

# Geostatistical Analysis of Topsoil Clay and Hydraulic Conductivity Case Study from Al-Jabal Al-Akhdar Region

Ahmed Y. Habel<sup>1</sup>

## ABSTRACT

The present study aimed at applying cokriging to predict the values of the primary variable hydraulic conductivity, which is sparsely sampled and hard to measure, using the information content of topsoil Clay content, which is densely sampled and easy to measure, taking into consideration the fact that these two variables are correlated. Regression analysis of both clay percentage and hydraulic conductivity indicated a negatively highly correlated two variables, which satisfies the need to carry out cokriging analysis of hydraulic conductivity using the information content of clay %. The correlation coefficient for this analysis is -0.815. The semi-variogram for clay was fitted to the Gaussian model. While, the semivariogram for hydraulic conductivity was fitted to the Spherical model. The cross-semivariogram of clay and hydraulic conductivity is of the collocated type, which means that the estimation was performed using variables measured at the same location. The most important parameter in this estimation is the high  $R^2$  (0.76) obtained from the fitting process. This slightly high estimation regression coefficient comes helps in explain the spatial distribution of both variables. The advantage of cokriging over linear regression is that it takes into consideration the spatial variability of the surrounding points, rather than performing blindly the linear regression, which lacks this improvement.

**Key words:** Geostatistics, semivariogram, cokriging analysis, physical properties, spatial linear regression

## INTRODUCTION

Geostatistics has been applied to described the spatial variability using the semi-variogram and predict the values of soil attributes at un-sampled locations by different kriging techniques (Trangmar et al., 1985; Warrick et al., 1986; Habel, 1989; Webster and Oliver, 1989; Burrough, 1989; Webster, 1991; Goovaerts, 1992, 1998 and 1999; Bahnassy et at. 1995; Bahnassy and Morsy, 1996; Janik, 2005; Grego et al.,2006 and Janik, 2008), ecological properties (Banerjee and Gelfand, 2002), and categorical variables (Bogaert, 2002).

The term cokriging is used for spatial linear regression that uses data defined by different attributes. The data set will contain the primary variable of interest in addition to one or more secondary variables, which are spatially cross-correlated with the primary variable. Thus, the dataset will contain useful information about

the primary variable. The cross-correlation between variables utilized to improve these estimates, and to reduce the variance of the estimation error. The usefulness of the secondary variable is often enhanced by the fact that the primary variable of interest is under-sampled (Issacks and Srivastava, 1989). The spatial relationship between the values of the attribute is governed by the regionalized variable theory, which states that observations close to each other are more correlated than observations taken at a further distance (Journel and Huijbregts, 1978). This means that points spatially close to the estimation points should be given higher weights than those further away (Cressie, 1993). The coregionalized variable theory deals with the same situation as the regionalized variable theory, but the variables under consideration are correlated, and behave the same (McBratney and Webster, 1983 and 1986). Consequently, the cross-semivariogram can be modeled as a joint function between the two variables (Issacks and Srivastava, 1989). The linear coregionalization model allows for different ranges of spatial correlations for each variable (Wackernagel, 1994 and 1995).

Due to computation and notation difficulties related to cokriging system (Journel and Huijbregts, 1978; Myers, 1982; and Deutsch and Journel, 1998), a limited number of researches have been carried out utilizing cokriging as a best linear unbiased estimator (B.L.U.E.). Habel (1989) used geostatistical approach to describe spatial variability of bulk density and moisture content in some Poland soils. Danielsson et al, (1998) applied cokriging to estimate the total amounts and the spatial distribution for organic carbon, nitrogen and phosphorus in the Gulf of Riga surficial sediments, using loss on ignition as a covariable. Goovaerts (1998) used different methods of kriging and cokriging to model the spatial distribution of pH and electrical conductivity in two transects in forest and pasture soils. Rivoirard (2001) indicated that the cokriging could be collocated or multi-collocated depending on the configuration of data and the location at which the value estimated. Bahnassy (2002) applied collocated cokriging to study the spatial distribution of topsoil sodicity using the information content of soil salinity. Morsy (2004) estimated the SAR and EC contents in the surface and subsurface layers. Moustfa (2005) used cokriging to estimate topsoil free

---

<sup>1</sup>Soil & Water Department, Faculty of Agriculture –  
Omar Al Mukhtar University, El- Bieda - Libya  
Received January 22, 2013, Accepted March 28, 2013

iron oxides content in some Libyan soils. Janik (2005) used kriging method to explain the variability of moisture content in two localities in Poland-Pszczew (Wielkopolskie Province) and Paw<sup>3</sup>owice (Dolnoœl'skie Province), and in one in Slovakia - Tatranska Kotlina. Janik (2008) proposed a method for the estimation of variability of two selected parameters of soil – bulk density and infiltration rate - on the basis of soil moisture variability which is easy to determine.

