Comparison between Kriging and Lognormal Kriging for Sulfur Data

Najla Sedek Yehya

**Department of Mathematics, College of Education for Pure Science,
University of Mosul, Mosul, IRAQ.**

E-mail: najla.sedek @uomosul.edu.iq

Abstract

This paper deals with the spatial estimation to be optimal estimate when the data have a normal distribution. Ordinary kriging technique used in this research and lognormal kriging after take logarithm of original data. Variogram function uses in this research to obtain the best model of covariance function. The objective of this research is to assess the resistance of ordinary kriging and lognormal kriging to outliers. The data adopted of this paper from hydrogeological hydro chemical study of Mosul Governorate /Iraq. From the results, the errors in the estimated value are critical for the log variance. These results show that ordinary lognormal kriging is more effect than ordinary kriging technique under outlier.

Keywords: spatial data, ordinary kriging, lognormal, variogram function.

1. Introduction

The lognormal distribution is a special form of distribution that has only one mode, but is more skewed than the negative binomial. (Eldeiry, Garcia, 2010) uses kriging system to predict the value by types of interpolation methods. When logarithms of counts follow a normal frequency distribution, the original counts must follow a discrete lognormal distribution. The logarithmic transformation of counts often provides a useful approximate model to normalize data in a negative binomial distribution. We can analyze how strength is distributed across nodes in a similar way to the analysis of degree distributions. Indeed, analysis of human structural connectivity networks constructed with diffusion indicates that both the strength distribution and degree distribution of brain networks have a similar form (Hagmann et al., 2007).

In the biology sciences in particular, the application of the normal logarithm as well as the logarithmic distribution have been dealt with in a number of different aspects, in addition to its application in the environment, geology, plants and meteorology, see (Crow, et. al, 1988) It is also applied in astrophysics, see (Parravano, et. al, 2012). A log-normal distribution can be defined as the distribution of a random variable. Recently, Laplace logarithm distributions have been proposed to model growth rates as stock prices and currency exchange rates Also presented is a random vector distribution following the logarithm of the normal distribution of order, which is an exponential generalization of the

force of a normal distribution, given by (Kitsos and Tavoularis,2009). This is a family of dimensional generalized normal distributions.

Researchers have also provided collection and dissemination of soil information in some scientific projects (such as Rossiter et al., 2015). As soil information can be obtained from agricultural fields, it is of great interest to benefit from data and explore important applications (Pie, et al., 2010; Mao, et al., 2014, Shiliang, et al., 2015;).

2. Methods of geostatisics

2.1 Variogram function

Let $H(u)$and $H(u+h)$be two random variables at two points (u) and $(u+h)$separated by the vector (h). The variability between these two quantities is characterized by:

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

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

$$2\gamma(h) = \frac{1}{n(h)}\sum_{i=1}^{n(h)}[H(u) - H(u+h)]^2 \quad (1)$$

2.2 Ordinary kriging. The estimator is:

$$\hat{H}_k(u_o) = \sum_i \omega_i H(u_i), \quad (2)$$

where $\omega_i = \lambda_{ok_i}$.

Kriging weights ω are obtained by solving the system:

$$\sum_i \omega_i \gamma(u_i, v_j) + r = \gamma(v_j, u_o), j = 1, \dots, n$$
$$\sum_i \omega_i = 1, \quad (3)$$

where $\gamma(h)$ is the variogram of $H(u)$ and r is the Lagrange multiplier. The symbol μ is reserved for the mean of the logs. If stationarity can be assumed, then the spatial covariance $c(h)$ also exists and is related to the variogram by:

$$\gamma(h) = c(0) - c(h).$$

The kriging system can be rewritten as:

$$\sum_i \omega_i c(u_i, v_j) + r = c(u_i, u_o), j = 1, \dots, n$$
$$\sum_i \omega_i = 1,$$

