*Turkish Journal of Computer and Mathematics Education* *Vol.12 No.10(2021), 6343- 6355*

*Research Article*

# Effect of Soil Data Spatial Distribution for Multivariate Spatial Prediction under Uncertainty

**Ghanim Mahmood Dhaher**

Email:ghanim-hassod@uomosul.edu.iq

**Abstract:** This paper is intended to explain the multivariate spatial prediction in geostatistical analysis. The purpose of this paper to develop an empirical methodology for spatial insurance by studying the fundamental causes for differences in between distributed locations. The objective of this work is to access an approach of optimal model of prediction under uncertainty. Cokriging technique used in this paper through empirical estimation of variograms and cross-variograms in all directions of compass. Multivariate technique applied to obtain fitting theoretical models of covariance function, with their properties. The data adopted from Mosul quadrangle/Iraq, include two sets of groundwater viability of soil data., each set contains (100) sample. First set primary variable, magnesium (Mg), and second set is secondary variable (co-variable) is chlorine (Cl). The results of this work shown to suggest fit models with best prediction, through the small variations and constraints of weights clear support of accuracy of cokriging prediction. The outcome showed the models of data adopted nearest of origin models of covariance functions. In conclusion, the prediction of cokriging found that the origin data are the nearest the prediction of cross variogram. The computations are carried out by Matlab language .

**Keywords**: cokriging, uncertainty, cross-vriogram function , soil data.

The main characteristic in mining field and spatial statistics is study of phenomenon based on a set of observationsspatial prediction defined as to get the model of regionalized variables and to study the phenomenon includes the required predicate in unmeasured common properties. The problem statement of this study is to improve interpolation techniques performance, which are seeking to obtain a minimum error in prediction. in 1963 Matheron developed kriging's theory basing the study on krige's Master. The name "kriging" is method of interpolation prediction. In general cases, estimate parameters of covariance functions or models, this problem started with the French engineer Matheron in 1971. Other studies took universal kriging in estimation process as Dalezios et. al (2002), Brus et. al (2007).

Kriging's technique includes many subjects and multi spaces in addition to using a spatial model to determine weights during estimation, one of the most important aspects of the geographic statistical method is the integration of quantitative data. From classical multivariate statistics, we notice that models developed from two or more variables often produce better estimates. We can extend the classic multivariate techniques in the geostatistics world and use two or more regional variables in the geostatistics estimation. These techniques have a benefit over kriging techniques because they contain multivariate under uncertainty by Delobbe and Holleman (2006), Trapero et. al (2009),Chiles and Delfiner (2012), Erdin et. al (2012).For example, by analyzing our exploratory data, we may find a good correlation between a characteristic measured at the sites of the studied phenomenon and a characteristic of a specific chemical element. In such a case, we may wish to use the chemical information to provide the best estimates obtained from the phenomenon data alone. Even when the primary data count is (good), it is possible to use a sample-dense secondary characteristic such as soil salinity in the interpolation process such as Dobesch et. al (2007), Xie et. al (2011). Examples of cokriging two correlated variables sample at different locations in various applications such as Eldeiry and Garcia (2010),Wackernagel (2010), Wackernagel (2013).

## 2. Methodology

### 2.1 The variogram function

Inaccuracy in results and interpretation back to a lack of accuracy among the observations during using the correlation coefficient . Therefore, south African scientist Krige found the experimental of varioram function, usually used to build the best mathematical model.Let $y(s)$and $y(s+h)$be two random variables at two points $(s)$ and $(s+h)$separated by the vector $h$The variability between these two quantities is characterized by

$$2\gamma(s,h) = E\{[y(s) - y(s+h)]^2\}$$

In all generality, the variogram function $2\gamma(s,h)$ is a function of both the point$(s)$and the vector$(h)$. And variogram function is defined as:

$$2\gamma(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} [y(s_i + h) - y(s_i)]^2 \qquad (1)$$

Where $y(s_i)$ and $y(s_i + h)$are random variables at locations, $(s_i)$ and $(s_i + h)$ respectively where $i = 1,2,\dots,n$ and $(h)$ is the distance between the observations of data. The total number of experimental pairs $y(s_i), y(s_i + h)$ of data with distance$(h)$ is denote $n(h)$. To describe the spatial correlation of the variogram function and the cross- variogram is defined as:

$$2\gamma_{ij}(h) = \frac{1}{n(h)} \sum_{i=1}^{n(h)} [y_i(s_k + h) - y_i(s_k)][y_j(s_k + h) - y_j(s_k)] \qquad (2)$$

