© Universiti Tun Hussein Onn Malaysia Publisher's Office

IJIE

Journal homepage: http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

Ordinary Kriging as a Method to Determine the Clay Mapping Distribution in Highland Area of Sabah

Mohammad Radzif Taharin[1*], Adriana E. Amaludin[1], Mohd Adib M. Razi[2]

[1]Faculty of Engineering,
Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, MALAYSIA

[2]Faculty of Civil and Environmental Engineering,
Universiti Tun Hussein Onn, 86400 Parit Raja, Johor, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/ijie.2019.11.08.027
Received 05 December 2018; Accepted 07 September 2019; Available online 15 December 2019

Abstract: Ordinary Kriging (OK), is one of the geostatistical methods, which were used in the variation types of mapping, which related to the soil. Compliment by semivariogram models, OK has become one of the most sought out method for the digital mapping, which applied Geographical Information System (GIS) as a main approach. In this method, four semivariogram models, which are spherical, exponential, circular and gaussian would be applied to determine the best model for the mapping purposes, with RMSE as a performance indicator. This method would be implemented in determining the clay mapping distribution for Kundasang area in Sabah, which has been known for the highland tourism and susceptible for the geohazard such as soil mass movement and landslides. Through this method, the clay percentage distribution mapping would be produced.

Keywords: Ordinary Kriging, semivariogram, clay distribution map

1. Introduction

Geographical Information System (GIS) has become one of the most popular tools for determine the topography and the landscape of the earth in recent times. Compliment with Global Positioning System (GPS) coordinates, GIS could manage to detect any unexpected event on the location, regarding to the natural disaster with great precision [1].

GIS have been used for many purposes, especially in the geohazard early warning system. Early warning of rainfall-induced shallow landslides and debris flow is one of the many examples [2]. Even landslides susceptibility mapping could also produced by using GIS [3]. Certain functions in the GIS software have made the task more convenient than before. For example, functions like geostatistics, which contain several methods of interpolation such as Radial Basis Function (RBF), Inversed Distance Weightage (IDW) and Kriging. IDW have been used previously to determine the peat location in Gold Coast before [4].

One of the most popular methods for geostatistics is Ordinary Kriging. This method has been implemented in many places that related to the study that involved in the soil mapping, especially in the agricultural purposes. This method would be applied to determine the clay percentage distribution in one of the most popular tourist attractions in Malaysia, which is Kundasang, Sabah.

Kundasang, Sabah (Fig. 1) is well known for the highland tourism in Malaysia, especially for the Mount Kinabalu, which heights is 4095.2 meters above sea level. This area is popular due to its topography and cool climate, which contributes to the crops production such as vegetables and fruits, and dairy products. Despite blessed with the serene

*Corresponding author: radzif1@gmail.com

environment, this area is susceptible to the geohazard such as soil mass movements and landslides, which were occurred in high frequency.

Located on two formations, which are Crocker Formation, and Trusmadi Formations, soil in Kundasang is known for low cohesion and high angle of internal friction values [5]. Since cohesion value is usually associated with clay, it is essential to determine the clay percentage distribution of the area, to get the actual view of one of the soil properties. This study would determine the clay percentage distribution for 70 boreholes (Fig. 2), from ground level towards 10 meters depth, divided into four equal sections, which are 0 to 2.5 meters, 2.5 to 5 meters, 5 to 7.5 meters, and 7.5 to 10 meters.

Fig. 1 - Kundasang area, with Mount Kinabalu as a background.

Fig. 2 - Boreholes location in Kundasang Town area.

2. Ordinary Kriging Method

Kriging methods were found by D.G. Krige, in 1953, where the prediction is needed to identify which area has the highest calorie of coal content. Krige introduced this method and by doing this, high calorie of coal content managed to be found and the prediction made is showing 96% of accuracy [6]. By implementing this method, lots of time has been saved in terms of trial and error exploring and excavating the unknown area, without guarantee of compensating the time spent with the expected outcome.