2.3 Ordinary lognormal kriging. Assume that $H(u)$ has a two-parameter lognormal distribution. The three-parameter lognormal is a simple extension of the two-

parameter case. So $F(u) = lnH(u) \sim \mathrm{N}(\mu, \sigma^2)$. The mean, variance, and spatial covariance function of $H(u)$, which is denoted by M, Σ, and $c(h)$ respectively, are related to the mean, variance, and covariance of $F(u)$ as follows:

1- mean: $E[H(u)] = \mathrm{m} = \log\left(\frac{-\sigma^2}{2}\right)$

2- variance: $var[H(u)] = \log\left(1 + \frac{-\sigma^2}{m^2}\right) = C(0)$

3- covariance: $Cov(h) = \log\left(1 + \frac{-\sigma^2}{m^2}\right)$

4- variogram $\gamma(h) = C(0) - C(h)$

$$\hat{H}_{oL_k}(u_o) = \exp\left[F(u) + \frac{\sigma^2_{oLk}}{2} + r\right] \quad (4)$$

Where $C(h)$ denotes the covariance of $F(u)$.

The skewness of the distribution can be gauged by calculating the coefficient of variation η for $H(u)$,

$$\eta^2 = \frac{variance}{mean^2} = \exp(\sigma^2) - 1. \quad (1)$$

The ordinary lognormal kriging estimator $\hat{H}_{oL_k}(u_o)$ is given by:

$$\hat{H}_{oL_k}(u_o) = \exp\left[\sum_i \omega_i F_i + \frac{\hat{\sigma}^2_{oLk}}{2} + r\right] \quad (2)$$

where $\omega_i = \sum_i \lambda_{oLk_i}$.

Weighting factors ω_i , Lagrange multiplier r and kriging variance σ^2_{oLk} are obtained by solving the following system:

$$\sum_i \omega_i P(u_i, v_j) + r = P(v_j, u_o), j = 1, \dots, n$$
$$\sum_i \omega_i = 1, \quad (3)$$

Clearly, the values of σ^2_{oLk} and r in this system depended linearly on the value of the sill of the variogram of the logs ($P(0) = \sigma^2$) but the weights ω_i do not.

The outlier effect index. In this research, the effect of a single outlier in a group of otherwise similar values. This being the simplest case, is quantified. Using the terminology proposed by Barnett & Lewis (1978) this large value would be considered as "discordant" (i.e., an "outlier") if the underlying population is assumed to be normal,

whereas it would not if the population is skew (lognormal or otherwise). In that case it would merely be part of the " long tail",J.

All of the sample value except one are assumed to be equal. For case where the mean is know (i.e., ordinary kriging), let $H_1 = \alpha H$ (where $H_j = H\ for\ j \neq 1$). When the mean is know, the first value H_i is equal to $\alpha_1\mu$ (or $\alpha_1 M$ for the lognormal case) whereas the rest of the values are $\alpha_2\mu$ (or $\alpha_2 M$). In both cases, the variogram model is assumed to be know. The method used here follows and extends the technique proposed by Bou fassa (1986).

2.4 Ordinary kriging. When these values for H_i are substituted in Eq. (2) we get:

$$\hat{H}_{ok}(u_o) = \mathrm{H}[1 + \omega_1(\alpha - 1)], \qquad (4)$$

and the ratio r_{ok} of the effect is:

$$r_{ok} = \frac{\hat{H}_{ok}(u_o)}{H} = 1 + \omega_1(\alpha - 1)\ . \qquad (5)$$

In particular if the values of the outlier H_1 is increased by 10% (i.e., increased $\alpha\ to\ 1.1\alpha$) then we get:

$$r_{ok} = 1 + \omega_1(1.1\alpha - 1)\ . \qquad (6)$$

The relation between r_{ok}, ω_1 and α is illustrated in Fig.(1).

