Malaysian Journal of Soil Science 2022 Vol. (26): 17-26

Spatial Variability of Soil Properties Using R Language

Amany M. Hammam

Central Laboratory Environment Quality Monitoring (CLEQM) – National Water Research Center (NWRC) – P.O. Box 13621/6, El-Kanater, Egypt

Corresponding author: Amany2010@hotmail.it

ABSTRACT

Accurate assessment of the spatial variability of soil properties is key component of the agriculture ecosystem and environment modelling. The main objective of the present study is to measure some of the soil properties and their spatial variability. Conventional analytical methods and geostatistical methods were used to analyse the data for spatial variability. During a period of 2018 – 2019, soil samples (n=23) were collected in the field through random sampling in the eastern part of Menoufia governorate (30˚ 50′ N and 31˚ N). Soil properties of soil Cation Exchange Capacity (CEC), electric conductivity (EC), percentage of soil Clay, Silt and Sand were estimated using the geostatistical approach methods. An ordinary kriging (OK) interpolation was used for direct visualization of soil properties. The semivariograms of the four soil properties were fit with Gaussian curve, except EC was fit with exponential curve. The results showed the effectiveness of statistical analysis and interpretation in sense of the obtained data. Cross-validation of variogram models through OK representing in ME showed that the spatial prediction of the selected soil properties is high. The present study suggests that the OK interpolation could potentially revealed the spatial distribution of soil properties and the sample distance in this study for interpolation.

Keywords: **Soil properties, spatial variability, semivariograms, ordinary kriging, mean error**

INTRODUCTION

Soils are the physical and synthetic results of the interactions among hydrology, geology, geomorphology, climate and biosphere. The interpretation and understanding of the resulting soil patterns and their spatial analysis were a major interest in the reconstruction field of environmental change and in cases of emphasising the earth surface system. Soil maps and their specified reports are considered essential information sources in relation to land resources. Thus, spatial distribution of soil properties has intensively been studied by many researchers (Wang *et al.* 2013; Monda *et al.* 2017; Xu *et al.* 2018)

In general, traditional geostatistical methods focuses on the prediction and description of the quantitative variables, so it is so hard to interpolate qualitative data using these techniques. Various techniques are used for mapping spatial distribution of qualitative variables, but it still requires high technical support and specialized knowledge (Fabiy *et al.* 2013; Wang *et al.* 2017). Nowadays, kriging is widely considered as a technique that can be used for predicting continuous soil characteristics at un-sampled locations. Bhunia *et al.* (2016) indicated that unequal to continuous variables, the qualitative variables cannot be considered in determining as mere linear combinations of neighbouring observations.

R environment is an open-source used for manipulating data, statistical analysis, and data visualization. R is considered as a modular system where it contains principle packages and other standard packages that are loaded after starting R. Besides, to achieve specific tasks, there are several thousand packages, such as, Gstat, GoeR, Lattice, etc. (Pebesma 2004). This paper applies ordinary kriging (OK) as a method using R language and Gstat package to estimate the prediction of soil properties at the observation points.

MATERIALS AND METHODS

The current research was carried out in different regions in Menoufia governorate, especially the centre and western parts of the area, in 2018-2019, to study the spatial distribution of some soil properties such as Electric Conductivity (EC); Cation Exchange Capacity (CEC); Sand; Silt and Clay contents (Figure 1).

Soil samples were collected randomly from surface soils (0-30 cm depth) from 23 locations across the eastern part of the study area. The exact location of the soil samples was precisely defined in the field using GPS and then, plotted on the map (Figure 1). XY plot using R language clarified samples distribution in the study area (Figure 2). Soil samples were collected using hand auger and prepared for analysis (air-dried, crushed and passed through a 2 mm sieve). In accordance to Adepetu *et al. (2000), s*ampling units were selected randomly and independently and were irrespective of any judgment regarding spots previously taken. Soil CEC, EC, Clay, Silt and Sand were analysed in accordance to Richards (1954) and Van Reeuwijk (2002).

R language 4.2.1 software was applied and semi-variogram was used in order to evaluate the spatial distribution pattern of each soil property. Semi-variogram was calculated using the following Equation (Behera *et al.* 2018).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(X_\alpha + h]^2$$

Where $\gamma(h)$, $N(h)$, $Z(X_\alpha)$ and $Z(X_\alpha + h)$ represents semi-variance for the lag distance h, number of sample pairs that separated by lag distance h, the measured value at α sample location and the measured value at point $\alpha + h$ sample location, respectively.

