Environ Monit Assess (2020) 192:300
https://doi.org/10.1007/s10661-020-08281-7

Application of geostatistics for grid and random sampling schemes for a grassland in Nigde, Turkey

Gülden Gök · Orhan Atakan Gürbüz

Received: 11 September 2019 / Accepted: 7 April 2020

Abstract Soil quality prediction maps are important tools for environmental scientists and policymakers. An 18 ha grassland was selected to create soil quality prediction maps. A total of 30 sampling points were selected, and samples were collected from top soil (0–20 cm depth). Twelve of the sampling points were selected randomly and 18 of the sampling points were selected based on a square shaped grid plan. The soil samples were then analyzed for pH, electrical conductivity (EC), organic matter (OM), water content, dissolved total carbon (DTC), dissolved organic carbon (DOC), dissolved inorganic carbon (DIC), and dissolved total nitrogen (DTN). Ordinary kriging (OK) and ordinary cokriging (OCK) spatial interpolation methods were used for the prediction of spatial distribution. The prediction errors showed that the parameters in the grid sampling scheme showed better prediction in the OK technique. The highest reduction in prediction errors was obtained in the DOC in grid sampling scheme after using OCK.

G. Gök (✉)
Department of Environmental Engineering, Aksaray University, E-90 Highway 7th Km, 68100 Aksaray, Turkey
e-mail: gokgulden@gmail.com

O. A. Gürbüz
Nigde Vocational School of Social Sciences, Nigde Omer Halisdemir University, Asagi Kayabasi Campus, Suleyman Fethi Avenue, Nigde, Turkey

Keywords Soil sampling · Soil parameters · Spatial statistics · Dissolved organic carbon · Ordinary kriging · Ordinary cokriging

Introduction

Digital soil maps or, in other words, prediction maps (Dharumarajan et al. 2019) have been used since 1990 (Zhang et al. 2017). Especially during the first quarter of the twenty-first century, statistical models have been used to create soil maps which are useful for environmental scientists (Li and Heap 2014) and other stakeholders in overcoming environmental issues (Keskin and Gruwald 2018). There is an important increase in soil quality mapping due to new policies around efficient land use management and sustainable development. The soil quality parameters are effective indicators when testing the sustainability of any land (Gong et al. 2015), particularly when we consider that economic development and increased quality of life can cause serious problems with regard to sustainable land management (Mandal et al. 2010).

When making such assessments, spatial distribution maps are often preferred to point source information (Li and Heap 2014). Due to the demand for continuous spatial information, geostatistical prediction techniques have been used in different branches of study since the 1960s (Oliver and Webster 2014), and geostatistical methods enable the estimation of unsampled points through modeling the spatial correlation between measured and predicted variables (Wu et al. 2009).

OK is one of the most used geostatistical techniques (Hou et al. 2017) when creating spatial distribution maps for different soil parameters such as pH, OM, EC (Houlong et al. 2016), soil organic carbon (Chabala et al. 2017), soil total nitrogen (Wu et al. 2009), and heavy metals (Liu et al. 2010). OCK, that is another geostatistical interpolation method, uses an auxiliary variable to decrease prediction errors of the model (Ersahin 2003), and the correlated variables have been used to decide auxiliary variables for more accurate predictions (Adhikary et al. 2017).

One of the most important steps of geostatistical analysis is the sampling scheme. In general, there are two sampling designs which are the following: statistical and geometric (Biswas and Zhang 2018). In simple sampling, sampling points are selected randomly (Brus and Gruijer 1997), although it has not been used by many studies (Biswas and Zhang 2018). Furthermore, one of the most prevalent sampling types is grid sampling wherein the distance between sampling points kept at a constant (Brus 2019). A study in 2016 done by Houlong et al. compared the two sampling plans for OK and found that simple random sampling showed better interpolations for pH, total phosphorus, and available phosphorus. In the same study, the grid sampling plan showed less prediction errors for OM, total nitrogen, and cation exchange capacity.

Since application of geostatistical modeling is important for generating soil quality prediction maps, in this study, we chose a grassland just outside downtown Nigde. The objectives of this study were to show the spatial distribution of soil quality parameters of the study area, to compare two sampling schemes for OK, and finally, in order to reduce prediction errors, to apply OCK for correlated soil parameters.

Materials and methods

Study area

The study site in question is located in the province of Nigde, which is in the central region of Turkey; the site is located around 2 km outside the city. The coordinates of the study area are 34°41′59.46″E and 37°56′59.67″N. The surface area of the field is around 18 ha. There is one bus terminal and one gas station to the north of the site. One afforestation site is also located on the south side. A cement factory is located to the southwest, and there is additionally the D805 intercity road and the O21 south Nigde road located on each side of the study area. The slope of the study site is increasing from east to west. Furthermore, the field can be classified as a grassland. Nowadays, the site has been selected by Governorship of Nigde as one of the afforestation sites of territory, and *Pinus brutia* (Turkish pine) has been planted on the study site (Fig. 1).

The climate of the study site is denoted as a steppe climate and has a mean precipitation of 338 mm. The difference in precipitation between the wettest and driest months is about 50 mm. The average annual temperature is around 10.1 °C and the coldest month was January with an average temperature of −0.2 °C. The hottest month of the area was July with a mean temperature of 21.3 °C (AM Online Projects n.d.).

Soil sampling and analysis

When determining the sampling points, three softwares were used throughout the research. Google Earth Pro was one of them and was used for identifying the borders of the grassland. Secondly, ArcGIS v. 10.5 employed when projecting geographical coordinates and, lastly, PNNL's Visual Sample Plan (VSP) software was used to identify the grid sampling point locations and random sampling point locations (Saito and McKenna 2007). A total of 30 soil sampling points were selected (Fig. 1), and their coordinates were transferred to a GPS device. The distance between the points in the grid sampling scheme ($N = 18$) was 100 m.