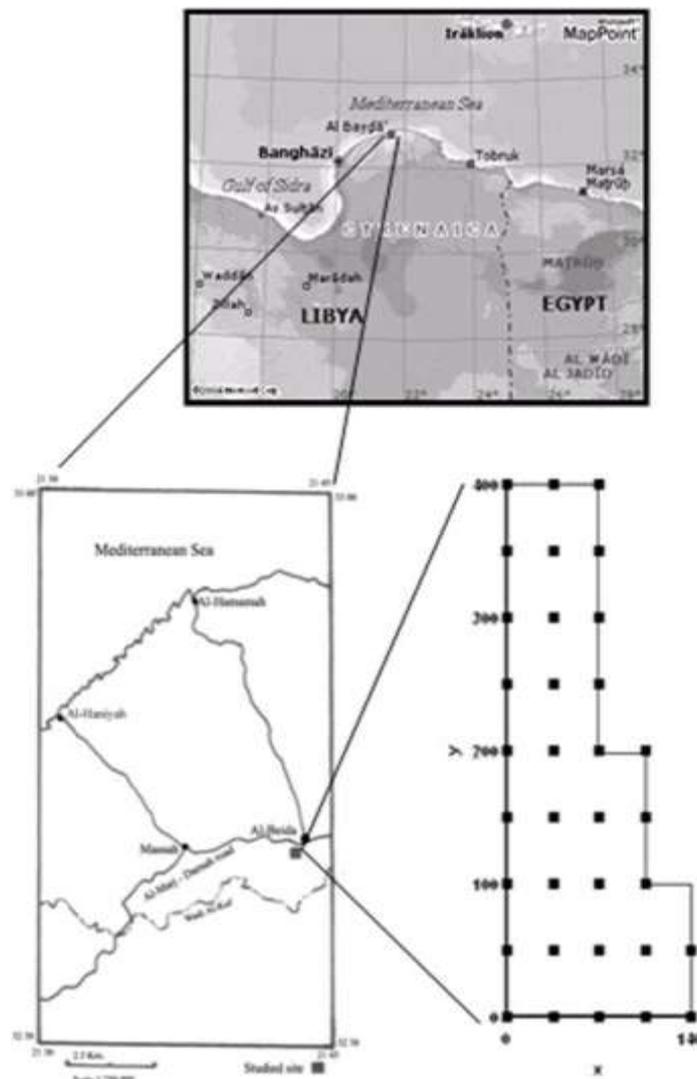
The current study aimed at applying cokriging to predict the values of the primary variable hydraulic conductivity, which is sparsely sampled and hard to measure, using the information content of topsoil Clay content, which is densely sampled and easy to measure, taking into consideration the fact that these two variable

are correlated. The cokriged hydraulic conductivity is compared to the kriged hydraulic conductivity and the standard error of estimation for both methods was matched.

## METHODOLOGY

### The Study Site

The studied site is located about 5 km. to the west of Al-Beida city and named as Ein El-Hiaa area (Fig. 1). The area characterized by the presence of short Wadies and some of them are branched from wadi al-Kuf. The studied site is part of one of these short wadies coarse. Soils in the studied site are mainly Terra Rossa that includes Typic Haploxeralfs and Typic Haploxerepts (Soil Survey Staff, 1998).



**Fig. 1. Location of study area and soil observations**

**Sampling Scheme and Soil Analysis**

Thirty-four soil observations collected over the study area using systematic soil survey design. The field is about 400 m long and 140 m width. The lag between observations was constant and was 100m in the SN direction and 35m in WE direction. The topsoil was analyzed for Clay content and hydraulic conductivity as described in Page et. al. (1982). The samples locations were georeferenced to the UTM coordinate system. The spatial configuration of the soil observations used for Clay and hydraulic conductivity showed in Fig.1.

**Descriptive Statistical Analysis**

The data for Clay content and Hydraulic conductivity were analyzed for basic statistics including mean, variance, standard deviation, minimum, maximum, skewness, and kurtosis. The histogram for both variables was obtained, and the correlation between the two variables was calculated.

**Semivariogram and Cross-semivariogram Analysis**

The semivariogram is defined as half of the average squared difference between two attribute values separated by vector h, for one variable (Burrough and McDonnell, 1998):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z(x_i) - Z(x_i + h)\}^2$$

where N(h) is the number of pairs at lag h, Z(xi) is the value of the attribute at location (xi) and Z(xi + h) is the value of the attribute at location (xi + h) separated by distance h. The separation vector h is specified with some direction and distance (lag) tolerance. This semivariogram is used to model both clay content and Hydraulic conductivity cm/h, and then fitting them to one of the known semivariogram functions (Gaussian, Exponential, and Spherical). In case of using two variables (cokriging) the cross-semivariogram is calculated as follows:

$$\gamma_{uv}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{Z_u(x_i) - Z_u(x_i + h)\} \{Z_v(x_i) - Z_v(x_i + h)\}$$

where Zu (Clay content) and Zv (Hydraulic conductivity cm/h) are the two variables. This equation is used to model Hydraulic conductivity cm/h using the information content of Clay, then fitting the obtained model to one of the known cross-semivariograms represented by Gaussian, Spherical, and Exponential functions.

**Cokriging**

Kriging is a means of spatial prediction that can be used for soil properties. It is a form of weight local averaging. It is optimal in the sense that it provides

estimate of values at unrecorded places without bias and with minimum and known variance. Kriging depends on first computing an accurate semi-variogram, which measures the nature of spatial dependence for the property. Estimates of semi-variance are then used to determine the weights applied to the data when computing the averages, and are presented in the kriging equations.