2.5 Ordinary lognormal kriging. Similarly, substituting value for H_i into Eq.)4(gives:

$$\hat{H}_{oL_k}(u_o) = H(\alpha^{\omega_1})\exp\left[\frac{\sigma'_{oLk}}{2} + r\right] \qquad (7)$$

$$r_{oLk} = (\alpha^{\omega_1})\exp\left[\frac{\sigma^{o}_{oLk}}{2} + r\right] = D(\alpha_{\omega_1})\ , \qquad (8)$$

where $\mathrm{D} = \exp\left[\frac{\sigma^2_{oLk}}{2} + r\right]$ is the bias correctio factor. In graphical terms, it is the ratio of the model of the lognormal distribution to its mean value. 10% increase in H_i would then give:

$$r_{oLk} = D(1.1\alpha)^{\omega_1}\ , \qquad (9)$$

that is, the effect would be attenuated if ω_1 is less than 1, which is usually the case. gt would be magnified in the unlikely case where the kriging weight is greater than 1.(This sometimes occurs with " double points" or when using a Gaussian variogram with no nugget effect). The relation between r_{oLk}, ω_1 and α is illustrated

2.6 Evaluation of kriging methods

To evaluate the performance of prediction method, we computed the mean error(ME), root mean square error(RMSE), and the coefficient of determination (R^2)

$$ME = \frac{1}{n}\sum_{i=1}^{n} |H(u_i) - \hat{H}(u_i)| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} [H(u_i) - \hat{H}(u_i)]^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} |H(u_i) - \hat{H}(u_i)|^2}{\sum_{i=1}^{n} |H(u_i) - \overline{H(u_i)}|^2} \quad (16)$$

Where $H(u_i)$ is the measured value of H at location u_i, $\hat{H}(u_i)$ is the predicted value of location, and $\overline{H(u_i)}$ is the mean of the measured value.

3. Results and Analysis

3.1 Data study

Due to the importance of sulfur in the soil, sulfur is a non-metallic chemical element that belongs to the oxygen group and is one of the most reactive chemical elements. Sulfur has some toxic effects at times. In addition, it may cause eye irritation, skin toxicity, and suffer from some problems and risks of inhalation. The data contain (100) samples of real values with their locations of sulfur (S). These data are a real spatial data from hydrogeological hydro chemical study of Mosul Quadrangle/Iraq, (Hatem, 2007). Table (1) below content the data statistic of sulfur (S).

Table (1): data statistic for sulfur (S)

Stat. / Data	Min	Max	Median	Mode	Std
sulfur (S)	**0.2000**	**36.5000**	**6.75000**	**3.70000**	**6.1525**

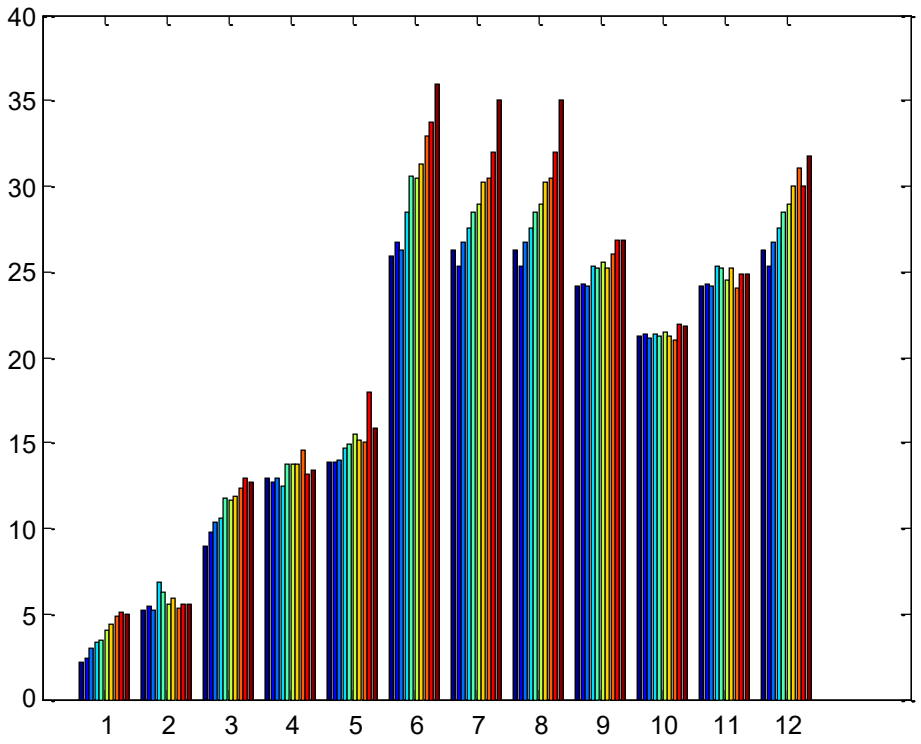

Figure 1: Histogram of S data

'