When( $i = j$) then variogram function equation (1) is reached, when ( $i \neq j$), then it is the cross-variogram. Random function is called second order stationary if and only if

$$E(y(s) = \boldsymbol{\mu} , \boldsymbol{cov}[y(s+h), y(s)] = c(h) \qquad (3)$$

Can be obtained on the relationship between covariance function $c(h)$, variogram function $2\gamma(h)$ and variance $c(0)$as in the following formula:$\gamma(h) = c(0) - c(h)$

## 2.2 Zone of Uncertainty

To access to zones of uncertainty, we must determine real spatial data locations with a certain probability provides in the field of study. kriging technique is the most important advantages using, it has the ability to identify areas of uncertainty. also, zones of uncertainty needs some information of the same data benefit the researcher. Let $y(s_o)$ be a variable with sit $(s_o)$, and $y(s_i), i = 1,2, \dots, n$, are observations of locations $(s_i)$ . Uncertainty round of $y(s_o)$ in unknown location $(s_o)$ is the model of the function of cumulative distribution of random variable$(s)$:

$$F(s, y) = Prob\{Y(S) \le y(s)\}$$

Where $F(.)$ is the function of the probability with unknown location and the probability between 0 and 1 give us the map of zones with and without zoned of uncertainty.

## 2.3 The Assumptions

Let $(m)$variables and variable $Y_\beta(s)$ the vector , represent number of $(n_\beta)$, $(n_\beta > 0)$

$y = (y_1(s), y_2(s), \dots, y_m(s))'$ , we can represent these the$\beta = 1,2, \dots, m$ as

$$y_\beta(s) = (y_\beta(s_1), y_\beta(s_2), \dots, y_\beta(s_{n\beta}))'$$

Where $y_1(s)$ are called primary variable and $y_\beta(s)$called secondary variables, then the vector of the random functions is:

$$y = (y_1(s_1), \dots, (y_1(s_{n1}), (y_2(s_1), \dots, (y_2(s_{n2}), \quad y_m(s_1), \dots, \quad y_m(s_{nm}))'$$

The weights vector that correspond to the random vector functions is

$$\omega = (\omega_1, \omega_2, \dots, \omega_m)' , \omega_\beta = (\omega_{\beta 1}, \omega_{\beta 2}, \dots, \omega_{\beta m})' \text{ then,}$$

$$\omega_\beta = (\omega_{11}, \omega_{12}, \dots, \omega_{1n_1}, \omega_{21}, \omega_{22}, \dots, \omega_{2n_2}, \omega_{m1}, \omega_{m2}, \dots, \omega_{mn_m})'$$

## 2.4 Multivariate geostatistics

Integrating secondary variable in spatial prediction often effective way to introduce secondary data in kriging algorithms is using secondary information as a trend.
In kriging, the primary variable is decomposed into a mean component

$$y_i(s) = m(s) + R(s) \quad (4)$$

the mean-component can be used to model the trend because secondary data is often smooth in nature,

$$E(y_i(s)) = m(s) = \varphi(y_2(s)) \quad (5)$$

from necessary of any function needs to be knowledge of data. this require that the secondary data is needs to be determined and available everywhere..

In the original soft data, spatial variance is largely neglected in kriging system. The property of spatial correlation between the secondary variables and the primary variable is characterized by continuity. Whereas, for the secondary data expressed in the correlation diagram, the spatial relationship between the secondary and the primary, therefore, we use the cross- covariance to obtain the complete spatial correlation.

$$\rho_{y_1 y_2}(h) = \frac{cov(y_1(s), y_2(s+h)}{\sqrt{var(y_1)var(y_2)}} \quad (6)$$

for a regionalization characterized by a set of *K* spatially intercorrelated random variables : { $y_k(s)$, k=1 to K}

The first and second order moments of these variables, assuming stationarity, are:

$$E[y_k(s)] = m_k, \quad \forall s \quad (7)$$

$$cov_{k\grave{k}}(h)] = E[y_k(s_{s1}) - m_k][y_{\grave{k}}(s_{s2}) - m_{\grave{k}}], \quad for\ s_{s1} - s_{s2} = h \quad (8)$$