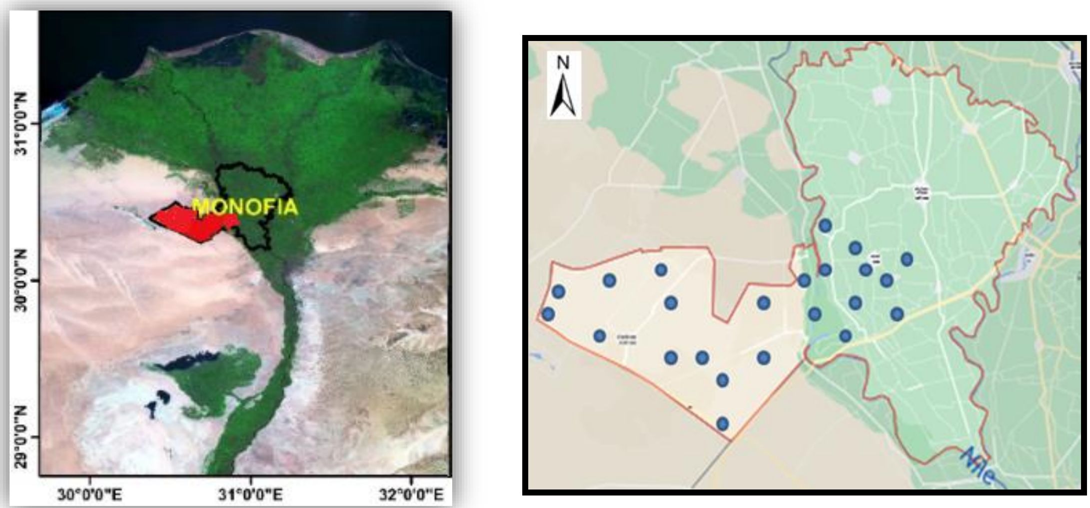

Figure 1. Samples collected from different regions in Menoufia Governorate

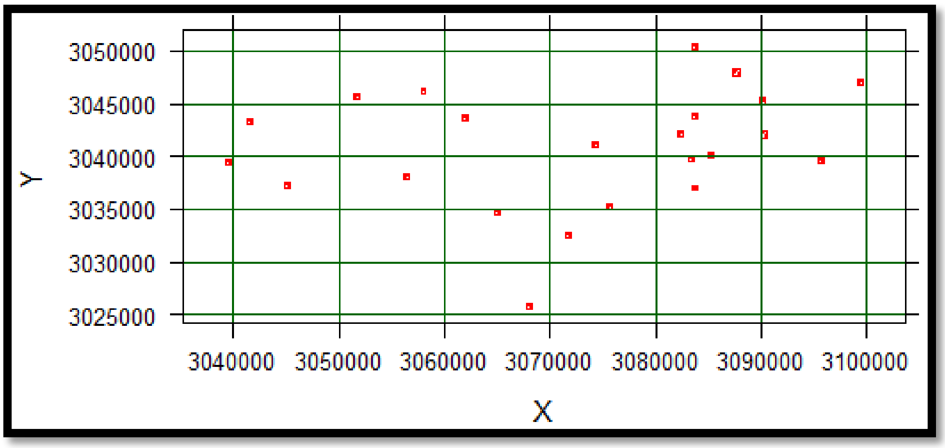

Figure 2. XY plot using R clarified samples distribution

Spatial dependence Mean error (ME) is considered one of several tools that usually used for evaluating semi-variogram models. In general, Gundogdu and Guney (2007) proposed that the best fit model was obtained to have mean error "ME value close to zero". The equation of criteria is as follows (Johnston *et al.,* 1996).

$$ME = \frac{1}{N}\sum_{i=1}^{N}[Z(x_i) - \hat{Z}(x_i)]$$

Where(x_i): refers to the observed value; $\widehat{Z}(x_i)$: the predicted value;N : the number of values for location i.

Cambardella *et al.* (1994) reported that semi-variogram model is based on the differences between nugget and sill or nugget to sill ratio, which may be strong (<0.25), moderate (0.25 – 0.75) and weak (> 0.75).

Interpolation mapping was applied using ordinary kriging technique, which is considered as a more reliable method than other methods used (Meul and Van Meirvenne 2003), for determining the soil properties values at unsampled locations. Interpolation mapping is an unbiased predictor for the random process like reducing impact of outliers (Triantafilis *et al.* 2001). Kriging technique works on weighting the surrounding measured values to derive a prediction for unmeasured locations (Meul and Van Meirvenne 2003).