3.2 Variogram function

We applied the experimental variogram function according equation (1), to plot the curves of variogram function and by using the data for sulfur (S), in all directions $\left(\theta = 0^{\circ}, 90^{\circ}, 45^{\circ}, \text{and} 135^{\circ}\right)$. Table (2) below shows the results of the experimental variogram function in all direction of compass.

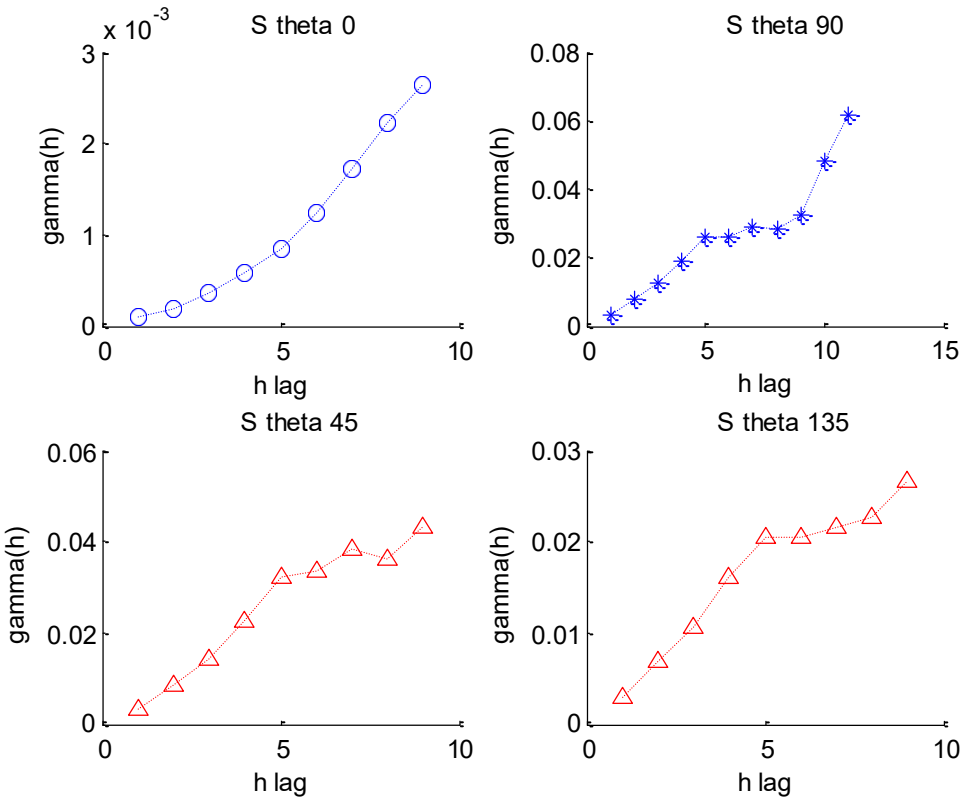

Figure 1: curves of variogram function in all directions for sulfur (S)

Figure (1) show the results of variogram function in all directions $\left(\theta = 0^{\circ}, 90^{\circ}, 45^{\circ}, \text{and} 135^{\circ}\right)$ for sulfur (S) between lag h and gamma(h). And the results of variogram function in all directions of compass ,where gamma1 for theta 0, gamma2 for theta 90, gamma3 for theta 45, and gamma4 for 135. (see Table (2) below),