Simple cokriging technique is similar to a simple kriging technique when the expectation is constant and known in all locations of the study. Also ordinary cokriging is a similar to ordinary kriging technique when the expectation is constant and unknown in the area of study, whereas the differences are in the numbers of variables, where there is primary variable and to predict the secondary variables so as to improve the value of prediction. Cokriging uses the multivariable, auxiliary variable and co- variable or secondary variable, with two constraints of the weights.

$$\sum_{i=1}^{n} \omega_i = 1, \quad \sum_j^m \sum_{i=1}^{n} \alpha_{ij} = 0 \quad (9)$$

$$\hat{y}(s_o) = \sum_{i=1}^{n} \omega_i y(s_i) \sum_{j=1}^{m} \sum_{i=1}^{n} \alpha_{ij} c(s_{ij}) \quad (10)$$

Where $(\omega_i)$ kriging weights, $(\alpha_{ij})$cokriging weights, and $c(s_{ij})$ are correlated to each other. The linear combination of cokriging predictor is defined as:

$$(11)\ \hat{y}(s_o) = \omega' y$$

Where $\hat{y}(s_o)$ represent real value of prediction. And by using the matrices form we get:

$$\hat{y}_{cok} = C_o^{'} C^{-1} y - C_o^{'} C^{-1} I\ (IC^{-1} I)^{-1} I' C^{-1} y + f'(IC^{-1} I) I' C^{-1} y \quad (12)$$

The variance of cokriging refers to the accuracy of prediction, and with the condition of unpaisedness that is equivalent$(\omega' I = f')$, then

$$\sigma_{cok}^2 = C_{oo} - \left(\omega' C_o + L_1\right)$$

$$\sigma_{cok}^2 = C_{oo} - \left(C_o^{'} C^{-1} C_o - (C_o^{'} C^{-1} I - f\right) (IC^{-1} I)^{-1} I' C^{-1} C_o - ((f'(IC^{-1} I)^{-1})(I' C^{-1} C_o - f) \quad (13)$$

## 2.5 Cross validation

- In order to obtain effective a prediction we used The value of g measures by using sample mean. where g- value is defined as:

$$g - value = \left[1 - \frac{\sum_{i=1}^{n} [y(s_i) - \hat{y}(s_i)]^2}{\sum_{i=1}^{n} [y(s_i) - \overline{y}(s_i)]}\right] \quad (14)$$

where $y(s_i)$are the observations of the variables,$\hat{y}(s_i)$ the predictor values and are $\overline{y}$ thesamplemean, to evaluate the complete prediction must gvalue is equal to 1 , while the prediction is less accurate when g value is a negative value, while if g value is a positive value that means a more positive prediction and g value of zero refer to the sample mean should be used.

- By using the cokriging variance we can defined the accuracy of prediction (MSEcok)and calculated by:

$$MSE_{cok} = \left[\sum_{i=1}^{n} \frac{[y(s_i) - \hat{y}(s_i)]^2}{\sigma_{cok}^2}\right] \quad (15)$$

## 3. Results and Analysis

### 3.1 Data study

On the practical side, the data used in this research were from the soil data. The data consisted of two groups. The first group of the primary variable is magnesium (Mg), content (100) samples of real values with their locations, and the second group of the co-variable or secondary variable, it contains (100) samples of chlorine (Cl). These data are a real spatial data from locations in Mosul/Iraq.(Hatem (2007))

Table (1): data statistic for Mg and Cl

| Stat. / Data | Min | Max | Median | Mode | Std |
|---|---|---|---|---|---|
| Mg | 0.2000 | 36.5000 | 6.75000 | 3.70000 | 6.1525 |
| Cl | 1 | 41 | 9.4000 | 7.8000 | 10.9614 |

Table (1) content the data statistic for two sets of data (Mg and Cl) fromMosul quadrangle. First set contains (100) sample of magnesium (Mg), while the second set contains (100) sample of chlorine (Cl) variables.