Kriging method has being expanded by George Matheron and become very popular and being used for many purposes not restricted to earth science and engineering only, but in other purposes as well such as medical and vector diseases control [7]. There are several types of Kriging methods, namely Ordinary Kriging, Simple Kriging, Universal Kriging, Indicator Kriging, Disjunctive Kriging, Probability Kriging and Cokriging. But the most popular method is the Ordinary Kriging (OK), due to its simplicity and accuracy.

OK is the most applied and the simplest geostatistical estimator model because of the fewest number of assumptions involved in it such as no trend, isotropy and variogram can be defined with mathematic model, compared to other methods. In OK, where the population mean is true but unknown constant, could be predicted according to the given data [7]. The mean could be determined according to the smoother pattern of data as shown in Fig. 3.

2.1 Semivariogram Models

Before OK method could be applied, semivariogram (Fig. 4) would be developed to determine the best model for each depth sections, which would be processed for the mapping development of the clay percentage. Semivariogram modelling is a key step between spatial description and spatial prediction, which is produced through the least square fit of the data. At relatively short lag distances, the semivariance is small but increases with the distance between the pairs

of sample points. At the range, the semivariance levels off to a relatively constant value referred to as the sill. The equation of the semivariogram is shown in Eq. (2).

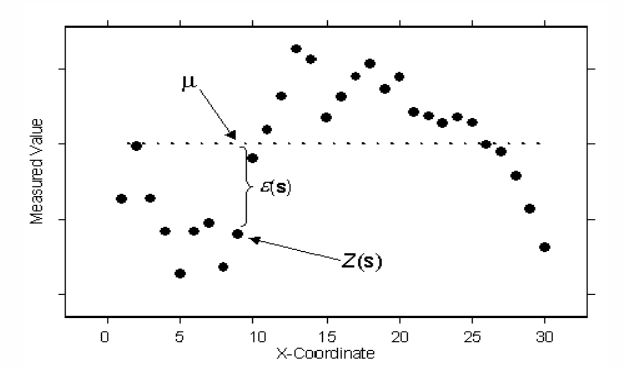


Fig. 3 - Prediction using Ordinary Kriging Method.

$$Z = \mu + \varepsilon(s) \tag{1}$$

where Z (s)= value that need to be obtained, μ = mean or the average value that need to be found, and $\varepsilon(s)$ = random error which will reduce the value error.

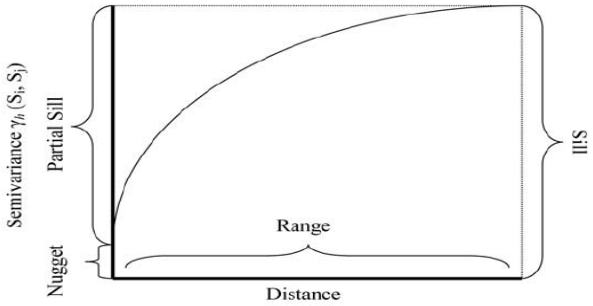


Fig. 4 - Semivariogram, which comprised of nugget, range and sill.

$$\gamma(h) = \frac{1}{2}\sum_{i=1}^{n}\left[z(x_i) - z(x_j + h)\right]^2 \tag{2}$$

where $\gamma(h)$ = semivariance between point x_i and x_j discrete by lag distance h, n = numbers of pairs of sample points separated by distance h, z = Attribute value (depends on the parameter that used for this method), and h = distance separating x_i and x_j.