Table 2: results of variogram function in all directions for sulfur (S)

gamma1	0.0001	0.0002	0.0004	0.0006	0.0008	0.0012	0.0017	0.0022	0.0026
gamma 2	0.0032	0.0077	0.0126	0.0193	0.0259	0.0260	0.0290	0.0282	0.0324
gamma 3	0.0033	0.0084	0.0144	0.0226	0.0322	0.0336	0.0384	0.0361	0.0431
gamma 4	0.0030	0.0069	0.0107	0.0161	0.0205	0.0205	0.0217	0.0226	0.0266

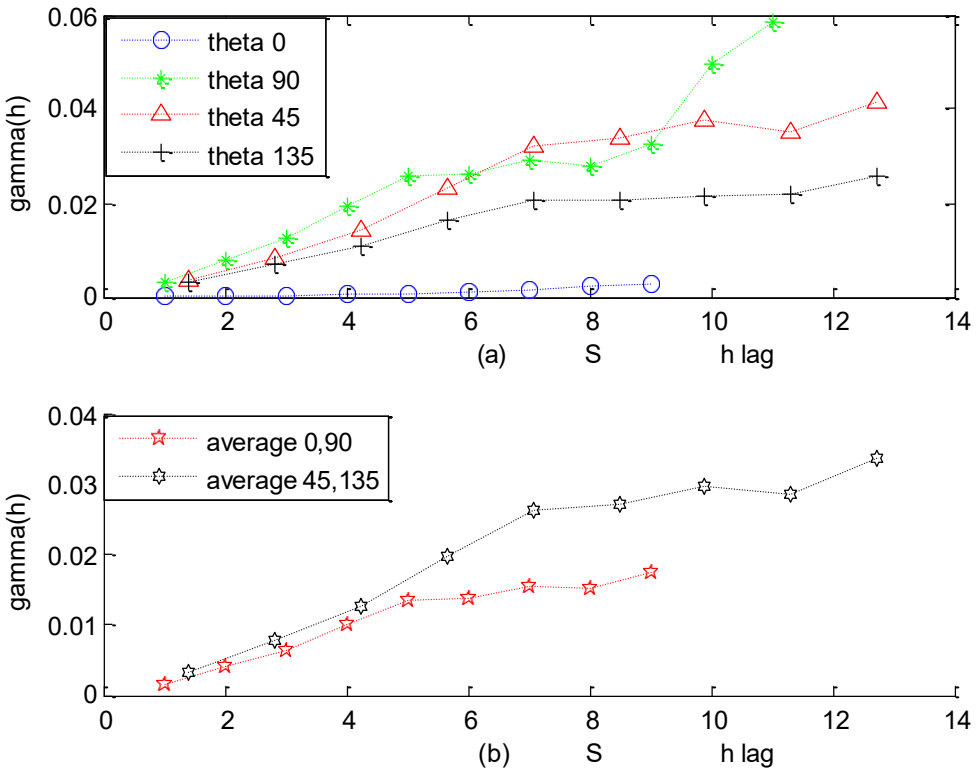

Figure 2: curves of variogram function (a) in all directions for sulfur (S), (b) average of variogram function

Figure (2) illustrates curves of variogram function (a) in all directions $(\theta = 0^{\circ}, 90^{\circ}, 45^{\circ}, \text{and} 135^{\circ})$ for sulfur (S), (b) average of variogram function $(\theta = 0^{\circ}, 90^{\circ})$ because have the same lag, and $(\theta = 45^{\circ}, 135^{\circ})$ have the same lag. And by statistics of average variogram function of $(\theta = 0^{\circ}, 90^{\circ})$, we show from the curves of average variogram function that nugget effect = 0.0017, sill= 0.0176, and range=8 on x-axis, while the curve of average variogram function of $(\theta = 45^{\circ}, 135^{\circ})$ illustrate nugget effect =0.0033, sill= 0.0336, and range is (11.31) on x-axis. (see Table (3) below)

Table 3: results of average variogram function

gamma5	0.0017	0.0040	0.0064	0.0101	0.0133	0.0137	0.0153	0.0151	0.0176
gamma6	0.0033	0.0078	0.0125	0.0197	0.0264	0.0271	0.0297	0.0285	0.0336