In other words, by using the kriging method, the weights are dependent not only on distance between the measured points and the predicted locations, but also on the overall spatial arrangement of the measured points. The spatial autocorrelation must be quantified in case of using the spatial arrangement in weights, Thus, in ordinary kriging, the weightage is based on any fitted model to the measured points, the distance to the predicted location, and the spatial relationships among the measured and the predicted location. In this study, different semi-variogram models were evaluated depending on the nugget to sill ratio. Ordinary kriging validation was also evaluated based on the value of mean error (ME) for the selected soil properties in this study.

RESULTS

Descriptive analysis

Descriptive statistics for the soil physical properties in 30 cm soil depth of the cultivated field are presented in Table 1.

Table 1. Descriptive statistics for the soil properties.

	Minimum	**Maximum**	**1st Quarter**	**Median**	**3rd Quarter**	**Mean**	**Standard Deviation**	**Coefficient of variation**	**Kurtosis**	**Skewness**
CEC ($cmol.Kg^{-1}$)	11.45	38.66	16.15	25.30	27.45	22.94	8.034	0.35	-0.77	0.19
EC (dS/m)	0.24	1.15	0.29	0.45	0.707	0.547	0.291	0.53	-0.52	0.83
Sand %	33.40	84.50	42.35	62.71	80.80	62.71	19.417	0.31	-1.93	-0.17
Silt %	10.90	36.10	13.90	21.70	23.05	19.84	6.975	0.35	-0.35	0.59
Clay %	3.30	35.80	5.35	6.90	30.70	17.46	13.227	0.76	-1.99	0.19