Four models of semivariogram models, which are spherical, exponential, circular, and gaussian were put into test for each depth section, with Root-Mean-Squared-Error (RMSE) as a performance indicators for this test. Through RMSE, the accuracy of the semivariogram or graph could be determined by the lowest values, that managed to reach most points, as the best model for the mapping purposes. Equations for the all four models are shown in Eqs. (3) and (4) for spherical, Eq. (5) for exponential, Eqs. (6) and (7) for circular, and Eq. (8) for gaussian, where *Co* is nugget, *C* is the semivariogram line, α is range, and h is distance.

$$\gamma(h) = Co + C\left(\frac{3}{2}\frac{h}{\alpha} + \frac{1}{2}\left(\frac{h}{\alpha}\right)^3\right) \qquad 0 \le h \le \alpha \tag{3}$$

$$\gamma(h) = Co + C \qquad h > \alpha \tag{4}$$

$$\gamma(h) = Co + C\left(1 - \exp\left(-\frac{h}{r}\right)\right) \qquad h > 0 \tag{5}$$

$$\gamma(h) = Co + C\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{\alpha}\right) + \sqrt{1 - \frac{h^2}{(\alpha)^2}}\right) \qquad 0 \le h \le \alpha \tag{6}$$

$$\gamma(h) = Co + C \qquad h > \alpha \tag{7}$$

$$\gamma(h) = Co + C\left(1 - \exp\left(-\frac{h^2}{r^2}\right)\right) \quad h > 0 \tag{8}$$

2.2 Interpolation Procedure

- Select a point among surrounding points to determine the unknown value. Determine the distances and value differences between surrounding points of the selected point.
- Arrange in the tabulated form, according to the distances and the squared value differences, before determining the semivariance value (half of squared value differences).
- Binning the distance and the semivariance according to the distance categories. Graph will be plotted based on the average pairs distance and average semivariogram and the equation will be produced. The equation is representing the semivariogram of the study.
- Distance between surrounding points will be inserted in the equation. Arrange it in the form of 6x6 matrixes and insert 1 at the end column and bottom row as unbiased predictor (matrix M). This matrix will be inversed (M^{-1}).
- Insert each distance between surrounding points to the selected point, and arrange it in the form of matrice 6x1 (M_A). Insert 1 at the bottom of the Matrix B as unbiased predictor. Multiply M^{-1} and M_A. The product of the multiplication is a new 6x1 matrix (M_B).
- Arrange the known values that representing surrounding points in the column form. Multiply the values in the M_B and values directly. Obtain the total sum values of multiplication, which will become the Kriging Predictor.
- Multiply directly values in the M_A and M_B and obtain the total sum. Square root the total sum to obtain the kriging error.

2.3 Mapping Procedure

For the mapping procedure, ArcGIS was used as a tool to execute semivariogram and OK in one single process. However, this process required user to select OK function prior to the semivariogram due to the programming set by the software developer to expedite the selected Kriging methods with the selected semivariogram models. The process could be seen in the Fig. 5.

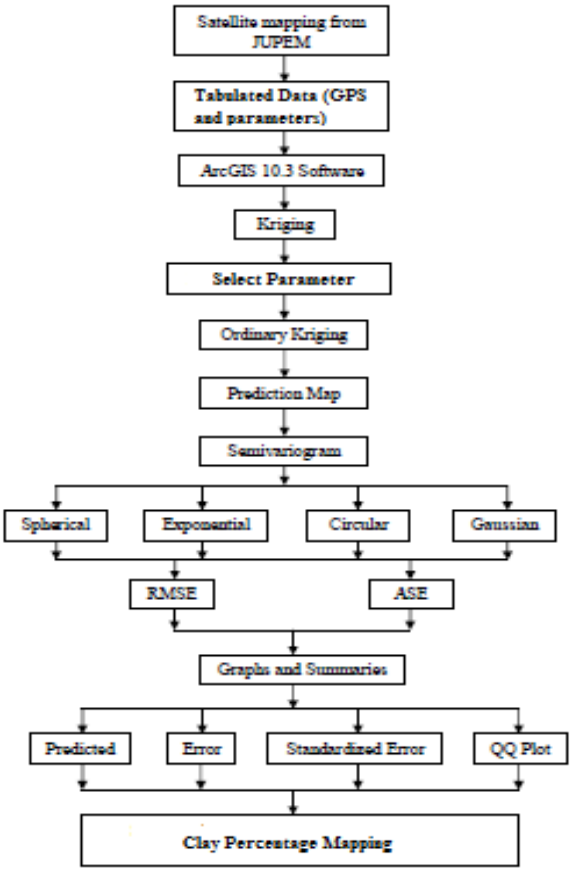
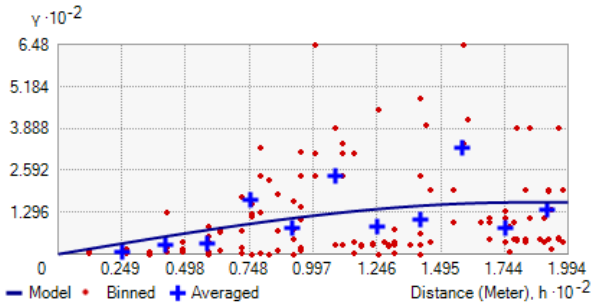