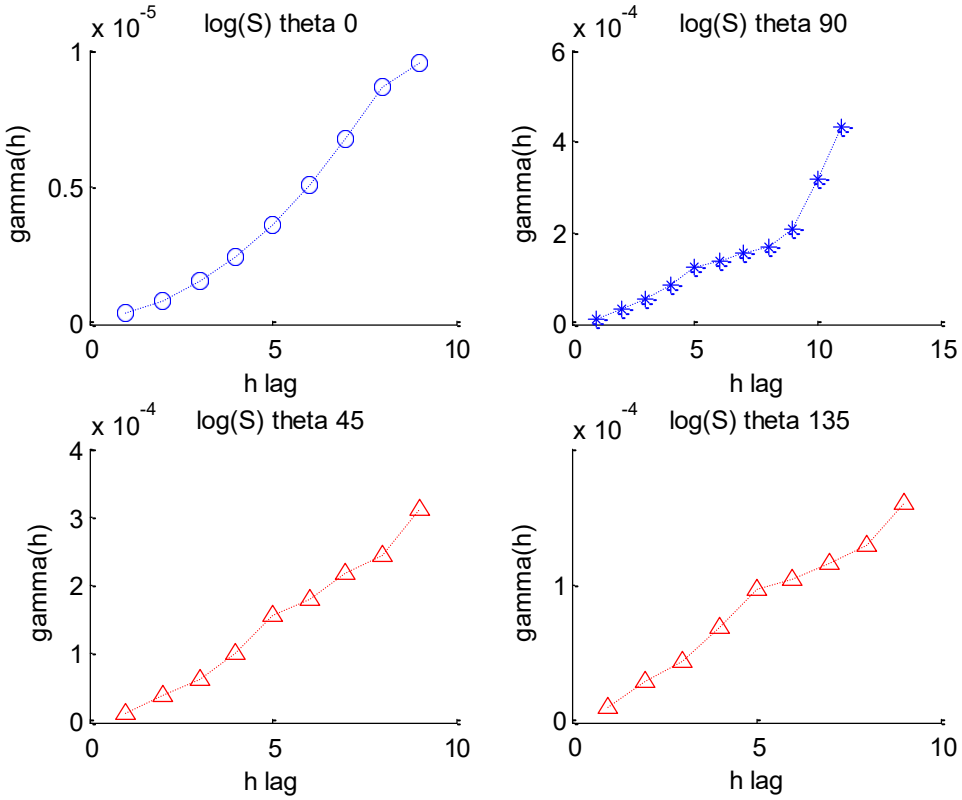

Figure 3 : curves of variogram function in all directions for log (S)

Figure(3) shows curves of variogram function in all directions $\left(\theta = 0^{\circ}, 90^{\circ}, 45^{\circ}, \text{and} 135^{\circ}\right)$ for sulfur (S),

Table 4: results of variogram function in all directions for log (S)

gamma1 1.0e-005 *	0.0392	0.0864	0.1534	0.2478	0.3578	0.5096	0.6743	0.8661	0.9509
gamma 2 1.0e-003 *	0.0128	0.0346	0.0548	0.0854	0.1241	0.1365	0.1567	0.1707	0.2089
gamma 3 1.0e-003 *	0.0137	0.0391	0.0636	0.1017	0.1548	0.1781	0.2175	0.2448	0.3107
gamma 4 1.0e-003 *	0.0115	0.0295	0.0444	0.0695	0.0972	0.1038	0.1160	0.1286	0.1604

Table (4) illustrate the results of variogram function in all directions for log(S), Where gamma1 (theta 0), gamma2 (theta 90), gamma3 (theta 45), and gamma4 (theta 135)