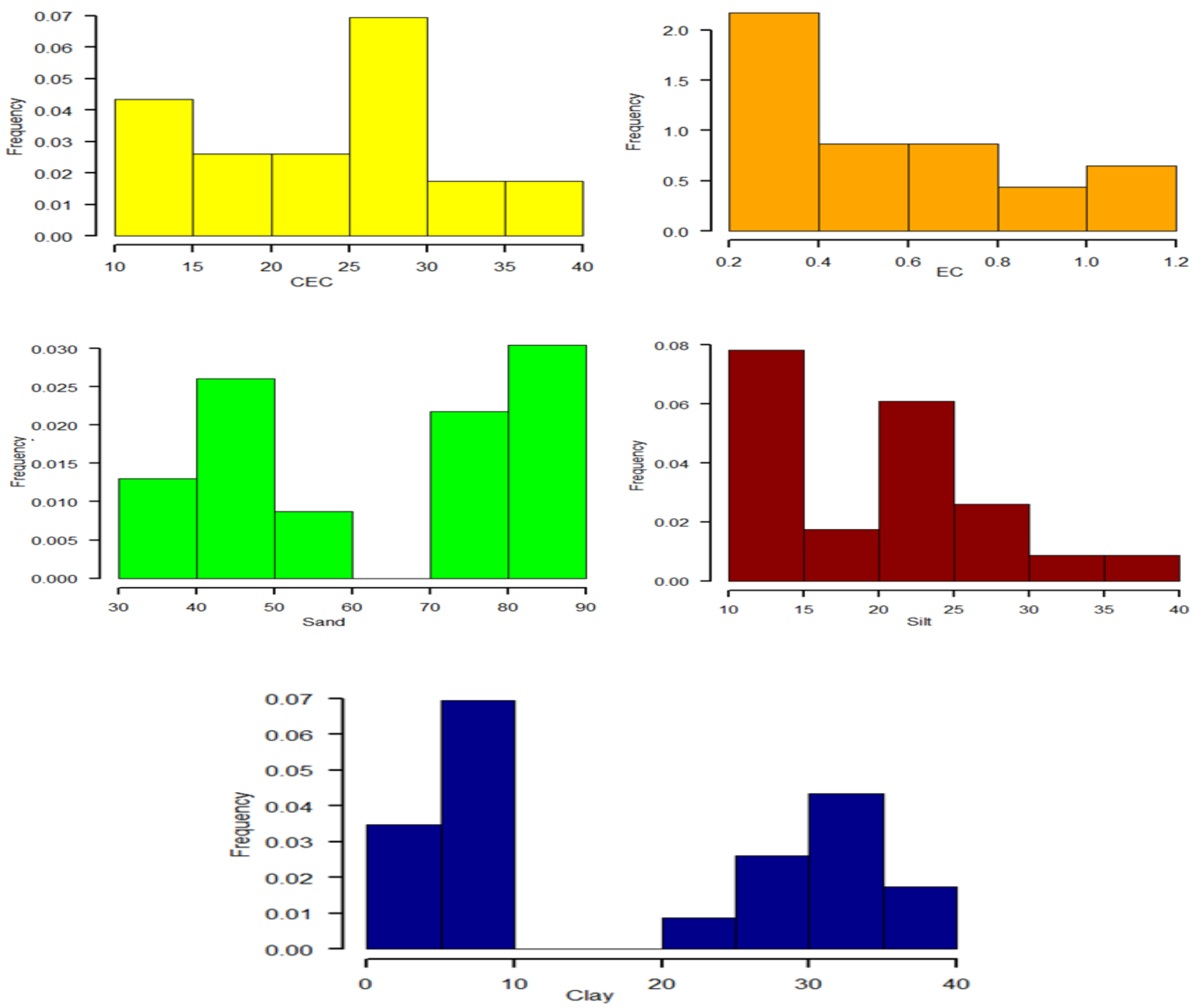

Figure 3. Frequency distribution of some soil properties

In 30 cm soil depth of the cultivated field, soil CEC varied between 11.45 and 38.66 ($cmol_c$ kg^{-1}), soil EC (0.24 to 1.15 dS/m), sand (33.40 to 84.50%), silt (10.90 to 36.10%) and clay (3.30 to 35.80%) values showed variations among the sampling points in the field (Table 1).As shown in Table 1, the C_0 /C_0+ C ratio values for CEC, EC, Sand, Silt and Clay were 31.93,0, 2.31, 41.27, and 1.41, respectively. The nugget/sill ratio was in between 25 and 75 % of CEC and Silt, respectively, which indicated a moderate spatial correlation.

Frequency distributions shown in Figure 3, in a graphical format, clarify the number of observations within a given interval for the selected soil properties. Frequency distributions are frequently used for the purpose of summarizing the categorical variables.

Geostatistical analysis

Table 2 clarifies the soil properties where variables characteristic was generated from semivariogram models. C_0 is represent the nugget variance; C is the variance structural, and C_0+C shows degree of the spatial variability.

Values varied from 5.43 for EC and 32.20 for Clay. The lowest value of Nugget (0.00) was observed in EC, however, silt recorded the largest nugget with a value of 19.08. Sill values varied among the selected soil properties in this study, where the largest value was 688.6 for sand, and the smallest one was 0.08 for soil EC.

As shown in Table 1, the C_0 /C_0+ C ratio values for CEC, EC, sand, silt and clay were 31.93; 0; 2.31; 41.27; and 1.41, respectively. The nugget/sill ratio was in between 25 and 75 % of CEC and silt, respectively, which indicated a moderate spatial correlation. The spatial dependence of the soil properties was moderate to strong relation.

The semivariogram of the selected soil properties for the soil surface are shown in Figure (4). It represents the experimental variogram that fitted to the studied soil properties. All soil characteristics fitted by using Gaussian model with exception for EC, as it fitted with Exponential model.

Table 2. Calculated semi-variograms properties of soil properties

Soil properties	Model	Range	Nugget (C_0)	Sill ($C_0 + C$)	*Nugget/sill ratio %	Spatial dependence
CEC	Gaussian	11.16	17.75	55.58	31.93	Moderate
EC	Exponential	5.43	0.00	0.08	0	Strong
Sand	Gaussian	21.40	15.91	688.6	2.31	Strong
Silt	Gaussian	16.79	19.08	46.23	41.27	Moderate
Clay	Gaussian	32.20	7.62	541.1	1.41	Strong

Note: Nugget/sill ratio (%) = $[C_0/(C_0 + C)] \times 100$ (Cambardella et al, 1994)

