aznalcóllar mine spill. *Journal of Environmental Quality*. **35**, 37-49. <https://doi.org/10.2134/jeq2004.0390>. (2006)
4. Zhang, Z. et al. Effect of sampling classification patterns on soc variability in the red soil region, china. *Soil and Tillage Research*. **110**, 2-7. <https://doi.org/10.1016/j.still.2010.05.007>. (2010)
5. Sun, W. et al. Effect of sampling density on regional soil organic carbon estimation for cultivated soils. *Journal of Plant Nutrition and Soil Science*. **175**, 671-680. <https://doi.org/10.1002/jpln.201100181>. (2012)
6. Duan, M. & Zhang, X. Using remote sensing to identify soil types based on multiscale image texture features. *Computers and Electronics in Agriculture*. **187**, 106272.

- <https://doi.org/10.1016/j.compag.2021.106272>. (2021)
7. Molla, A. et al. Optimal spatial sampling design for monitoring potentially toxic elements pollution on urban green space soil: a spatial simulated annealing and k-means integrated approach. *Science of The Total Environment*. **802**. <https://doi.org/10.1016/j.scitotenv.2021.149728>. (2021)
 8. Bhunia, G., Shit, P. & Maiti, R. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences*. **17**, 114-126. <https://doi.org/10.1016/j.jssas.2016.02.001>. (2016)
 9. Bogunovic, I., Mesic, M., Zgorelec, Z., Jurisic, A. & Bilandzija, D. Spatial variation of soil nutrients on sandy-loam soil. *Soil and Tillage Research*. **144**, 174-183. <https://doi.org/10.1016/j.still.2014.07.020>. (2014)
 10. Fisonga, M., Wang, F. & Mutambo, V. The estimation of sampling density in improving geostatistical prediction for geotechnical characterization. *International Journal of Geotechnical Engineering*. 1-8. <https://doi.org/10.1080/19386362.2018.1526484>. (2018)
 11. Wang, J., Yang, R. & Feng, Y. Spatial variability of reconstructed soil properties and the optimization of sampling number for reclaimed land monitoring in an opencast coal mine. *Arabian Journal of Geosciences*. **10**, 46. <https://doi.org/10.1007/s12517-017-2836-0>. (2017)
 12. Long, J. et al. Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography. *Ecological Indicators*. **93**, 562-571. <https://doi.org/10.1016/j.ecolind.2018.05.044>. (2018)
 13. Zhao, Z., Yang, Q., Sun, D., Ding, X. & Meng, F. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. *Computers and Electronics in Agriculture*. **169**, 105172. <https://doi.org/10.1016/j.compag.2019.105172>. (2020)
 14. Zhang, Z., Sun, Y., Yu, D., Mao, P. & Xu, L. Influence of Sampling Point Discretization on the Regional Variability of Soil Organic Carbon in the Red Soil Region, China. *Sustainability*. **10**, 3603. <https://doi.org/10.3390/su10103603>. (2018)
 15. Sckh, A. & Djib, B. How many sampling points are needed to estimate the mean nitrate-n content of agricultural fields? A geostatistical simulation approach with uncertain variograms - sciencedirect. *Geoderma*. **385**, 114816. <https://doi.org/10.1016/j.geoderma.2020.114816>. (2021)
 16. Behera, S. & Shukla, A. Spatial distribution of surface soil acidity, electrical conductivity, soil organic carbon content and exchangeable potassium, calcium and magnesium in some cropped acid soils of india. *Land Degradation and Development*. **26**, 71-79. <https://doi.org/10.1002/ldr.2306>. (2015)
 17. Song, X., Zhang, G., Liu, F., Li, D. & Zhao, Y. Characterization of the spatial variability of soil available zinc at various sampling densities using grouped soil type information. *Environmental Monitoring and Assessment*. **188**, 600. <https://doi.org/10.1007/s10661-016-5615-6>. (2016)
 18. Liu, D., Wang, X., Nie, L., Liu, H. & Wang, W. Comparison of geochemical patterns from different sampling density geochemical mapping in altay, xinjiang province, china. *Journal of Geochemical Exploration*. **6**, 106761. <https://doi.org/10.1016/j.gexplo.2021.106761>. (2021)