Soil samples were collected in November 2018. Samples were collected from the top soil (0–20 cm) via using a shovel and placed in sampling bags.

The Hach HQ411D branded digital laboratory EC and pH meter was used for analyzing the pH and EC of sampled soil. Water content and OM were also analyzed through gravimetric methods. For water, content samples were dried for 24 h at 104 °C. For OM, samples were burned for 2 h at 600 °C.

The soil samples were dried and sifted through a 0.5 mm sieve before DTC, DOC, DIC, and DTN analysis. Before starting the analysis, 10 g of samples were taken, and 160 ml of purified water was added. In order to set the pH level at 5.0, 0.5 N glacial acetic acid was added to samples, and then samples were mixed for 24 h for eluate extraction. At the end of the 24 h, purified water was added to the marked part, and filtration was performed (Lowenbach 1978). After filtration, samples were placed

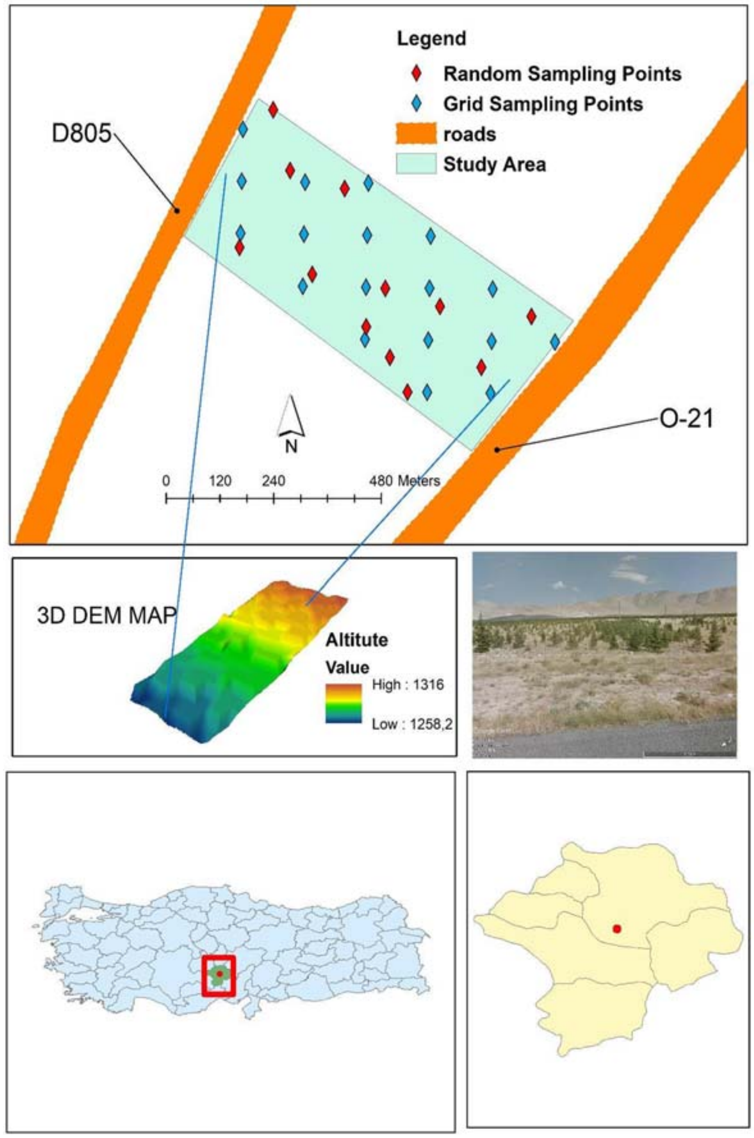

Fig. 1 , DEM map, distribution of the sampling locations, and location of the study site distances among points in simple random sampling scheme ($N = 12$) were varied between 60 and 700 m.

in centrifuge tubes and stored at 4 °C. The stored samples were diluted 50 times before analyzed in the Shimadzu Total Organic Carbon Analyzer and the Total Nitrogen Analyzer Kit for catalytic oxidation (EPA Method 9060a).

Statistical methods

IBM SPSS v. 18 was used to estimate the deterministic statistics for pH, water content, EC, OM, DTC, DOC, and DTN from 30 sampling points. Maximum value, minimum value, standard deviation, mean value, skewness, and kurtosis values were determined for the soil parameters. To test the distribution of variables, Kolmogorov-Smirnov and Shapiro-Wilk tests were used, and the Pearson correlation coefficient was also estimated to test the statistical correlation of soil parameters (Paz-Gonzalez et al. 2001).

Spatial statistics

To show the characteristics of a random field, spatial autocorrelation and spatial stratified heterogeneity are important statistical indicators that are used together. The term spatial autocorrelation explains the relationship between characteristics of spatial data and their location. Closer sites have similar attributes and vice versa (Wang et al. 2016). Moran's I statistic is an index of spatial autocorrelation and calculated as follows (Eq. 1):

$$I = \left[\frac{n}{\sum_{i=1}^{n}(y_i - y)^2}\right] \times \left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - y)\left(y_j - y\right)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}\right] \qquad (1)$$

where ij is the locations, n is the number of parameters, y is the overall mean, and y_i and y_j are means at specific locations (O'Sullivan and Unwin 2010).