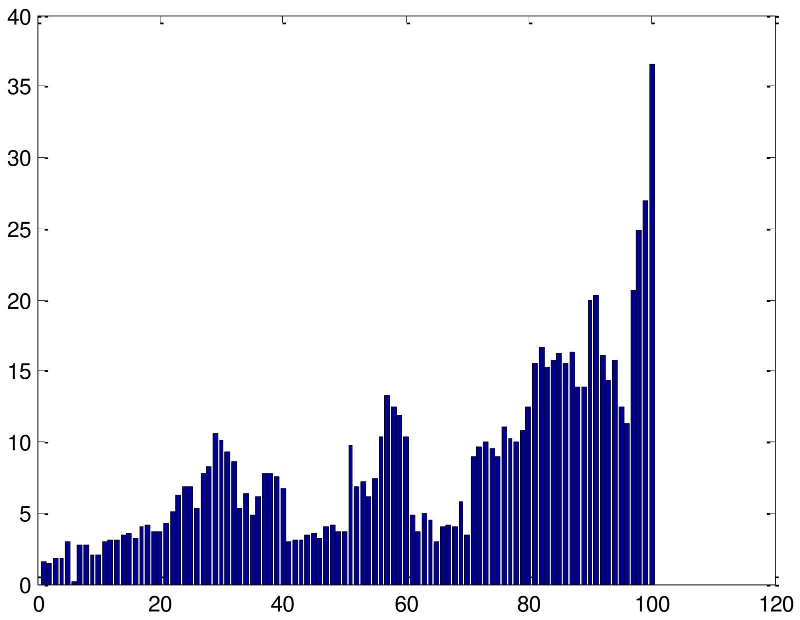


Figure 1: Histogram of Mg data

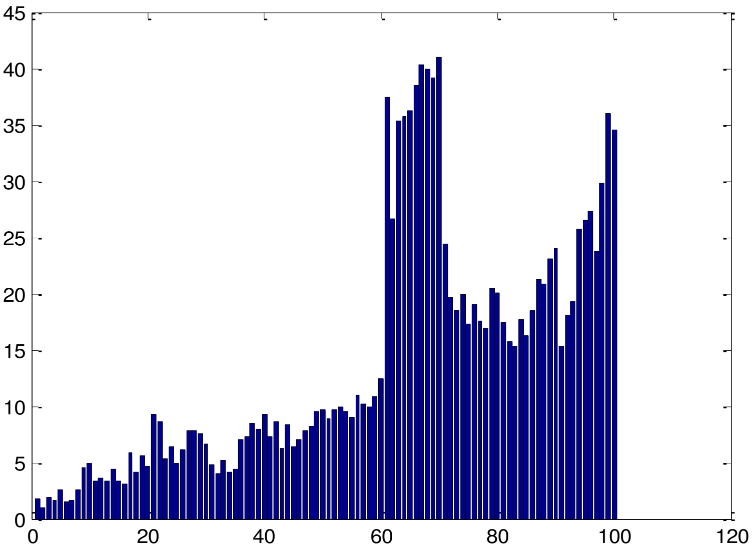


Figure 2: Histogram of Cl data

## 3.2 Variogram function

By using isotropic property, we applied the experimental variogram function according equation (1), to plot the curves of variogram function and by using the data of the first group for Mg, in all directions ($\theta = 0°, 90°, 45°, and\ 135°$), as shown inFigure (3)