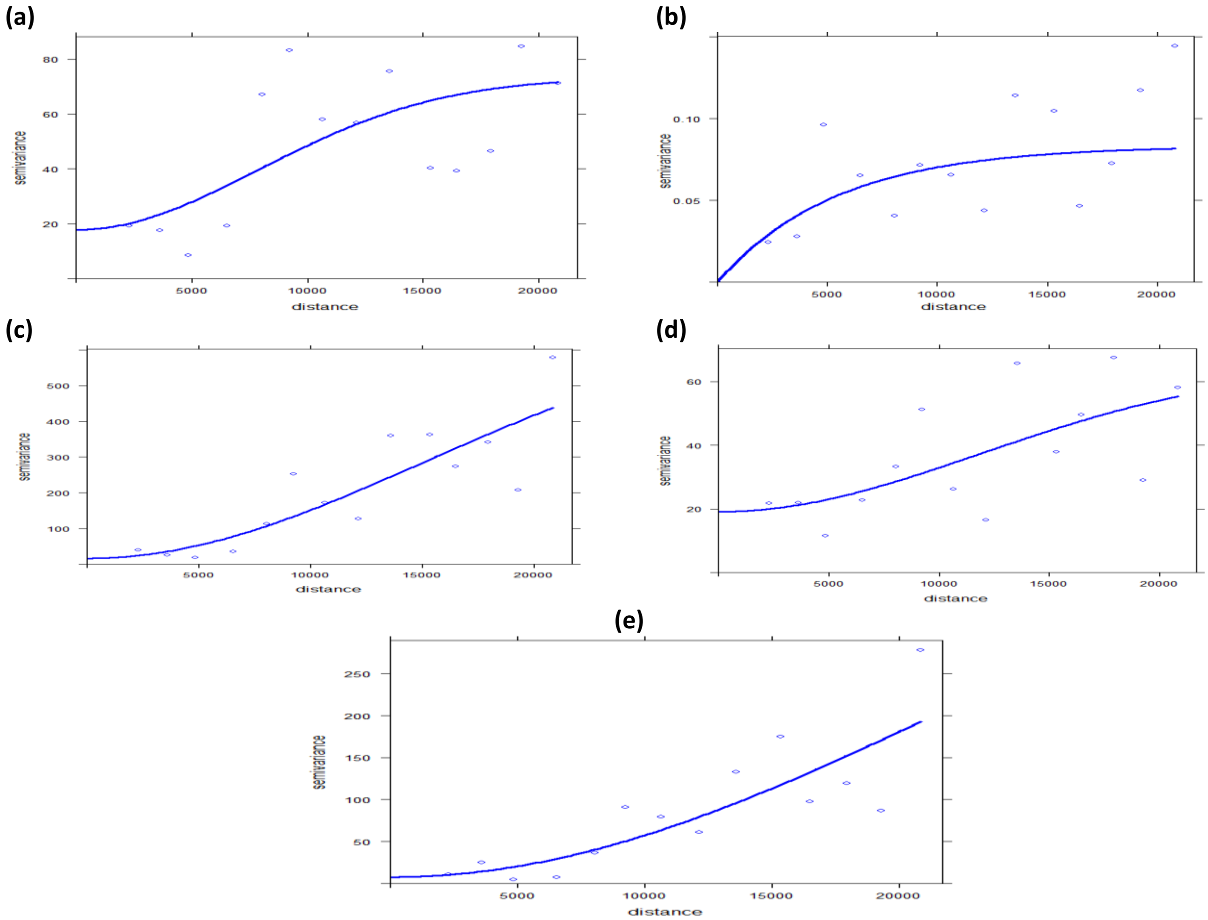

Figure 4. Semivariogram parameters of best-fitted theoretical model, (a) Soil CEC, (b) Soil EC, (c) Sand, (d) Silt and (e) Clay

Spatial distribution of soil properties

Figure 5a–e represents the spatial distribution of soil properties of CEC, EC, Sand, Silt and Clay using ordinary kriging (OK) as an interpolation method. The spatial correlation map of the soil properties CEC, EC, Sand, Silt and measured Clay was produced, compared and analysed for the results. Spatial variability maps among the soil CEC, EC, Sand, Silt and predicted Clay was prepared using R language to highlight the spatial dependence of the selected soil properties (Figure 5a-e).

Concentration of Soil CEC was observed in the eastern part of the study area. The spatial map of Soil EC is generated from the measured EC value from the collected samples in the study sites. In the central and eastern portion of the study site, higher EC was concentrated. The highest values of sand were in the central and western parts of the study area, while the highest values of silt was concentrated in the eastern parts. Clay was spatially distributed in the north-eastern part of the studied area. Future studies are needed in order to clarify the spatial variability on larger scale and to better understand the factors controlling the spatial distribution of soil properties.

Values of ME for CEC, EC and Sand were 0.08, 0.006 and 0.37, respectively. But for silt and clay, ME values were -0.16 and -0.06, respectively.