The measure of spatial autocorrelation at the local level could identify the heterogeneity via spatial cluster patterns. Local Moran's I index (LISA) can be estimated as (Eq. 2):

$$I_i = \frac{z_i - z}{\sigma^2}\sum_{j=1,\ j\neq 1}^{n}\left[W_{i,j}\left(z_j - z\right)\right] \qquad (2)$$

where z_i is the value of variable at the ith location, z_j is the value of variable at the jth location, $w_{i,j}$ is the weighted distance between i and j, and σ^2 is the variance of variable (Fu et al. 2014).

Interpolation methods

The spatial distribution of parameters was interpolated through OK which is a geostatistical technique that uses semi variance modeling in order to estimate unknown values. The equation of OK is as follows (Eq. 3):

$$Z^{'}(x_0) = \sum_{i=1}^{n} \quad \lambda_i Z(x_i) \qquad (3)$$

where λ_i is the kriging weight, $Z^{'}(x_0)$ is the interpolated value of DOC of the unsampled location, and $Z(x_i)$ is the known value of DOC from sampled location (Chabala et al. 2017)

OCK is an advanced modeling technique delivered from OK. OCK is a geostatistical interpolation technique that uses two correlated parameters together to estimate unknown values and their spatial distribution. OCK estimation is as follows (Eq. 4):

$$Z^{'}(x_0) = \sum_{i_1}^{n} \quad \omega_{i1} Z(x_{i1}) + \sum_{i_2}^{m} \quad \omega_{i2} V(x_{i2}) \qquad (4)$$

(with) $\sum_{i_1=1}^{m}\omega_{i1} = 1; \sum_{i_2=1}^{m}\omega_{i2} = 0$ where $Z^{'}(x_0)$ estimated value of the first parameter, ω_{i1} and ω_{i2} are the kriging weights according to sampling points, Z and V are the first and second parameter, respectively, and n and m are the number of sampling points for both parameters (Adhikary et al. 2017).

Data validation

Data validation procedure for this study included leave-one-out cross validation (LOOCV), average standardized error (ASE), root mean square error (RMSE), and root mean square standardized error (RMSSE) (Houlong et al. 2016) (Chabala et al. 2017) (Gong et al. 2014).

LOOCV statistical method was used for the precision of the models that were used for interpolation of measured data. This cross-validation technique eliminated a known variable and tested the model results with predicted variable of the eliminated point (Chabala et al. 2017). Arc GIS v 10.5 automatically uses LOOCV for interpolation methods that were used in this study (Gong et al. 2014).

ASE, RMSE, and RMSSE are described in Eq. 5, Eq. 6, and Eq. 7, and these prediction error values for measured data were estimated in IBM SPSS v. 18.

$$ASE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \quad \sigma(i)} \qquad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \quad [Z(x_i) - Z'(x_i)]^2} \qquad (6)$$

$$RMSSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \quad \left[\frac{Z(x_i) - Z'(x_i)}{\sigma(i)}\right]^2} \qquad (7)$$

where $Z(x_i)$ is the measured value of any parameter, and $Z'(x_i)$ is the predicted value of the same variable. σ(i) is the standard error for the location i, and N is the total number of samples (Houlong et al. 2016) (Chabala et al. 2017) (Gong et al. 2014).

When estimating the reduction of the prediction errors between OK and OCK, the relative reduction in RMSE (RRMSE) was used. The evaluated method was OCK, and the estimated one was OK. RRMSE can be defined by (Eq. 8) as:

$$RRMSE = \frac{RMSE_{OK} - RMSE_{OCK}}{RMSE_{OK}} \times 100\% \qquad (8)$$

where $RMSE_{OK}$ is the root mean square error of any parameter estimated after OK, $RMSE_{OCK}$ is the root mean square error of any parameter estimated after OCK, and $RRMSE$ is the percent reduction of error between OCK and OK (Pang et al. 2009).

Results and discussion

Descriptive statistics

The list of descriptive statistics (the skewness of pH, water content, OM, DTC, and DTN for grid sampling plan) can be seen in Table 1. The soil pH ranges from 8.16 to 8.59, and the average grassland pH was 8.41 for random sampling points. EC had the highest skewness among other parameters for grid sampling, and similar results were found by Korkanç et al. (2018) in the same region near Akkaya Dam in Nigde. According to the results of that study, the pH value of the grassland near Akkaya Dam was 8.57, and the pH value in a poplar grove near the dam was 8.02. The mean moisture holding capacity for the grassland area and the poplar plantation area were 15.92 and 15.10, respectively. The mean EC value was 191.99 µS/cm for the grassland.

The level of dispersion of the variables was determined through coefficient of variation (CV). If the CV value is between 0.1 and 0.9, then the distribution level will be moderate. CV values lower than 0.1 are considered as poor distribution, and where CV values are higher than 0.9, then it is considered as significant dispersion (Yao et al. 2019). The dispersion levels were moderate for DTC, DOC, DTN, and DIC in both sampling plans (see Table 1).

Table 1 also shows the Kolmogorov-Smirov (P_{k-s}) and Shapiro-Wilk (P_{s-w}) test results. According to the results of the normality test, the water content, pH, OM, DOC, DTC, DTN showed normal levels of distribution for points planned on the grid scheme. In simple random sampling, the values of pH, EC, DOC, DTC, and DTN were also distributed normally. In the geostatistical analysis, estimation methods give statistically significant results with normally distributed data (Webster and Oliver 2007). Also, the Pearson correlation matrix, the prediction models, and maps were estimated based on normally distributed parameters.