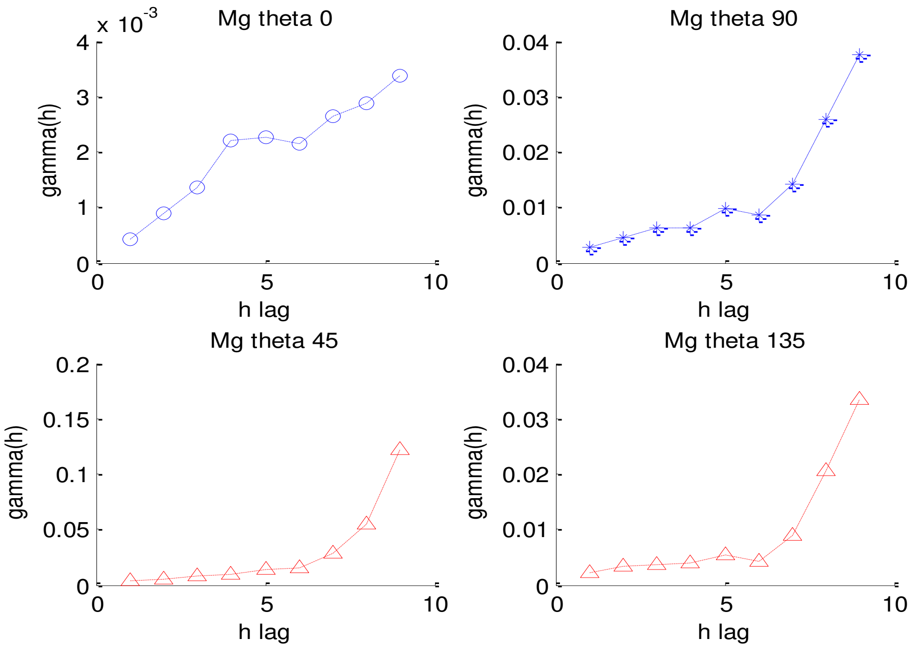


Figure 3: variogram functions of Mg data in all directions

Table (2) below shows the average of variogram function, because the two directions ($\theta = 0°, 90°$), haves the same lag $h$, $h = 1,2, \dots ,9,$, where we found the average of variogram, while the variogram function was found by two directions ($\theta = 45°, and\ 135°$), , where $h = 1.414, 2.828, \dots ,12.72$. Also, these two angles having the same lag ($h$)were found to be the average of variogram. These results of variogram give knowledge to define the properties of variogram function.

Table (2): Results of average variogram function of Mg data

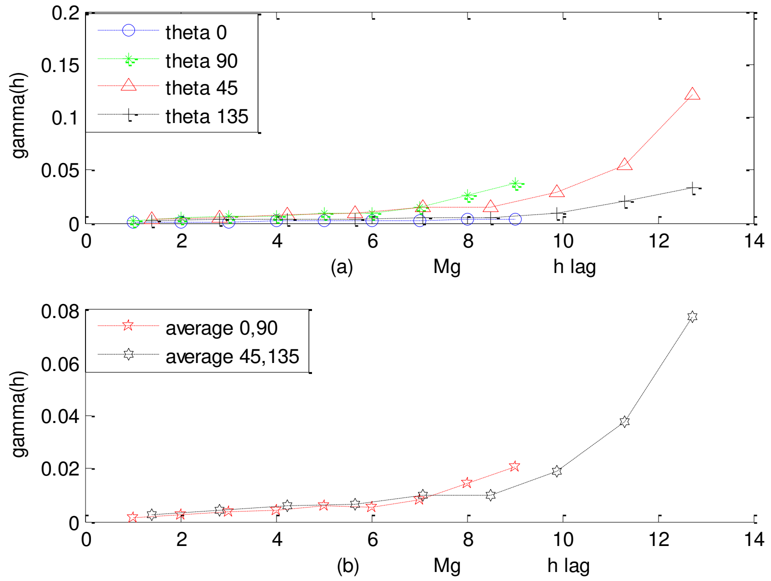


Figure 4: variogram functions of Mg data, (a) in all directions, (b) average variogram

Figure (4) explain the curves the experimental variogram functions, by using the data of the first group for Mg. Figure (4a) the curves of variogram in all directions($\theta = 0°, 90°, 45°, and\ 135°$),, while Figure (4b) shows the average of variogram functions in ($\theta = 0°, 90°$)and ($\theta = 45°, 135°$), because the same lag of h as mentioned previously.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gamma $0°, 90°$ | 0.0068 | 0.0093 | 0.0104 | 0.0138 | 0.0193 | 0.0264 | 0.0173 | 0.0222 | 0.0309 |
| Gamma $45°, 135°$ | 0.0131 | 0.0182 | 0.0202 | 0.0266 | 0.0373 | 0.0509 | 0.0316 | 0.0445 | 0.0588 |