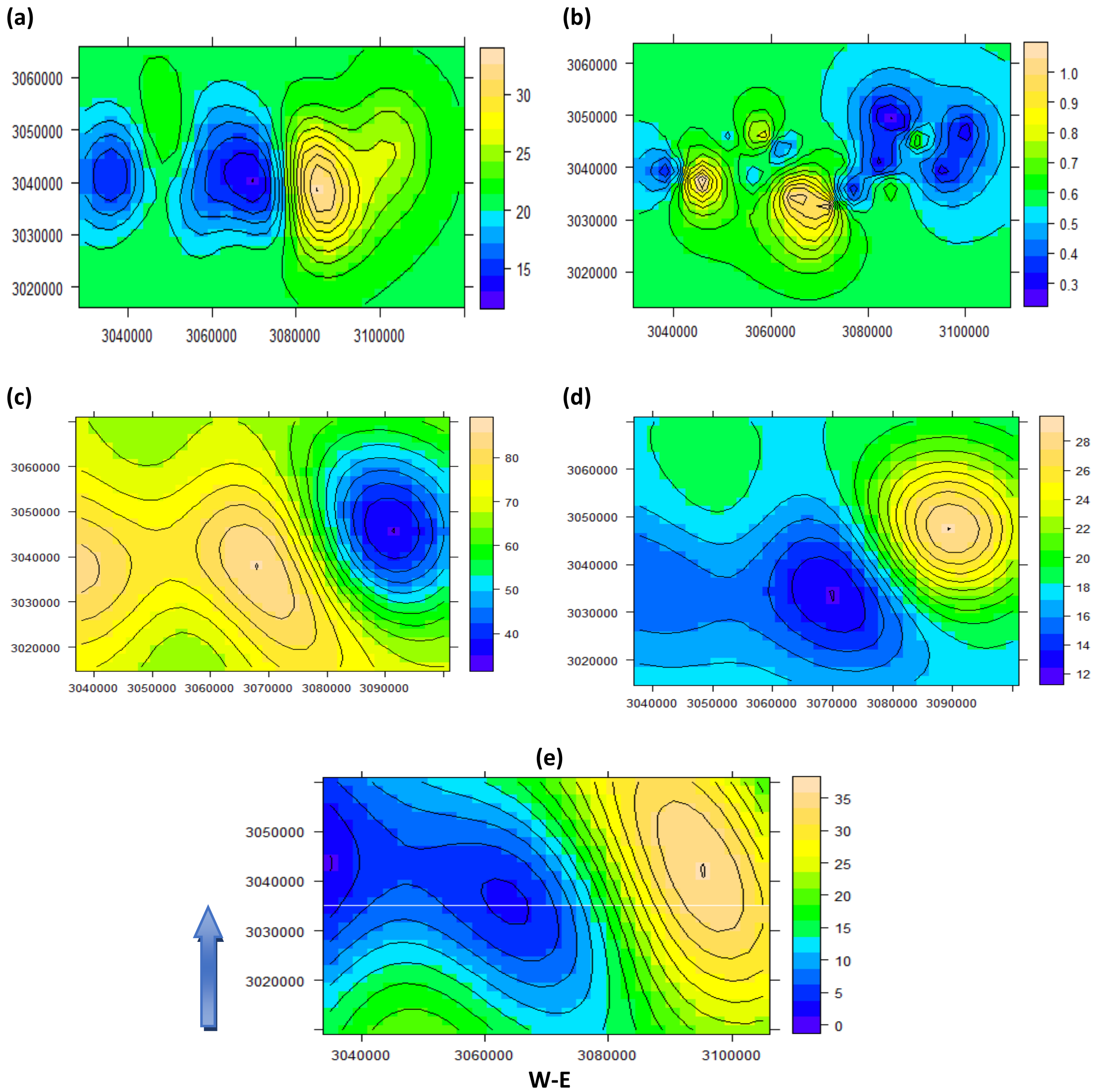

Figure 5. Spatial distribution of some soil properties (a) CEC, (b) EC, (c) Sand, (d) Silt and (e) Clay

DISCUSSION

In this study, spatial analysis was applied to selected soil samples from different regions in the Menoufia governorate, especially the centre and western parts. The descriptive analysis clarified that, all skewness values of the soil characteristics used in this study ranged from -0.17 to 0.83. Gia Pham *et al.* (2019), indicated that the distributions of all soil variables were only skewed when the skewness value is less than 1.0.

As shown from the results, C_0+C highlights degree of the spatial variability, influenced by both structural and stochastic factors. Golaszewski (2002), reported that depending on the range of effectiveness, the sill (C_0+C), and the nugget effect (C_0) for each parameter, the degree of autocorrelation was in relation to the spatial dependencies (Nugget/Sill ratio) among sampling points.