Table 2 shows the Pearson correlation matrix. According to Table 2 based on grid sampling plan, there was a positive relationship between water content OM and DOC-DTC. The level of correlation between water content and OM was significant at 0.05. There was a strong correlation between DOC and DTC at 0.01. Among the points in the simple random sampling scheme, the only

Table 1 Descriptive statistics of 30 sampling points

Sampling plans	Soil Parameters	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis	CV	p_{k-s}	p_{s-w}
Grid sampling plan $N=18$	Water content [%]	0.03	0.07	0.05	0.01	0.66	2.97	0.19	0.18	0.11
	pH	8.33	8.58	8.47	0.07	−0.49	−0.23	0.01	0.20	0.81
	EC [µS/cm]	108.30	557.00	175.08	99.91	3.65	14.34	0.57	0.00	0.00
	OM [%]	5	8	6	0.01	−0.15	−0.20	0.12	0.20	0.95
	DOC [mg/L]	1.17	3.23	2.00	0.50	0.35	1.35	0.25	0.01	0.06
	DTC [mg/L]	1.32	3.35	2.12	0.49	0.43	1.48	0.23	0.01	0.06
	DIC [mg/L]	0.11	0.15	0.12	0.01	0.62	−0.97	0.12	0.01	0.04
	DTN [mg/L]	0.21	0.93	0.58	0.17	−0.33	0.31	0.30	0.20	0.83
Simple random Sampling plan $N=12$	Water content [%]	0.03	0.07	0.04	0.01	1.20	0.91	0.29	0.01	0.03
	pH	8.16	8.59	8.41	0.12	−0.63	0.60	0.01	0.20	0.78
	EC	114.50	165.80	136.99	15.59	0.46	−0.30	0.11	0.20	0.70
	OM[%]	5	11	7	0.01	1.69	4.39	0.19	0.19	0.04
	DOC [mg/L]	1.20	2.78	2.06	0.44	−0.24	−0.04	0.21	0.20	0.99
	DTC [mg/L]	1.77	3.42	2.36	0.48	0.90	0.63	0.20	0.20	0.43
	DIC [mg/L]	0.12	2.21	0.30	0.60	3.46	12.00	2.01	0.00	0.00
	DTN [mg/L]	0.29	0.99	0.55	0.20	1.18	1.04	0.37	0.20	0.11

CV coefficient of variation, $p_{k\text{-}s}$ *p* value of Kolmogorov-Smirov test, $p_{s\text{-}w}$ *p* value of Shapiro-Wilk test

Table 2 Pearson correlation matrix

Sampling plan	Parameters	Water content	pH	OM	DOC	DTN	DTC
Grid sampling plan $N = 18$	Water content	1.00					
	pH	0.18	1.00				
	OM	0.54*	− 0.32	1.00			
	DOC	0.13	− 0.15	0.38	1.00		
	DTN	0.24	0.09	0.20	0.46	1.00	
	DTC	0.14	− 0.15	0.38	1.00**	0.45	1.00
Sampling plan	Parameters	pH	EC	DOC	DTN	DTC	
Simple random sampling plan $N = 12$	pH	1.00					
	EC	− 0.48	1.00				
	DOC	− 0.11	− 0.09	1.00			
	DTN	− 0.10	0.05	0.65*	1.00		
	DTC	− 0.41	− 0.20	0.14	0.30	1.00	

*Correlation is significant at 0.05 level

**Correlation is significant at 0.01 level

correlation was found between DOC and DTN. These values were used in the OCK model in order to obtain a precise spatial distribution. A similar correlation was done by Yao et al. (2019) between soil organic carbon stock values and topographic factors. Yao et al. (2019) estimated that there was a positive relationship between elevation values and soil carbon stock values. Xu et al. (2017) also found a positive and statistically significant relationship between soil organic carbon and soil nitrogen.

Spatial autocorrelation and cluster/outlier analysis

Global and local Moran's I were estimated for DTC, DOC, and DTN for grid and random sampling schemes as can be seen in Table 3. According to Table 3, Moran's I results showed that there was a possitive autocorrelation in DTN in grid sampling scheme and DTC in random sampling scheme. DTN in grid sample plan was the strongest possitive autocorrelation. The strongest neggative autocorrelation was DOC in simple random sampling scheme.

Table 3 Moran's I and LISA results

Sampling design	Soil parameters	Moran's I	z-score	p value	Pattern	LISA
Grid sampling plan $N = 18$	DOC [mg/L]	− 0.18278	− 0.704401	0.481183	Random	1; cluster:high 1;cluster:low 1; high outlier
	DTC [mg/L]	− 0.18424	− 0.715202	0.474484	Random	1; cluster:high 1;cluster:low 1; high outlier
	DTN [mg/L]	0.372094	2.380418	0.017293	Clustered	1; cluster low
Simple random sampling plan $N = 12$	DOC [mg/L]	− 0.01587	0.291199	0.770899	Random	Not significant
	DTC [mg/L]	0.154509	0.976397	0.328868	Random	1;low cluster
	DTN [mg/L]	− 0.13628	− 0.183441	0.854452	Random	2;low outlier