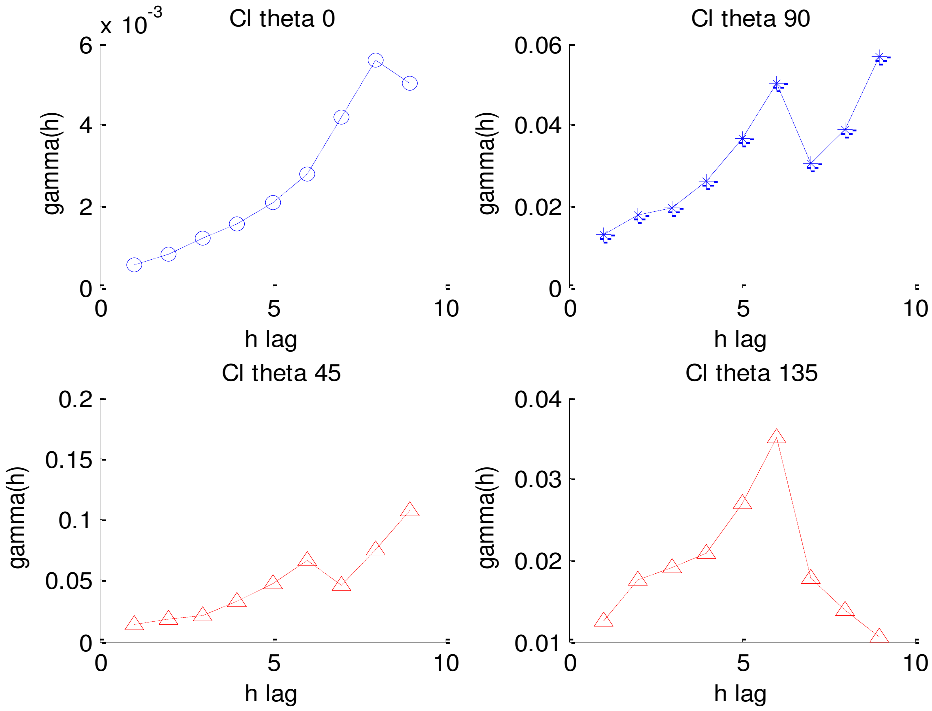


Figure 5: curves of variogram function of Cl data

Figure (5) show the curve of the experimental variogram function according equation (1), by using the data of the second group for Cl, in all directions$(\theta = 0°, 90°, 45°, and\ 135°)$,.

Table (3) shows the average of variogram function, using the data of Cl. Because the two directions $(\theta = 0°, 90°)$haves the same lag $h$, $h = 1,2, \ldots, 9$, where we found the average of variogram, while the variogram function was found by two directions $(\theta = 45°,\ 135°)$, , where $h = 1.414, 2.828, \ldots, 12.72$. Also, these two angles having the same lag $(h)$ were found to be the average of variogram function.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Gamma $(0^0, 90^0)$ | 0.0016 | 0.0027 | 0.0039 | 0.0042 | 0.0060 | 0.0055 | 0.0084 | 0.0144 | 0.0205 |
| Gamma $(45^0, 135^0)$ | 0.0027 | 0.0041 | 0.0057 | 0.0064 | 0.0098 | 0.0099 | 0.0189 | 0.0376 | 0.0676 |

Table 3: results of average of variogram function of Cl data

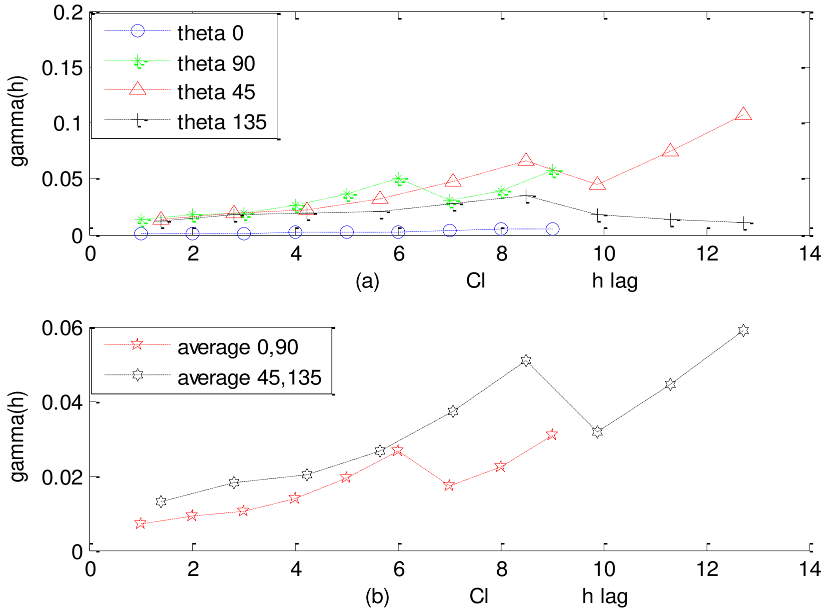


Figure 6: curves of average of variogram function of Cl data

Figure (6) explain the curves the experimental variogram functions, by using the data of the first group for Cl. Figure (6a) the curves of variogram in all directions $(\theta = 0°, 90°, 45°, and\ 135°)$,, while Figure (6b) shows the average of variogram functions in $(\theta = 0°, 90°)$, and $(\theta = 45°, 135°)$, because the same lag of h as mentioned previously.