A co-kriged estimate is a weighted average in which the value of U at location xo is estimated as a linear weighted sum of co-variables Vk. If there are k variables k = 1, 2, 3, . . . V, and each variable is measured at nv places, xik = 1, 2, 3... Nk, then the value of one variable U at xo is predicted by (Burrough and McDonnell, 1998):

$$Z'_U(x_o) = \sum_{k=1}^V \sum_{i=1}^{n_v} \lambda_{ik} Z(x_{ik}) \quad \text{for all } V_k$$

where λik is the weight assigned to variable k and Z(xik) is the value of the variable at location i.

To avoid bias, i.e. to ensure that: E[zu(xo) - z'u(xo)]=0

And the sum of weights λik = 1 for U = V and the sum of weights λik = 0 for V ≠ U

The first condition (sum of weights λik = 1) implies that there must be at least one observation of U for cokriging to be possible. The interpolation weights are chosen to minimize the variance:

$$s^2_u(x_o) = E[\{zu(xo) - z'u(xo)\}^2]$$

There is one equation for each combination of sampling site and attribute, so for estimating the value of variable j at site xo, the equation for the g-th observation site of the k-th variable is:

$$\sum_{j=1}^V \sum_{i=1}^{n_v} \lambda_{ij} \gamma_{ij}(x_{ij}, x_{gk}) + \Phi_k = \gamma_{uv}(x_o, x_{gk})$$

for all g=1 to nv and all k=1 to V, where Φk is the Lagrange's multiplier. These equations together make-up the cokriging system.

**Cross Validation**

Cross validation is a technique, which is used to compare estimated and true values using the information available in the data set. In cross validation, the estimation method is tested at the locations of existing samples. The sample value at a particular location is temporarily discarded from the sample data set; the value at the same location is then estimated using the remaining samples. Once the estimate is calculated, it is compared to the true sample value that was initially

removed from the sample data set. This procedure is repeated for all samples. This could be expressed as (Issaks and Srivastava, 1989):

$$\text{Error} = r = v' - v$$

Where  $v'$  is the estimated value and  $v$  is the true value. Mean square error (MSE) is calculated from the formula:

$$MSE = \frac{1}{n} \sum_{i=1}^n r^2$$

## RESULTS AND DISCUSSIONS

### Descriptive Statistical Analysis

The statistical analysis of the clay content and Hydraulic conductivity is shown in table (1). It is clear that Hydraulic conductivity has less variability than clay content as the standard deviation and variance are small. This is attributed to the greater number of soil samples with low Hydraulic conductivity values (Fig.2), which lowered the mean compared to the standard deviation. The histogram for both Clay content and Hydraulic conductivity is shown in figures (2 and 3). The distribution of Hydraulic conductivity variables is skewed, indicating the dominance of low values, with the presence of a very little high values that might have an impact of the final estimates (Isskas and Srivastiava, 1989). On the other hand, variance indicates that Hydraulic conductivity has spread on a wide range contrary to Clay content, which is distributed around a high number of samples with low values (Fig.3). Regression analysis of both clay% and hydraulic conductivity indicated a negatively highly correlated two variables, which satisfies the need to carry out cokriging analysis of hydraulic conductivity using the information content of clay %. The correlation coefficient for this analysis is -0.815. Yates and Warrick (1987) showed that if the correlation coefficient between a primary variable and the covariable exceeds 0.5, then the inclusion of the covariable is favorable, and cokriging performs better than kriging.

### Clay content and Hydraulic conductivity Semivariograms

The semivariogram for clay was fitted to the Gaussian model as shown in the following equation:

$$\gamma(h) = C_0 + C_1 \{1 - \exp(-\frac{3h^2}{a^2})\}$$

While, the semivariogram for Hydraulic conductivity was fitted to the Spherical model as shown in the following equation:

$$\gamma(h) = C_0 + C_1 \{1 - \exp(-\frac{h}{a})\}$$

Where  $C_0$  is the nugget,  $C_1$  is the sill,  $h$  is the separation distance (lag) in meters, and  $a$  is the range.

The parameters for the fitted semivariograms for both clay and Hydraulic conductivity are shown in table (2), and the semivariograms are shown in figures 4 and 5. The formulated equations for these two variables are as follows:

$$\gamma_{Clay}(h) = 55.21 \{1 - \exp(-\frac{3h^2}{(274.9)^2})\}$$

$$\gamma_{Kh}(h) = 0.0021 \{1 - \exp(-\frac{h}{259.2})\}$$

### Cross validation and unsampled sites

The process of cross validation between the estimated and the true value permits the evaluation of cokriging performance. Figures 6 and 7 shows the linear regression between the cokriged and actual values of clay and hydraulic conductivity. The regression equation resulted from the cokriging cross validation is as follows:

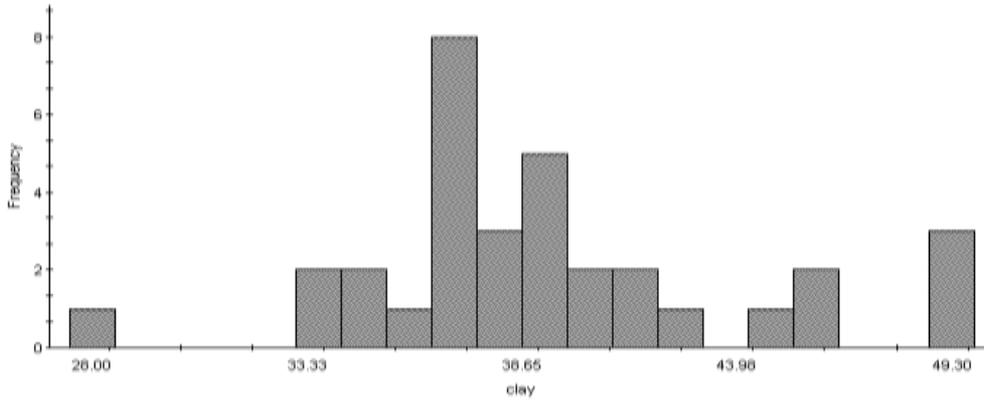
$$X (\text{estimated clay}) = 0.443 + 0.985 Y (\text{actual clay})$$

$$X (\text{estimated Kh}) = -0.032 + 1.140 Y (\text{actual Kh})$$

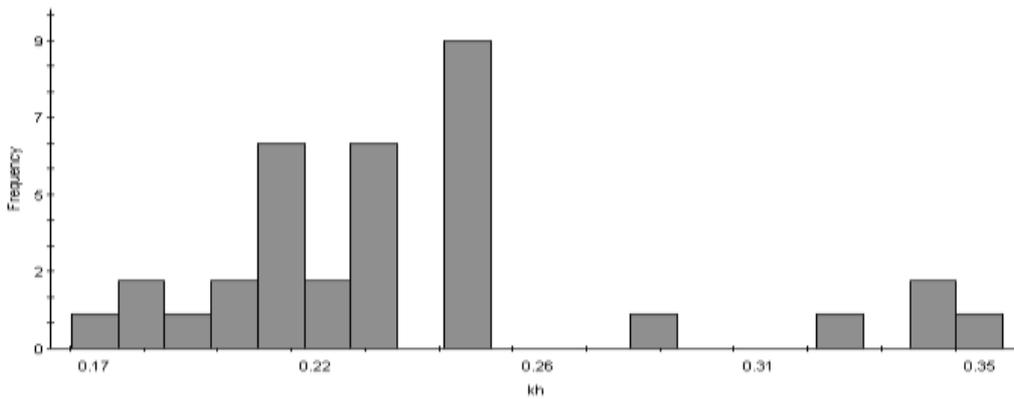
From the regression equations and spatial distribution the unsampled sites could be estimated and valued as shown in figures 8 and 9.

**Table 1. Descriptive statistical analysis for Clay content and Hydraulic conductivity**

Statistical Parameter	Clay content	Hydraulic conductivity cm/h
Mean	39.0412	0.2376
Standard Deviation	4.9649	0.0445
Variance	24.6504	0.0020
Minimum	28.00	0.170
Maximum	49.30	0.350
Skewness	0.42	1.10
Kurtosis	0.12	0.85
N (number of samples)	34	34



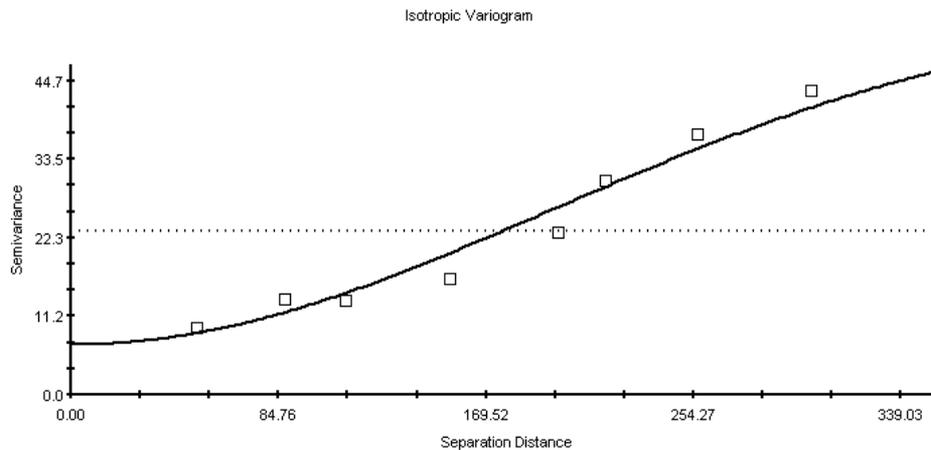
**Fig.2.**The histogram for clay content distribution



**Fig.3.** The histogram for Hydrulic conductivity distribution

**Table 2. Semivariograms types and parameters for clay and hydraulic conductivity**

Variable	Model	Nugget (C <sub>0</sub> )	Sill (C <sub>0</sub> +C <sub>1</sub> )	Range (a)	R <sup>2</sup>
Clay content	Gaussian	7.1	55.21	274.90	0.96
Hydraulic conductivity cm/h	Spherical	0.0006	0.0021	259.20	0.56