Tabi and Ognukunle (2007), mentioned that the range clarifies the distance in a field that determined properties are no longer spatially correlated. The shortest range (5.43 m) was observed for EC and the longest range (32.20 m) was observed for clay content. According to the results, the ranges of spatial influence for the soil physical properties were generally ≤ 32.20 m for clay, ≤ 21.40 m for sand, ≤ 16.79 m for Silt, ≤ 11.16 for CEC and ≤ 5.43 m for EC.

Any variable has a strong spatial dependency if the ratio of nugget/sill is given a value that equals to or less than 25%, a moderate ratio of the spatial dependency if the ratio is between 25 and 75%, and finally, a weak spatial dependency if the ratio recoded a value that is greater than 75% (Bo *et al.,* 2003). In regards to the previous study (Denton *el al.,* 2017), the classification of the spatial dependent variables that used, was pointed as a strongly spatial dependency if the ratio was <25, moderately spatial dependency if between 25 and 75%, and weak spatial dependency if it was >75%.

In the current study, the nugget/sill ratio was in between 25 and 75 % of CEC and Silt, respectively, which indicates a moderate spatial correlation. AbdelRahman *et al.* (2021) reported that variables with a moderate spatial dependence may be because of the soil homogeneity aspects. In this respect, range values are considered as essential measures for planning and experimental evaluations, the range can help in sampling procedure definition. In this sense, the range of the semivariogram was considered to estimate the minimum number of samples for characterizing the soil spatial variability pattern of soil chemical properties. AbdelRahman *et al.* (2021), added that low land environments were more heterogeneous in case of chemical properties.

The C_0 /C_0+ C ratio were less than 25 % in the three soil properties such as EC, Sand, and Clay which indicate a strong spatial correlation. Generally, a strong spatial dependency of soil properties is linked to the structural intrinsic factors like texture, parent material and mineralogy, but a weak spatial dependency is linked to the random extrinsic factors like ploughing, fertilization and may other related soil management practices (Zheng *et al.,* 2009).

As mentioned above, soil EC characterized by Exponential model was fitted without any nugget effect. These results were in agreement with another study that was applied in vineyards, the experimental semivariograms were best fitted to the theoretical models without any nugget effect (Mirás-Avalos *et al.,* 2020). AbdelRahman *et al.* (2021) suggested that the nugget effect is equal to the sill, when the variable under estimation is spatially independent.

Ordinary kriging method was focused on estimating values of soil properties at un-sampled locations taking into consideration the weighted local averaging method. An OK technique was applied for switching points of soil samples into continuous patterns of the selected soil properties. For an accurate prediction model, the absolute values of ME must be as small as possible. For the CEC, EC, sand, silt and clay, the absolute values of Mean Error (ME) resulted by Ordinary Kriging method are small and close to 0. Values of ME for CEC, EC and sand were 0.08, 0.006 and 0.37, respectively. But for silt and clay, ME values were -0.16 and -0.06, respectively. This observation is in conformity with Gia Pham *et al.*(2019), who reported that negative ME values indicate that the actual value recorded may be higher than the predicted value**.** These statistical values indicate that the prediction accuracy of

Ordinary Kriging estimate is high, and in agreement with those found by Gia Pham *et al. (*2019).

CONCLUSION

Understanding the spatial distribution and mapping accuracy of the soil properties for large scale areas are essential for precision farming, environmental monitoring, and modelling. This study used geostatistical models that were fitted using interpolation techniques in R language for five selected soil properties; cation exchange capacity (CEC), electrical conductivity (EC), Sand, Silt and Clay in different regions in Menoufia Governorate. The study confirms that this methodology can be used to investigate the spatial variability of soil properties. The results showed that the best fit semivariogram model of EC was Exponential model, where Gaussian model was the best fit model of Sand, Silt, Clay and CEC. According to the spatial distribution map, five zones of spatial variables were identified. The nugget/sill ratio was between 25 and 75 % for CEC and Silt, respectively, indicating a moderate spatial correlation, whereas a strong spatial correlation was found with a ratio of less than 25 % in the other soil properties. The results show that the spatial distribution and the spatial dependence level of soil properties can be different even within the same area. It also shows that the effectiveness of R environment in sense of data interpretation. Cross-validation of variogram models through OK representing in ME showed that the spatial prediction of the selected soil properties is high. However, future studies are needed to a better understand the spatial variability on a larger scale and to better understand the factors that control spatial variability of soil properties.

REFERENCES

Adepetu, J.A., H. Nabhan, and A. Osinubi. 2000. Simple soil, water and plant testing techniques for soil resource management. *International Institute of Tropical Agriculture.* Food and Agriculture Organization of The United Nations, Rome.