Table (4) below shows the results of the experimental variogram functions, by using the data for Mg and Cl. The two columns of variogram function in all directions $(\theta = 0°, 90°, 45°, and\ 135°)$, for Mg and Cl, while shows the third column is result of cross variogram of Mg and Cl.

Table(4): results variogram and cross varigram functions of (Mg,Cl )data

| **Mg** | **Cl** | **Cross(Mg, Cl)** |
|---|---|---|
| **22.8335** | **98.7452** | **7.7403** |
| **25.8454** | **132.0987** | **17.0148** |
| **31.5339** | **123.7253** | **26.2601** |
| **40.1776** | **111.0988** | **38.4092** |
| **59.8202** | **140.4894** | **43.7184** |
| **55.8966** | **114.3383** | **40.4860** |
| **43.7589** | **90.5646** | **26.7150** |
| **36.8287** | **131.9825** | **38.1440** |
| **31.3665** | **141.0141** | **35.3195** |
| **25.7040** | **79.6113** | **34.9735** |

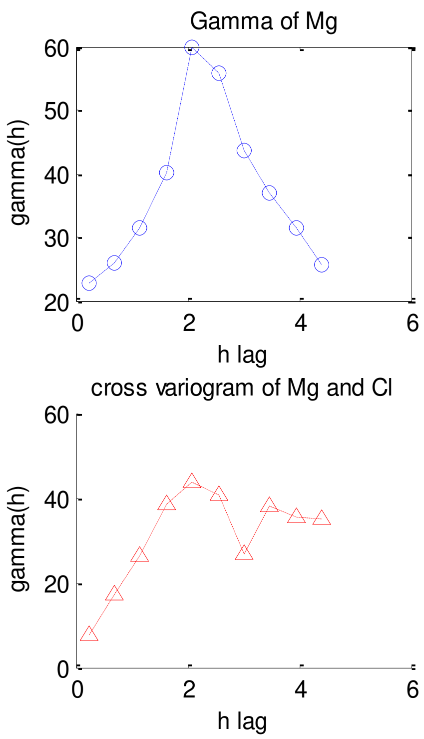


Figure 7: results of variogram and cross varigram functions of (Mg,Cl )data

Figure (7) shows the curves the experimental variogram functions, by using the data of the first group for Mg and second group for Cl. The two curves (blue) of variogram in all directions ($\theta = 0°, 90°, 45°, and\ 135°$), for Mg and Cl, while shows the red curve is cross variogram of Mg and Cl. Figure (7) gives the average of variogram function. It was noted that curves in all directions are nearest to the spherical model that is defined as:

$$\gamma(h) = \begin{pmatrix} c_o + c & , & h > a \\ c_o + c[1.5\left(\frac{h}{a}\right) - 0.5(\left(\frac{h}{a}\right)^3 & , & h < a \\ c_o & , & h = a \end{pmatrix} \qquad (16)$$

Where $h = a$ and a is the range, $c_o$ is Nugget effect and $c_o + c$ is the variance, it was noted that the curve of variogram function for Mg cut the vertical axis at $c_o = 4.74$, and the range $a = 4.175$when the curve of variogram is stable, and the variance $c_o + c$ =43.72.

These results of variogram give the knowledge to define the properties of the variogram function for Mg and Cl data, with results of prediction, Refer to Table (6).

Table (6): results of prediction for Mg and Cl

### 3.3 Prediction of cokriging technique

Cokriging technique used to predict unmeasured location for the primary variable(Mg) by using the secondary variable(Cl). This prediction used for five random locations by applying equation (9) to obtain the weights for each variable. The results of weights were obtained where the nearest data has the bigger weights while, the furthest data has the smallest weights and is near to zero. And the sum of primary variable weights is equal to one while