**Fig. 4.** The semivariogram for clay

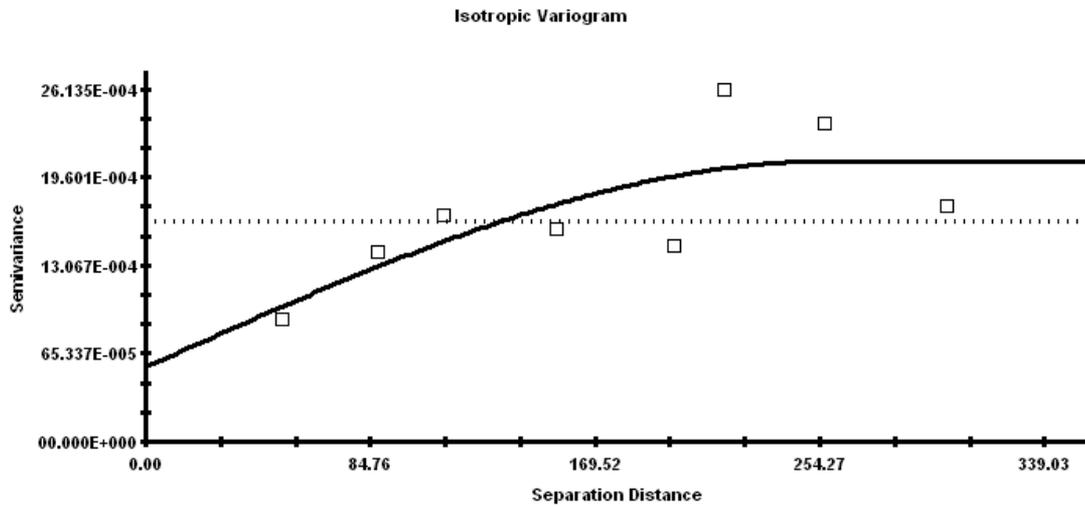


Fig. 5. The semivariogram for hydraulic conductivity

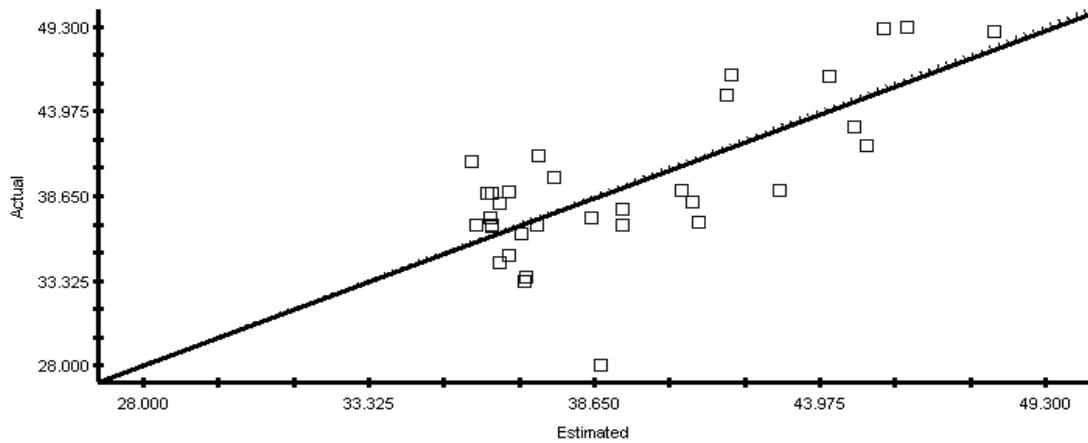


Fig. 6. The regression for clay content

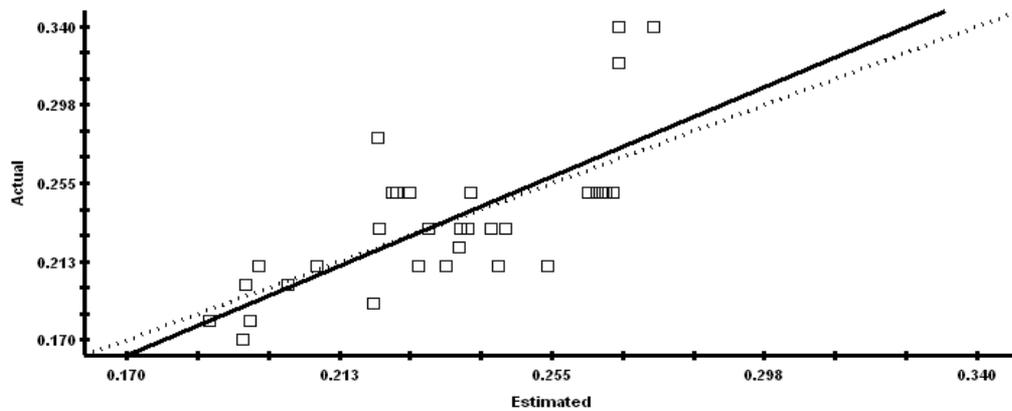