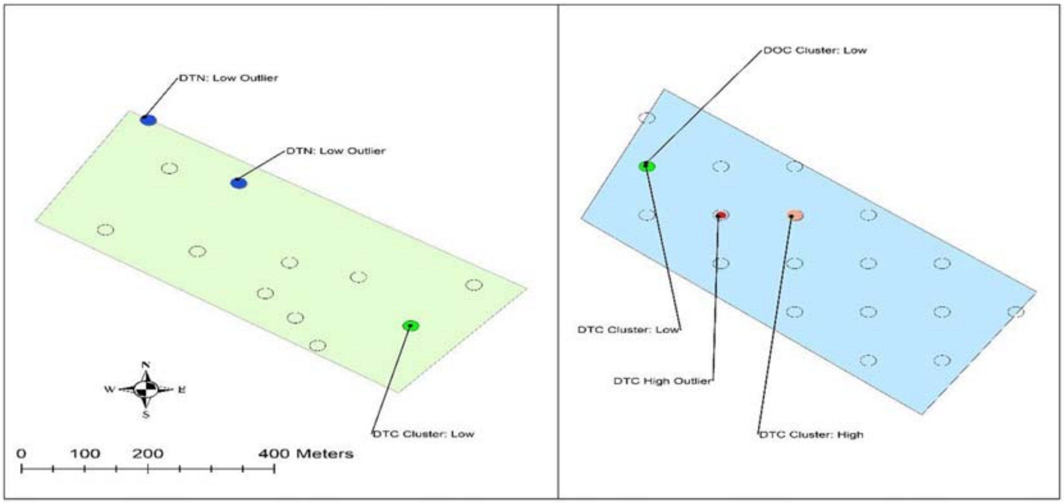

Fig. 2 Spatial outliers and spatial clusters map

As it can be seen in Fig. 2 in simple random sampling plan, a low-low cluster was observed for DTC and two low outliers located to the north part of the study field. High-low clusters and a high outlier for DTC for grid sampling plan were detected in the middle and upper left of the study area. A low-low cluster was also found for DOC in grid sampling scheme.

Geostatistical analysis of soil quality parameters

Tables 4 and 5 exhibit the semivariogram model parameters for the grid and random sampling scheme for soil variables. Semivariogram models were used so as to fit experimental datasets to the semivariogram function which helps in the estimation of unsampled locations (Fanchi 2010). The Gaussian model was the best fit for the semivariogram modeling for pH. The Stable model was selected for the rest of the soil variables. For randomly sampled values, the Gaussian model was the best fitting model for DTC (Table 5).

The range values show the border of the autocorrelation. There will not be any spatial correlation for distances greater than the range value (Pouladi et al. 2019). The maximum range value was 218.2 m for DOC (Table 4) for grid sampling, and DTN had the highest range among others for random sampling (Table 5).

Nugget/sill ratio (Bhunia et al. 2018) is also known as nugget semivariance, and with this ratio, the degree of spatial correlation can be tested. A nugget/sill ratio lower than 0.25 represents strong spatial correlation. An average spatial dependency could be seen with a nugget/sill ratio between 0.25 and 0.75. The weakest spatial dependency is found when the nugget/sill ratio is higher than 0.75

Table 4 Semivariogram model parameters for grid sampling scheme ($N = 18$) for OK

Soil parameter	Nugget	Range [m]	Partial sill	Sill	Lag size	Nugget/sill	Model
pH	0.0014	203.5	0.0014	0.0028	25.5	0.5	Gaussian
DOC	0.1712	218.2	0.0841	0.2553	27.27	0.67	Stable
DTN	0	203.5	0.0226	0.0226	27.27	0.00	Stable
DTC	0.1661	218.2	0.083	0.2491	27.27	0.67	Stable
Water content	0	211.7	0.0001	0.0001	25.85	0.00	Stable
OM	0	203.5	0.000006	0.000006	25.67	0.00	Stable

Table 5 Semivariogram model parameters for simple random simple sampling scheme ($N = 12$) for OK

Soil parameter	Nugget	Range [m]	Partial sill	Sill	Lag Size	Nugget/sill	Model
DOC	0.1289	251.7	0.0048	0.1337	31.47	0.96	Stable
DTC	0.101	369.8	0.133	0.234	46.22	0.43	Gaussian
DTN	0.0248	710.5	0.0455	0.0703	62.93	0.35	Stable
pH	0.0145	755.3	0	0.0145	62.94	1	Stable
EC	243.07	755.3	0	243.07	62.94	1	Stable
OM	0.00019	755.3	0	0.00019	62.94	1	Stable

(Cambardella et al. 1994). The strongest dependents among grid sampling variables were DTN, water content, and OM (Table 4). All other parameters dependency levels were moderate, and for randomly collected samples, the weakest spatial dependency could be seen in DOC.

The values of pH, OM, and EC in simple random sampling were not consistent with the locations of sample points. The semivariograms of these parameters were irrelevant with spatial correlation (pure nugget effect) (Monterı et al. 2015) (Table 5) so no geostatistical interpolation was used for them.

Figures 3 and 4 present the OK results for the grid sampling and simple random sampling plan respectively. In Fig. 3, the lowest percent of water content can be found in the eastern side of the grassland. The distribution of DOC and DTC could be seen clearly in Fig. 3. The majority of the DOC values ranged between 1.73 and 2.12 mg/L. In Fig. 4, the spatial distribution of