| Locations of prediction | $\hat{y}(s_0)$for Mg | $\sigma^2_{cok}$ | $MSE_{cok}$ | g-value | $\hat{y}(s_0)$for Cl | $\sigma^2_{cok}$ | $MSE_{cok}$ | g-value |
|---|---|---|---|---|---|---|---|---|
| (12.6 , 7.1) | 1.14 | 0.1134 | −0.243 | 0.966 | 1.80 | 0.137 | 0.100 | 1.002 |
| (6.9 , 10.7) | 3.15 | 0.0224 | −0.253 | 1.003 | 0.34 | 0.051 | −0.444 | 0.898 |
| (7.9 , 8.3) | 5.43 | 0.056 | −0.105 | 0.974 | 2.32 | 0.103 | −0.506 | 1.005 |
| (2.5 , 9.7) | 3.39 | 0.009 | −0.662 | 0.979 | 1.84 | 0.454 | −0.111 | 0.969 |
| (0.4 , 4.2) | 1.88 | 0.103 | −0.678 | 1.004 | 2.55 | 0.214 | 0.009 | 0.978 |

the sum of co-variable weights is equal to zero, that is the unbiased condition. The variance of cokriging (according equation (13)), is compute in these locations to get the accuracy of prediction process. Table (6) shows a comparison between both metals of Mg and Cl using cokriging variance. The observed points are display as a reference for the cokriging models. The cokriging model preformed as a trend surface. Most of the values of cokriging variance are very small and also MSECOR (according equation(15)). This proves the accuracy of the cokriging technique and, likely, supplies a good prediction. Also, most of the g values (according equation(14)) are closer to 1 of cokriging technique which proves the effectiveness of prediction process.

### 4. Conclusion

In this paper, the values at unmeasured locations in the field of study were estimated by using multivariate geographical statistics. Multivariate geostatistics uses to find the solution to spatial prediction with co-regionalized variables, to improve prediction of environmental modeling. Cokrigink techniques have a benefit over kriging techniques because they contain multiple variables. The variable in this paper is isotropic, illustrated by studying the varogram curves of the compass directions. The results obtained indicate small differences in the estimate. Weights were obtained where the closest data contains the largest weights, while the further data contains the smallest weights and is close to zero. The advantages of cokriging can be added to the technique of interpolation as auxiliary variables. The performance of the magnesium variable (primary variable) may be attributed to the multiple cutting of the chlorine variable (co-variable). The result of this work is to show that the chlorine data are more effective than the magnesium (primary variable). The behavior of all data is closest to the spherical model of the covariance function when using the covariate function. The accuracy of the cokriging technique gives a good prediction.

## References


- Brus DJ, Heuvelink GB (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138(1):86–95
- Chiles, J. and Delfiner, P. (2012). Geostatistics: Modelling Spatial Uncertainty, 2nd edition. John Wiley and Sons, New York. 576 pp.

- Dalezios NR, Loukas A, Bampzelis D (2002) Universal kriging of hail impact energy in greece. Physics and Chemistry of the Earth. Parts A/B/C 27(23):1039–1043
- Delobbe L, Holleman I (2006) Uncertainties in radar echo top heights used for hail detection. Meteorol Appl 13(4):361–374
- Dobesch H, Dumolard P, Dyras I (2007) Spatial interpolation for climate data: The use of GIS in climatology and meterology, iste ltd, 6 fitzroy square, london
- Eldeiry AA, Garcia LA (2010) Comparison of ordinary kriging, regression kriging, and cokriging techniques to estimate soil salinity using landsat images. J Irrig Drain Eng 136(6):355– 364
- Erdin R, Frei C, Künsch HR (2012) Data transformation and uncertainty in geostatistical combination of radar and rain gauges. J Hydrometeorol 13(4):1332–1346
- Hatem K.S. (2007) Hydrogeological hydro chemical study of Mosul Quadrangle sheet(NJ-38-13). State company of geological survey and mining, department of mineral investigation. Baghdad-Iraq.
- Trapero L, Bech J, Rigo T, Pineda N, Forcadell D (2009) Uncertainty of precipitation estimates in convective events by the Meteorological Service of Catalonia Radar Network. Atmos Res 93(1):408–418
- Wackernagel H (2013) Multivariate geostatistics: an introduction with applications. Springer Science & Business Media
- Wackernagel, H. (2010). Multivariate Statistics. Springer-Verlag, Berlin. 387 pp.
- Xie H, Zhang X, Yu B, Sharif H (2011) Performance evaluation of interpolation methods for incorporating rain gauge measurements into NEXRAD precipitation data: a case study in the Upper Guadalupe River Basin. Hydrol Process 25(24):3711–3720.