Fig.7. The regression for hydraulic conductivity

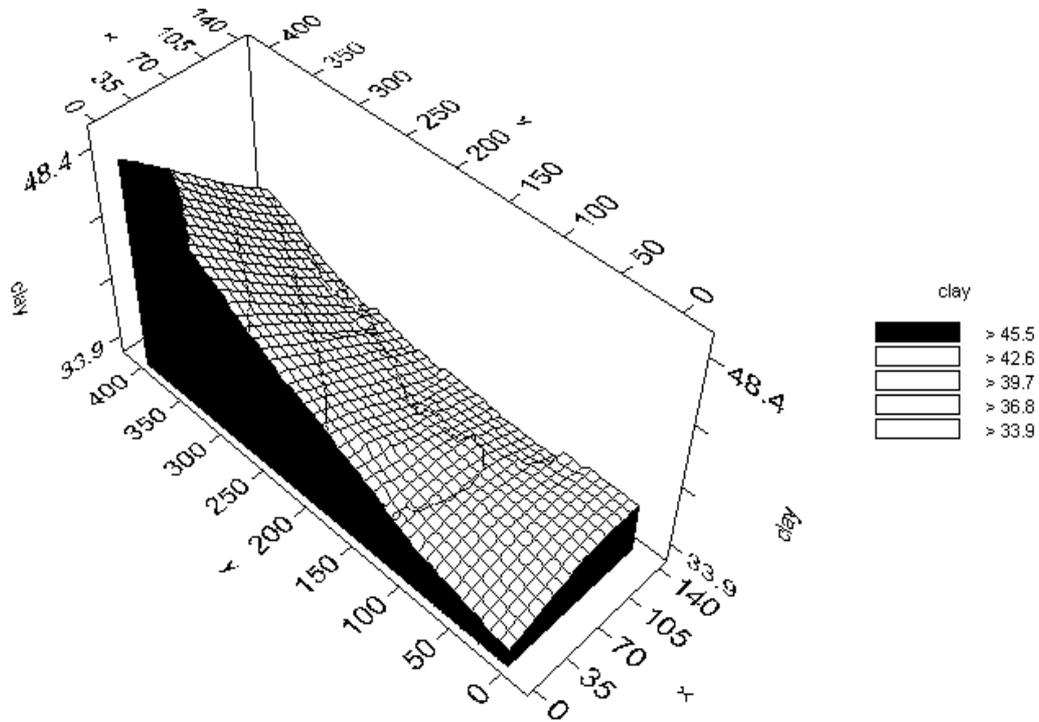


Fig. 8. Estimation of clay content in unsampled sites

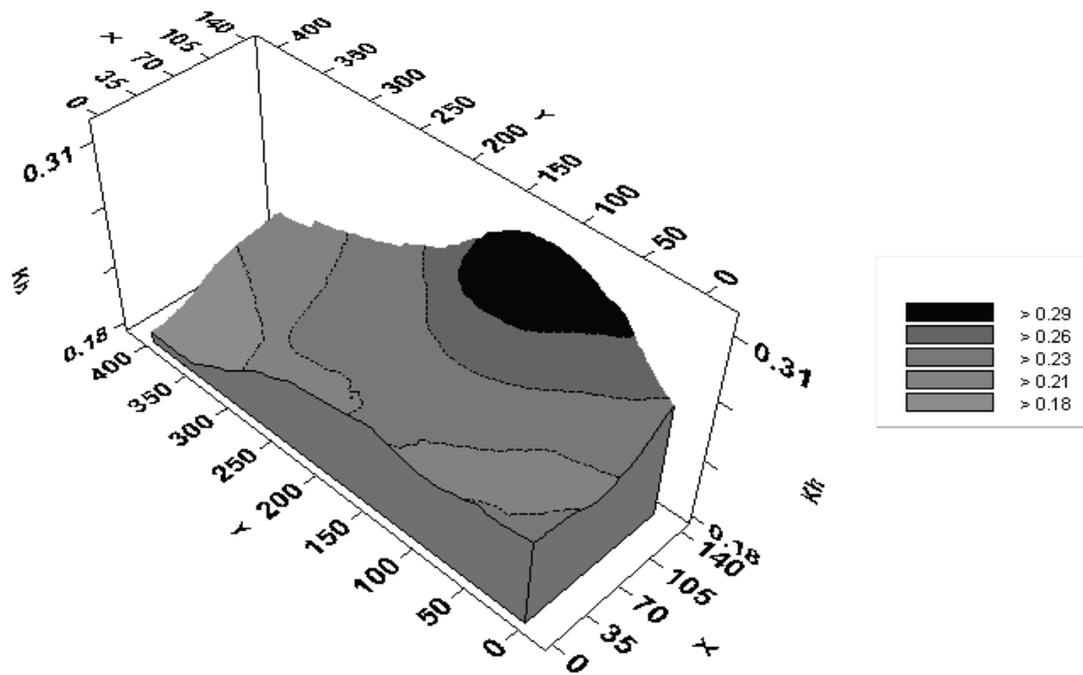
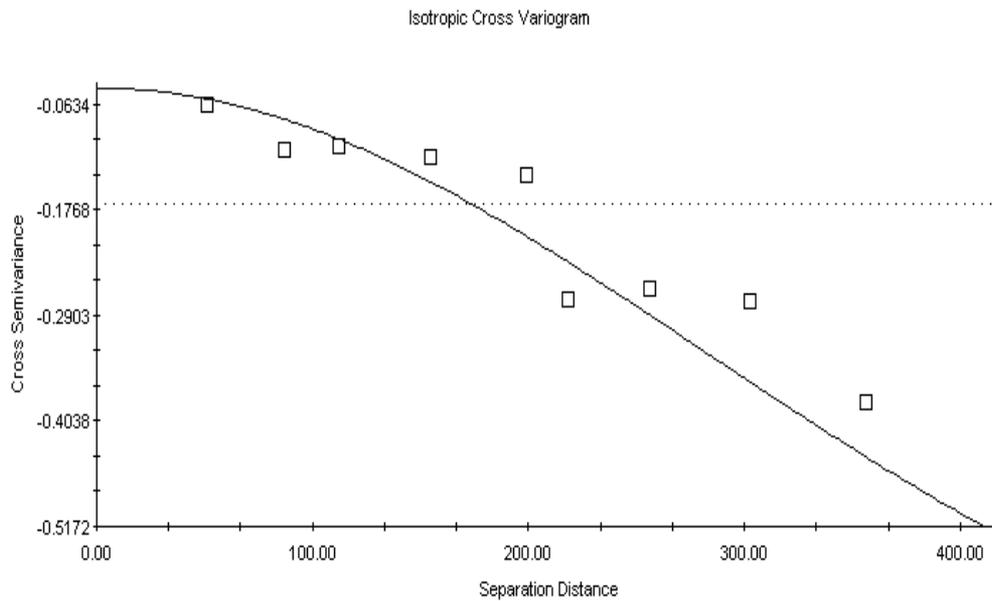