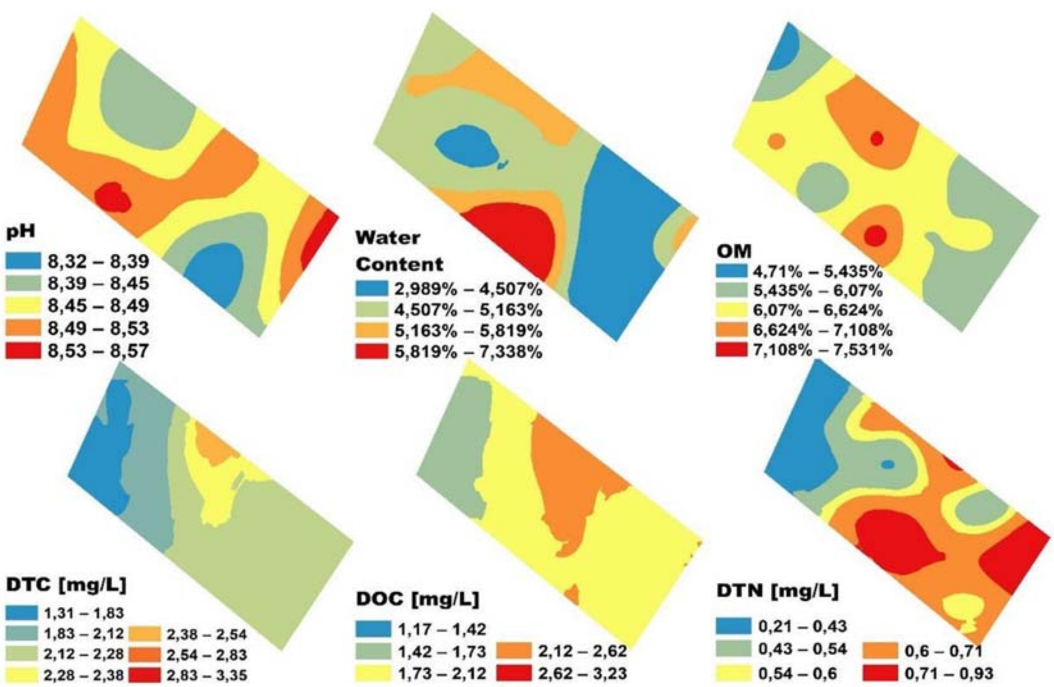

Fig. 3 OK interpolation of soil parameters based on grid sampling ($N = 18$)

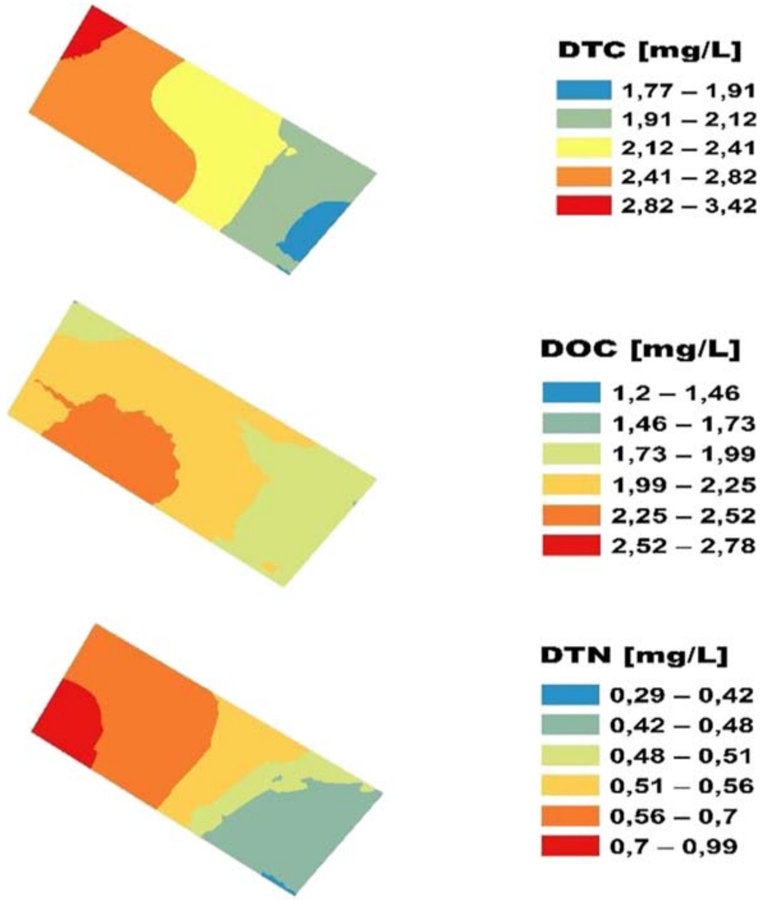

Fig. 4 OK interpolation of soil parameters based on simple random sampling ($N = 12$)

DTN and DTC were similar based on the contour colors. According to simple random sampling, the DOC values ranged between 1.99 and 2.25 mg/L in the central part of the field of study.

Accuracy of OK models

According to Table 6, the coefficient of determination (R^2) was the highest for DTN in the simple

Table 6 Cross-validation results of 30 sampling points

Sampling type	Parameters	Efficiency	Prediction errors		
		R^2	ASE	RMSE	RMSSE
Random	DTC	0.182	0.408	0.416	0.984
	DOC	0.0133	0.429	0.426	0.971
	DTN	0.001	0.186	0.224	1.174
Grid	DTC	0.01	0.496	0.491	0.9888
	DOC	0.015	0.502	0.497	0.986
	DTN	0.405	0.108	0.13	1.162
	pH	0.02	0.064	0.066	1.018
	Water content	0.09	0.007	0.009	1.249
	OM	0.00036	0.007	0.008	1.122

Table 7 Semivariogram model parameters for OCK

Sampling plan	Soil parameter	Auxiliary parameter	Nugget	Range	Partial sill	Sill	Lag size	Nugget/ sill	Model
Grid Sampling	OM	Water Content	0	201.34	0.00001	0.00001	25.17	0.00	Exponential
Grid Sampling	DOC	DTC	0.0066	201.34	0.2623	0.2689	25.17	0.02	Exponential
Simple Random Sampling	DOC	DTN	0.1289	251.75	0.0483	0.1772	31.47	0.73	Gaussian

random sampling scheme. The ASE values were higher than RMSE or RMSSE values which were lower than 1; there was an overestimation among modeled values (Houlong et al. 2016). In the grid sampling scheme, the predicted values of DTN were similar with the modeled values (Table 6). The distribution of DTN in the grid sampling plan was better predicted than in the simple random sampling plan. The prediction errors were also lower for water content and OM. The highest prediction errors could be seen in DOC values for both sampling plans. The lowest R^2 value was estimated from the grid sampling scheme for OM; however, its values were lowest in ASE and RMSE among other parameters.