AbdelRahman, M.A.E., Y.M. Zakarya, M.M. Metwaly, and G. Koubouris. 2021. Deciphering Soil Spatial Variability through Geostatistics and Interpolation Techniques. *Sustainability*, 13, 194. https://doi.org/10.3390/su13010194

Behera, S.K., R.K. Mathur, A.K. Shukla, K. Suresh and C. Prakash 2018. Spatial variability of soil properties and delineation of soil management zones of oil palm plantations grown in a hot and humid tropical region of southern India. *Catena*, 16 (5): 251- 259.

Bo, S., Z. Shenglu and Z. Qiguo. 2003. Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. *Geoderma*, 115(1-2): 85-99.

Bhunia, G.S., P.K. Shit, and R. Maiti. 2016. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *Journal of the Saudi Society of Agricultural Sciences* 17(2), 114-126.

Cambardella, C.A., T.B. Moorman, J.M. Novak, T.B. Parkin, D.L. Karlen, R.F. Turco, and A.E. Konopka.1994. Field scale variability of soil properties in central Iowa soils. Soil Sci. Soc. Am. J., 58(5): 1501-1511.

Denton, O.A., V.O. Aduramigba-Modupe, A.O. Ojo, O.D. Adeoyolanu, K.S. Are, A.O. Adelana, A.O. Oyedele, A.O. Adetayo, and A.O. Oke. 2017. Assessment of spatial variability and mapping of soil properties for sustainable agricultural production using geographic information system techniques (GIS), Cogent Food & Agriculture, 3:1, 1279366.

Fabiyi, O.O., O. Ige-Olumide, and A.O. Fabiyi. 2013.Spatial analysis of soil fertility estimates and NDVI insouth-western Nigeria: A new paradigm for routine soil fertility mapping. Res. J. Agric. Environ. Manag. 2, 403–411.

Gia Pham, T., M. Kappas, C. V. Huynh and L.H.K. Nguyen, 2019. Application of Ordinary Kriging and Regression Kriging Method for Soil Properties Mapping in Hilly Region of Central Vietnam. *ISPRS Int. J. Geo-Inf.8*, 147. https://doi.org/10.3390/ijgi8030147

Golaszewski, J. 2002. Geostatistical approach to data from field experiments with check plots. *Electronic Journal of Polish Agricultural Universities*, 5 :(2).

Gundogdu, K.S. and I. Guney. 2007. Spatial analyses of groundwater levels using universal kriging. *J. Earth System Sci.*, 116: 49-55.

Johnston, K., J.M.V. Hoef, K. Krivoruchko, and N. Lucas 1996. Using ArcGIS geostatistical analysis. *GIS Manual by EAR*I, New York, (p. 120-187).

Mondal, A., D. Khare, S. Kundu, S. Mondal, S. Mukherjee, and A. Mukhopadhyay. 2017. Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data. *Egypt. J. Remote Sens. Space Sci.*, 20: 61–70.

Meul, M. and M. Van Meirvenne. 2003. Kriging soil texture under different types of nonstationarity. *Geoderma*, 112: 217-233.

Mirás-Avalos, J.M., Fandiño, M., Rey, B.J., Dafonte, J. and Cancela, J.J. 2020. Zoning of a newly-planted vineyard: Spatial variability of physico-chemical soil properties. *Soil Systems*, 4, 62.

Pebesma, E. 2004. Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, Vol. 30, 27:683 – 691.

Richards, L.A. 1954. Diagnosis and improvement of saline and alkali soils. LWW.

Wang, S., Q. Zhuang, Q. Wang, X. Jin, and C. Han, 2017. Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China. *Geoderma*, 305, 250–263.

Tabi, F.O. and A.O. Ogynkunle. 2007. Spatial variation of some soil physic-chemical properties of an alfisol in Southwestern Nigeria. *Nigerian Journal of Soil and Environmental Research.* 7:82 – 91.

Triantafilis, J., I. Odeh, and A. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. *Soil Sci. Soc. Am. J.*, 65: 869-878.

Van Reeuwijk, L. 2002. International soil reference and information centre. The Netherlands. Variable charges in selected temperate region soils. *Geoderma*, 32:89–101.

Zheng, H., J. Wu, and S. Zhang. 2009. Study on the spatial variability of farmland soil nutrient based on the kriging interpolation. *AICI, International conference on artificial intelligence and computational intelligence*, November 7-8, Vol.4 (p.550 – 555). Shanghai, China.