Fig. 9. Estimation of hydraulic conductivity in unsampled sites

**Table 3. Cross-semivariograms parameters between clay and hydraulic conductivity**

Variable	Model	Nugget (Co)	Sill (Co+C1)	Range (a)	R2
Clay and Hydraulic conductivity	Gaussian	- 0.046	- 0.762	396.7	0.762



**Fig.10. The cross-semivariogram between clay and hydraulic conductivity**

**The Cross-semivariogram (Collocated semivariogram)**

The cross-semivariogram of clay and hydraulic conductivity is of the collocated type, which means that the estimation was performed using variables measured at the same location. Table (3) and figure (10) indicate the parameters of the fitted Gaussian cross-semivariogram between clay and hydraulic conductivity. The Gaussian joint semi-semivariogram is as follows:

$$\gamma_{Clay - Kh}(h) = -0.762\{1 - \exp(-\frac{3h^2}{(396.7)^2})\}$$

**CONCLUSIONS**

Regression analysis of both clay% and hydraulic conductivity was indicating a negatively highly correlated two variables, which satisfies the need to carry out cokriging analysis of hydraulic conductivity using the information content of clay percentage. The semi-variogram for clay was fitted to the Gaussian model, while, the semi-variogram for hydraulic conductivity was fitted to the Spherical model. The high R<sup>2</sup> (0.76) obtained from the fitting process explain the spatial distribution of clay and hydraulic conductivity. There are linear regression between the cokriged and actual values of clay and hydraulic conductivity. The regression equation resulted from the cokriging cross validation is as follows:

$$X \text{ (estimated clay)} = 0.443 + 0.985 Y \text{ (actual clay)}$$

$$X \text{ (estimated Kh)} = -0.032 + 1.140 Y \text{ (actual Kh)}$$

**REFERENCES**

Bahnassy, M. 2002. Geostatistical analysis of topsoil sodicity using collocated cokriging. J. Agric. Sci. Mansoura Univ.,

Bahnassy, M., A.S. Suliman, and Post, D.F. 1995. Geostatistical analysis of topsoil and its reflectance properties of the Maricopa Agricultural Center, USA. Alex. J. Agric. Res., 40(2): 225-242.

Bahnassy, M., and I.M. Morsy. 1996. Geostatistical study of surface and subsurface variables of bangar El-Sokkar soils. J. Agric. Sci. Mansoura Univ., 21(7):2745-2758.

Banerjee, S. and A.E. Gelfand. 2002. Prediction, interpolation and regression for spatially misaligned data. The Indian Journal of Statistics, Volume 64, Series A, Part 2: 227-245.

Bogaert, P. 2002. Spatial prediction of categorical variables: the Bayesian maximum entropy approach. Stochastic Environmental Research and Risk Assessment, 16: 425-448

Burrough, P.A., 1989. Fuzzy mathematical methods for soil survey and land evaluation. J. Soil Sci. 40. 477-492.

Burrough, P.A., and R. McDonnell, 1998. Principles of geographic information systems. Oxford University Press, New York.

Cressie, N. 1993. Statistics for spatial data. John Wiley and Sons Inc. New York.