OCK with auxiliary parameter for grid sampling and simple random sampling schemes

To increase the R^2 values, the OCK method was used for OM values in grid sampling and DOC values for both sampling plans. The auxiliary parameters were chosen as the Person correlation values as shown in Table 3. The best fitting models were exponential for OM in the grid sampling plan, exponential for DOC in the grid sampling plan, and Gaussian for DOC in the simple

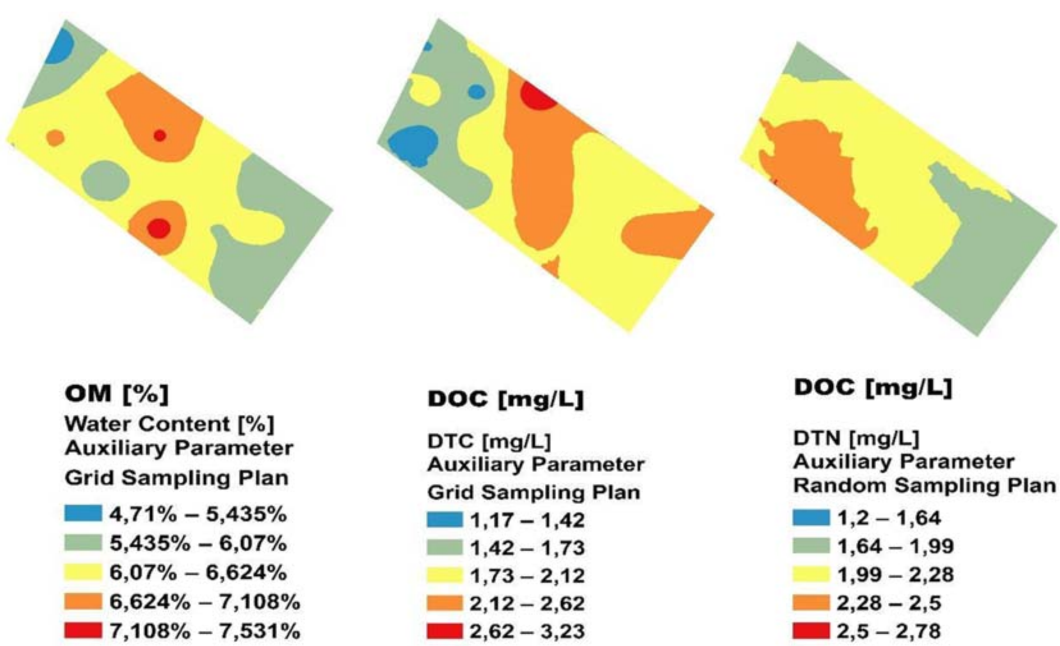

Fig. 5 OCK prediction results of OM and DOC for grid and simple random sampling plans

Table 8 Cross validation results of OCK interpolation for OM and DOC

Sampling type	Parameters	Auxiliary parameters	Efficiency	Prediction errors			Comparison with OK	
			R^2	ASE	RMSE	RMSSE	$RMSSE_{OK}$	RRMSE [%]
Grid sampling	OM	Water content	0.138	0.007	0.007	0.991	1.122	12%
Grid sampling	DOC	DTC	0.9521	0.3856	0.1837	0.4747	0.986	52%
Simple random Sampling	DOC	DTN	0.2776	0.4236	0.3801	0.8793	0.971	9%

sampling model. According to the nugget/sill ratios in Table 7, the spatial dependency was weakest for DOC in the simple random sampling plan. In the grid sampling scheme, the spatial dependency of OM was stronger than DOC.

Figure 5 shows the spatial distribution maps of OCK methods. According to the contour maps, the distribution of DOC in two sampling schemes showed similarities. The majority of DOC levels range between 1.42 and 2.48 mg/L in the east and central part of the study area. The distribution of OM was highest in the middle part of the study area with a ratio of 6%:6.6%.

Table 8 shows the data validation values of OCK for OM and DOC values. According to prediction errors, the OCK method worked best for DOC in the grid sampling plan with DTC with auxiliary parameter. The coefficient of determination of DOC in grid sampling was 0.9521, and reduction in RMSSE by using OCK was highest among other parameters.

Conclusion

The deterministic statistical results showed that the soil quality parameters of the grassland have heterogeneous distribution for many variables. pH, EC, DOC, DTC, and DTN in the simple random sampling scheme were normally distributed. In the grid plan, pH, OM, DOC, DTC, DTN values showed normal distribution. After modeling the semivariograms for parameters, DTC had the highest R^2 values in both sampling schemes. According to the cross validation results, the highest prediction error reduction was obtained from DOC with auxiliary parameter of DTC which is also with the highest R^2 value as well.

From this set of materials, we can conclude that grid sampling plan showed better results for OK. For further study goals, it is important to increase the number of samples by lowering the grid spacing to avoid pure nugget effects.

References

Adhikary, S. K., Muttil, N., & Yilmaz, G. A. (2017). Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes, 31*, 2143–2162.

AM Online Projects. (n.d.). *Climate-Data.Org*. OpenStreetMap Contributors. https://tr.climate-data.org/asya/tuerkiye/nigde/nigde-254/#temperature-graph. Accessed June 2019.

Bhunia, G. S., Shit, P. K., & Maiti, R. (2018). 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.

Biswas, A., & Zhang, Y. (2018). Sampling designs for validating digital soil maps: A review. *Pedosphere, 28*, 1–15.

Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma, 338*, 464–480.