19. Zhang, X. et al. Effect of sampling on spatial variability in soil salinity in the Yellow River Delta Area. *Resources Science*. **12**, 2375-2382. <https://doi.org/10.18402/resci.2016.12.16>. (2016)
20. Hai, N. et al. Effect of number sampling sites on characterization of spatial variability of soil organic matter. *Acta Pedologica Sinica*. **52**, 783-791. <https://doi.org/10.11766/trxb201410080505>. (2015)
21. Zhao, Q., Zhao, Y., Jiang, H., Li, M. & Tang, J. Study on Spatial Variability of Soil Nutrients and Reasonable Sampling Number at County Scale. *Journal of Natural Resources*. **27**, 1382-1391. <https://doi.org/CNKI:SUN:ZRZX.0.2012-08-013>. (2012)
22. Yan, P. et al. Spatial variability in soil ph and land use as the main influential factor in the red beds of the nanxiong basin, china. *PeerJ*. **7**. <https://doi.org/10.7717/peerj.6342>. (2019)
23. Kerry, R. & Oliver, M.A. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*. **140**, 383-396. <https://doi.org/10.1016/j.geoderma.2007.04.019>. (2007)
24. Cambardella, C.A., Moorman, T.B., Novak, J.M., Parkin, T.B. & Konopka, A.E. Field-scale variability of soil properties in central iowa soils. *Soil Science Society of America Journal*. **58**, 1501-1511. <https://doi.org/10.2136/sssaj1994.03615995005800050033x>. (1994)
25. Sokal, R.R. & Thomson, J.D. Applications of Spatial Autocorrelation in Ecology. *Springer Berlin Heidelberg*. 431-466. https://doi.org/10.1007/978-3-642-70880-0_12. (1987)
26. Wang, J.G., Zhou, W.J., Wang, B.W. & Chen, C. Soil Sample Density and Interpolation Accuracy on County Scale - A Case Study on Soil pH. *Hunan Agricultural Sciences*. **21**, 27-30. <https://doi.org/10.16498/j.cnki.hnnykx.2011.21.017>. (2011)
27. Li, Y. et al. Distribution of heavy metals in soils of the yellow river delta: concentrations in different soil horizons and source identification. *Journal of Soil and Sediments*. **14**, 1158-1168. <https://doi.org/10.1007/s11368-014-0861-0>. (2014)
28. 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*. **12**. <https://doi.org/10.3390/su12219033>. (2020)
29. Hong, S., Gan, P. & Chen, A. Environmental controls on soil pH in planted forest and its response to nitrogen deposition. *Environmental Research*. **172**, 159-165. <https://doi.org/10.1016/j.envres.2019.02.020>. (2019)
30. Rossel, R. & Mcbratney, A.B. Soil chemical analytical accuracy and costs: implications from precision agriculture. *Australian Journal of Experimental Agriculture*. **38**, 765-775. <https://doi.org/10.1071/ea97158>. (1998)
31. Emadi, M., Shahriari, A.R., Sadegh-Zadeh, F., Seh-Bardan, B.J. & Dindarlou, A. Geostatistics-based spatial distribution of soil moisture and temperature regime classes in mazandaran province, northern iran. *Archives of Agronomy and Soil Science*. **62**, 502-522. <https://doi.org/10.1080/03650340.2015.1065607>. (2016)
32. Hamidi, M.E., Larabi, A., Faouzi, M. & Souissi, M. Spatial distribution of regionalized variables on reservoirs and groundwater resources based on geostatistical analysis using gis: case of rmel-oulad

ogbane aquifers (larache, nw morocco). *Arabian Journal of Geoences*. **11**, 1-18.

<https://doi.org/10.1007/s12517-018-3430-9>. (2018)

33. Bogunovic, I. et al. Reducing sampling intensity in order to investigate spatial variability of soil pH, organic matter and available phosphorus using co-kriging techniques. A case study of acid soils in Eastern Croatia. *Archives of Agronomy and Soil Science*. **63**, 1852-1863. <https://doi.org/10.1080/03650340.2017.1311013>. (2017)
34. Oliver, M.A. & Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*. **113**, 56-69. <https://doi.org/10.1016/j.catena.2013.09.006>. (2014)
35. Burgess, T.M. & Webster, R. Optimal interpolation and isarithmic mapping of soil properties.II. Block kriging. *European Journal of Soil Science*. **31**, 333-341. <https://doi.org/10.1111/j.1365-2389.1980.tb02085.x>. (2010)

Figures

Figure 1

Prediction error of soil pH at different point sets

Figure 2

Expression diagram of soil pH spatial variation of different point sets. The letters a-l represent the spatial variation expression of soil pH at 908, 797, 700, 594, 499, 398, 299, 200, 149, 100, 75 and 50 sampling points respectively.

Figure 3

Overview of the study area