- Danielsson, A., R. Carman, L. Rahm and J. Aigars. 1998. Spatial Estimation of Nutrient Distributions in the Gulf of Riga Sediments using Cokriging. *Estuarine, Coastal and Shelf Science*, 46: 713–722
- Deutsch, C.V., and A.G. Journel. 1998. *CSLIB: Geostatistical Software Library and User's Guide: 2<sup>nd</sup> edition*. Oxford Univ. Press. New York.
- Gamma Design Inc. 2001. *GS+ Geostatistical software user manual*. Plainwell, Michigan, USA.
- Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biol Fertil Soils*, 27: 315–334.
- Goovaerts, P. 1992. Factorial kriging analysis: a useful tool for exploring tile structure of multivariate spatial soil information. *J. Soil Sci.* 43. 597-619.
- Goovaerts, P. 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89: 1-45
- Grego R.C., Vieira S.R., Antonio A.M., and Della Rosa S.K. 2006. Geostatistical analysis for soil moisture content under the no tillage cropping system. *Sci. Agric. (Piracicab Braz.)*, 63, 4, 341-350.
- Habel, A. 1989. Spatial variability of bulk density and moisture content in Hapludalfs and Haplaquolls of Koscian Lowland. M.Sc. Thesis. Fac. of Land Reclamation and Water Management. Agriculture Academy of Poznan, Poland.
- Isaaks, E.H., and R.M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York.
- Janik, G. 2005. Spatial variability of soil moisture in grassland. *Int. Agrophysics*, 19:37-45.
- Janik, G. 2008. Spatial variability of soil moisture as information on variability of selected physical properties of soil. *Int. Agrophysics*, 22:35-42.
- Journel, A.G., and Ch. J. Huijbregts. 1978. *Mining Geostatistics*. Academic Press, New York.
- McBratney, A.B. and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. *J. Soil Sci.*, 34: 137-162.
- McBratney, A.B., and R. Webster. 1986. Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. *J. Soil Sci.*, 37: 617-639.
- Morsy I. M. 2004. Geostatistical study of the spatial distribution for topsoil sodicity and subsoil salinity using cokriging, *Egypt. Alex. Sci. Exch.*, Vol. 25 No.1, pp: 167-185.
- Moustafa, A. M. 2005. Geostatistical analysis of topsoil free iron oxides content using cokriging. *J. Agric. Sci. Mansoura Univ.*, 30(4):2275-2291.
- Myers, D.E. 1982. Matrix formulation of co-kriging. *Mathematical Geology*, 14: 249-257.
- Page, A.L., Miller, R.H. and Keeny, R. 1982. *Methods of Soil Analysis*. Agronomy Monograph no. 9. ASA. Madison, Wisconsin, USA.
- Rivoirard, J. 2001. Which Models for Collocated Cokriging?. *Mathematical Geology*, 33: 117-131.
- Soil Survey Staff. 1998. *Soil Taxonomy: A Basic System of Soil Classification for Making and Interpreting Soil Surveys.*– USDA Handbook No. 436, U.S. Government Printing Office, Washington, D.C.
- Trangmar, B.B., R.S. Yost, and G. Uehara. 1985. Application of geostatistics to spatial studies of soil properties. *Advances in Agronomy*, 38: 45-94
- Wackernagel, H. 1994. Cokriging versus kriging in regionalized multivariate data analysis. *Geoderma*, 62: 83-9-2.
- Wackernagel, H. 1995. *Multivariate Geostatistics: An introduction with application*. Springer, New York.
- Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science. In: *Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods*. Agronomy Monograph no. 9, 2<sup>nd</sup> edition., pp. 53-82.
- Webster, R. and M.A. Oliver. 1989. Optimal interpolation and isarithmic mapping of soil properties: VI. Disjunctive kriging and mapping the conditional probability. *J. Soil Sci.*, 40: 497-512.
- Webster, R., 1991. Local disjunctive kriging of soil properties with change of support. *J. Soil Sci.* 42, 301-318.
- Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. *Soil Sci. Soc. Am. J.*, 51: 23-30.

## الملخص العربي

# التحليل الجيو-إحصائي لتوزيع الطين والتوصيل الهيدروليكي للتربة السطحية دراسة حالة لبعض

## أراضي الجبل الأخضر - ليبيا

أحمد يوسف هبيل

يكون من النوع جاوس Gaussian. في حين، ال semivariogram للتوصيل الهيدروليكي من النوع ال Spherical.

أما ال semivariogram لكل من الطين والتوصيل الهيدروليكي معاً، يشير إلى إمكانية إجراء تقدير باستخدام المتغيرات المقاسة في نفس الموقع. النتيجة الأكثر أهمية في هذا التقدير هو قيمة  $R^2$  العالية (0.76) والتي تم الحصول عليها بين المتغيرين. هذا الانحدار العالي يساعد في تفسير التوزيع المكاني لكل المتغيرات. إن أهم ميزة لل cokriging عن الانحدار الخطي هو أنه يأخذ بعين الاعتبار التغير المكاني من النقاط المحيطة بها، وهذا ما لا يأخذه الانحدار الخطي في الاعتبار.

تهدف هذه الدراسة إلى تطبيق طريقة cokriging للتنبؤ بقيم التوصيل الهيدروليكي والذي يحتاج إلى جمع عينات خاصة وصعوبة في القياس، وذلك باستخدام محتوى التربة السطحية من الطين، والذي يمكن تقديره بطريقة سهلة وبسيطة، مع الأخذ في الاعتبار حقيقة أن هذين المتغيرين بينهما درجة عالية من الارتباط. وأشار تحليل الانحدار بين نسبة الطين والتوصيل الهيدروليكي إلى وجود علاقة عكسية قوية، والتي تلي الحاجة إلى إجراء تحليل cokriging للتوصيل الهيدروليكي باستخدام محتوى التربة من نسبة الطين. معامل الارتباط لهذا التحليل هو (-0.815) أوضحت النتائج أن نموذج ال variogram للطين