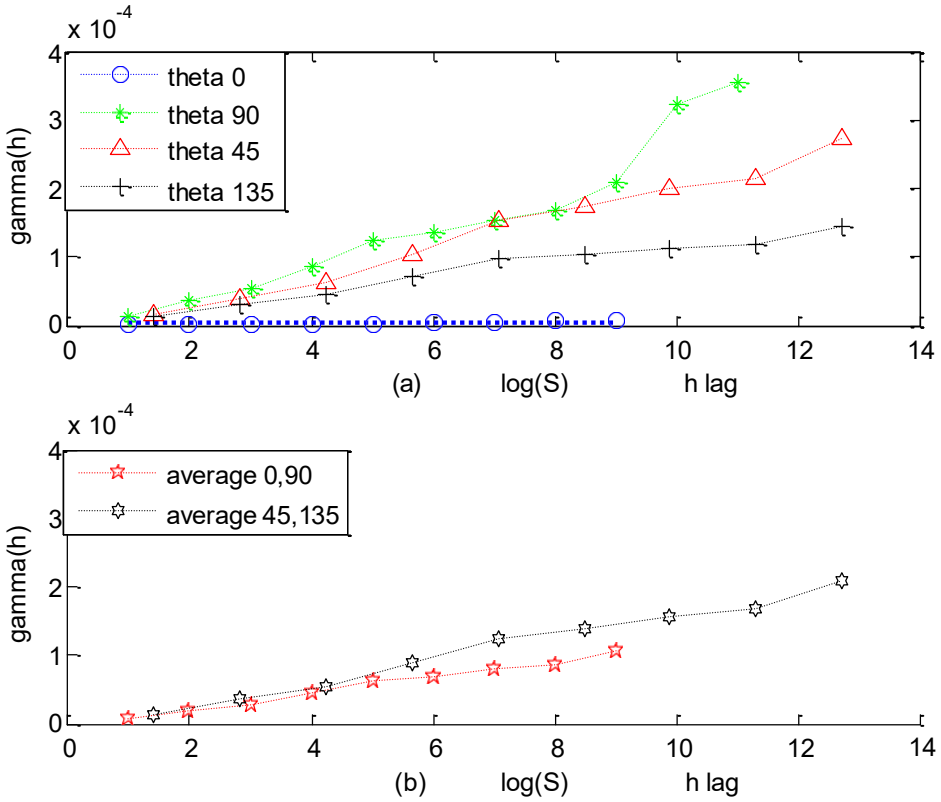

Figure 4 : curves of variogram function (a) in all directions for log (S), (b) average of variogram function

Figure(4) shows curves of variogram function, (a) in all directions for log (S), (b) average of variogram function $(\theta = 0^{\circ}, 90^{\circ})$ because have the same lag, and $(\theta = 45^{\circ}, 135^{\circ})$ have the same lag. And by statistics of average variogram function of $(\theta = 0^{\circ}, 90^{\circ})$, we show nugget effect = (0.0073) *1.0e-003, sill= (0.1073) *1.0e-003 , and range=8 on x-axis, while the curve of average variogram function of $(\theta = 45^{\circ}, 135^{\circ})$ illustrate nugget effect is (0.0143) *1.0e-003, sill= (0.2093) *1.0e-003, and range is (11.31) on x-axis.

Table 5: results of average variogram function for log (S)

gamma5 1.0e-003 *	0.0073	0.0181	0.0274	0.0446	0.0626	0.0694	0.0795	0.0864	0.1073
gamma6 1.0e-003 *	0.0143	0.0361	0.0541	0.0884	0.1245	0.1381	0.1568	0.1671	0.2093

It was noted that curves in all directions are nearest to the gaussian model that is defined as:

$$\gamma(h) = (c_o + c\ [1 - \exp(\frac{-|h|^2}{a^2})]) \quad , \quad h \neq 0 \qquad (17)$$

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 data, cut the vertical axis at $(c_o = (0.0073) * 1.0e - 003)$, and the range $(a = 11.31)$ when the curve of variogram is stable, and the variance $c_o + c = (0.2093) * 1.0e - 003$.