Brus, D. J., & Gruijer, J. J. (1997). Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma, 80*, 1–44.

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., et al. (1994). Field-scale variability of soil properties in Central Iowa soils. *Soil Science Society of America Journal, 58*, 1501–1511.

Chabala, L. M., Mulolwa, A., & Lungu, O. (2017). Application of ordinary kriging in mapping soil organic carbon in Zambia. *Pedosphere, 27*, 338–343.

Dharumarajan, S., Hegde, R., Janani, N., & Singh, S. K. (2019). The need for digital soil mapping in India. *Geoderma Regional, 16*, e00204.

Ersahin, S. (2003). Comparing ordinary kriging ad cokriging to estimate infiltration rate. *Soil Science Society of America, 67*, 1848–1855.

Fanchi, R. J. (2010). Integrated Reservoir Asset Management Principles and Best Practices. In *Reservoir Characterization* (pp. 196–198). Oxford-Burlington: Gulf Professional Publishing.

Fu, W. J., Jiang, P. K., Zhou, G. M., & Zhao, K. L. (2014). Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeosciences, 11*, 2401–2409.

Gong, G., Mattevada, S., & O'Bryant, S. E. (2014). Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environmental Research, 130*, 59–69.

Gong, L., Ran, Q., He, G., & Tiyip, T. (2015). A soil quality assessment under different land use types in Keriya river basin, southern Xinjiang, China. *Soil & Tillage Research, 146*, 223–229.

Hou, D., O'Connor, D., Nathanail, P., Tian, L., & Ma, Y. (2017). Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review. *Environmental Pollution, 231*, 1188–1200.

Houlong, J., Daibin, W., Chen, X., Shuduan, L., Hongfeng, W., et al. (2016). Comparison of kriging interpolation precision between grid sampling scheme and simple random sampling scheme for precision agriculture. *Eurasian Journal of Soil Science, 5*, 62–73.

Keskin, H., & Gruwald, S. (2018). Regression kriging as a workhorse in the digital soil mapper's toolbox. *Geoderma, 326*, 22–41.

Korkanç, S. Y., Şahin, H., Özden, A. O., & Özkurt, B. (2018). The effects of land use conversion on soil organic carbon and selected properties of soils: A case of Niğde province. *Turkish Journal of Forestry, 19*, 362–367.

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software, 53*, 173–189.

Liu, G.-S., Jiang, H.-L., Liu, S.-D., Wang, X.-Z., Shi, H.-Z., et al. (2010). Comparison of Kriging interpolation precision with different soil sampling intervals for precision agriculture. *Soil Science, 175*, 405–415.

Lowenbach, W. (1978). *Compilation and evaluation of leaching test methods. Resource Document*. Environmental Protection Agency. https://nepis.epa.gov/Exe/ZyPDF.cgi/300064UZ.PDF?Dockey=300064UZ.PDF. Accessed July 2019.

Mandal, D., Singh, R., Dhyani, S. K., & Dhyani, B. L. (2010). Landscape and land use effects on soil resources in a Himalayan watershed. *Catena, 81*, 203–208.

Monterı, J., Fernandez-Aviles, G., & Mateu, J. (2015). Spatial and spatio-Temporal geostatistical Modeling and kriging. In *Structural Analysis* (1st ed., pp. 35–38). Chichester: John Wiley & Sons.

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena, 113*, 56–69.

O'Sullivan, D., & Unwin, D. J. (2010). *Geographic Information Analysis*. Hoboken: John Wiley & Sons.

Pang, S., Li, T.-X., Wang, Y.-D., Yu, H.-Y., & Li, X. (2009). Spatila interpolation and sample size optimization for soil copper (Cu) investigation in cropland soil at county scale using cokriging. *Agricultural Sciences in China, 11*, 1369–1377.

Paz-Gonzalez, A., Castro Taboada, M. T., & Vieira, S. R. (2001). Geostatistical analysis of heavy metals in a one-hectare plot under natural vegetation in a serpentine area. *Canadian Journal of Soil Science, 81*, 469–479.

Pouladi, N., Moller, B. A., Tabatabai, S., & Greve, M. H. (2019). Mapping soil organic matter contents at field level with cubist, random forest and kriging. *Geoderma, 342*, 85–92.

Saito, H., & McKenna, S. A. (2007). Delineating high-density areas in spatial Poisson fields from strip-transect sampling using indicator geostatistics: Application to unexploded ordnance removal. *Journal of Environmental Management, 84*, 71–82.

Wang, J.-F., Zhang, T.-L., & Fu, B.-J. (2016). A measure of spatial stratified heterogeneity. *Ecological Indicators, 67*, 250–256.

Webster, R., & Oliver, M. A. (2007). Basic Statistics. In *Geostatistics for Environmental Scientists* (2nd ed., pp. 18–19). West Sussex: John Wiley & Sons, Chichester.

Wu, C., Wu, J., Luo, Y., Zhang, L., & DeGloria, S. D. (2009). Spatial estimation of soil total nitrogen using cokriging with predicted soil organic matter content. *Soil Science Society of America, 73*, 1676–1681.

Xu, Y., Pu, L., Liao, Q., Zhu, M., Yu, X., et al. (2017). Spatial variation of soil organic carbon and total nitrogen in the coastal area of mid-eastern China. *International Journal of Environmental Research and Public Health, 780*, 1–12.

Yao, X., Yu, K., Deng, Y., Zeng, Q., Lai, Z., et al. (2019). Spatial distribution of soil organic carbon stocks in Masson pine (Pinus massoniana) forest in subtropical China. *Catena, 178*, 189–198.

Zhang, G., Liu, F., & Song, X. (2017). Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture, 16*, 2871–2865.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.