3.3 Prediction of kriging technique

Kriging technique used to predict unmeasured location for the regionalized variable. This prediction used for five random locations by applying equation (6) 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 one that is the unbiased condition. The variance of kriging is compute in these locations to get the accuracy of prediction process. The observed points are display as a reference for the kriging models. The kriging model preformed as a trend surface. Most of the values of kriging variance are very small and also ME, RMSE and R^2 (see Table 6). This proves the accuracy of the kriging technique supplies a good prediction, which proves the effectiveness of prediction process. we need to understand how outlier work, that the kriging process has two steps: first step, build a semivariogram to model the spatial relationships between points, to find the correlation between data points based on their distance. And the second step is to use the semivariogram and a dataset to make predictions at new location. We use the whole data(outliers included) in the prediction, because the modeling is not corrupted by the outliers, but the prediction surface still accounts for the extreme values, for example outlier (0.0336) for gamma of sulfur data (S), while (0.2093) multiply *1.0e-003 for gamma of log (S).

Table 6: results of cross validation

ME	0.0345	0.0111	0.0244	0.0431	0.0126	0.0324	0.055	0.0554	0.103
RSME	0.73	0.67	0.545	0.887	0.645	0.741	0.678	0.857	0.963
R^2	0.932	0.884	0.794	0.887	0.964	0.890	0.797	0.869	0.978

4. Conclusion

Through our findings, we notice that the two-dimensional variogram function is very similar to the Gaussian mathematical model, and this is shown by the congruence of the theoretical and practical studies. Robust estimators can overcome the bias caused by outliers. This supports that the values of the normal kriging variance are very small and with negligible differences, as well as the variance of the kriging estimator is small which indicates the error is very small and that the weights are close to the same for one as the function diagram after prediction increases as the displacement h increases to a certain extent and then stabilizes, which indicates the compatibility of the mathematical model with the data of the applied example. The curve of origin data of sulfur is nearest of the curve after took logarithm of data and near than Gaussian model with their properties.

References

- Crow, E. L. and K. Shimizu, *Lognormal Distributions*, Marcel Dekker, New York, NY, USA, 1988.View at: Zentralblatt MATH | MathSciNet
- 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.
- Hagmann, P , Kurant, M, Gigandet, X, Thiran, P, Wedeen, V, Meuli, R , and Thiran, J. Ph. Mapping Human Whole-Brain Structural Networks with Diffusion MRI, 2007; 2(7): e597. PMCID: PMC1895920 Published online 2007 Jul 4. doi: 10.1371/journal.pone.0000597
- 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.
- Kitsos, C. P. and N. K. Tavoularis, "Logarithmic Sobolev inequalities for information measures," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2554–2561, 2009.View at: Publisher Site | Google Scholar | MathSciNet
- Mao, Y.M., Sang, S.X., Liu, S.Q., Jia, J.L., 2014. Spatial distribution of pH and organic matter in urban soils and its implications on site-specific land uses in Xuzhou, China. C. R. Biol. 337 (5), 332–337.
- Parravano, A, N. Sanchez, and E. J. Alfaro, "The Dependence of prestellar core mass distributions on the structure of the parental cloud," *The Astrophysical Journal*, vol. 754, no. 2, article 150, 2012.View at: Publisher Site | Google Scholar
- Pei, T., Qin, C.Z., Zhu, A.X., Yang, L., Luo, M., Li, B.L., Zhou, C.H., 2010. Mapping soil organic matter using the topographic wetness index: a comparative study based on different flow-direction algorithms and kriging methods. Ecol. Indic. 10 (3), 610–619.
- Rossiter, D. G., Liu, J, Carlisle, S., Zhu,A. Can citizen science assist digital soil mapping.December,2015.Geoderma 259-260:71-80, DOI:10.1016/j.geoderma.2015.05.006
- Shiliang Liu, Nannan An, Juejie Yang, Shikui Dong, CongWang, Yijie Yin(2015) Prediction of soil organic matter variability associated with different land use types in mountainous landscape in southwestern Yunnan province, China. journal homepage: www.elsevier.com/locate/catena Catena 133 (2015) 137–144.