Fig. 5 - The mapping procedure by using ArcGIS 10.3.

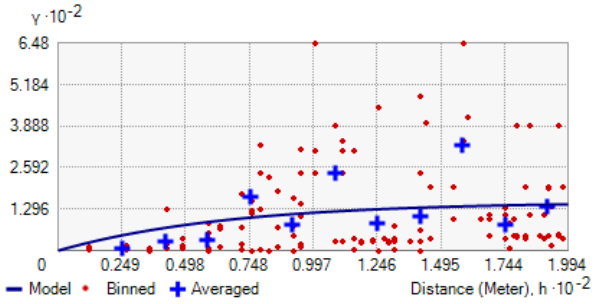
3. Results and Discussion

Fig. 6 to Fig. 9 are showing the semivariogram of clay percentage for the depth section of 0 to 2.5 meters. 18 boreholes were involved for this depth. The x-axis representing the distance in meters, while y-axis representing the average semivariance of clay percentage.



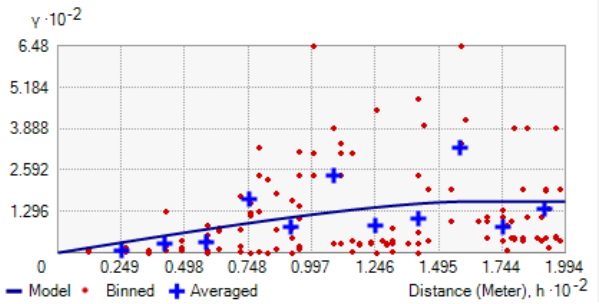
$$\gamma(h) = 0 + 160.03\left(\frac{3}{2}\frac{h}{(181.17)} + \frac{1}{2}\left(\frac{h}{181.17}\right)^3\right) \quad (9)$$

Fig. 6 - Spherical model of clay percentage for depth 0 to 2.5 meters.



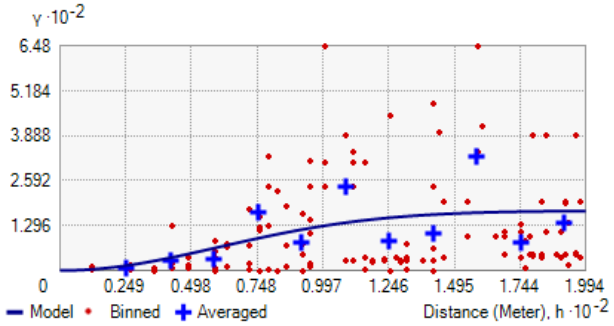
$$\gamma(h) = 0 + 152.17\left(1 - \exp\left(-\frac{h}{199.36}\right)\right) \quad (10)$$

Fig. 7 - Exponential model of clay percentage for depth 0 to 2.5 meters.



$$\gamma(h) = 0 + 160.05\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{158.45}\right) + \sqrt{1 - \frac{h^2}{(158.45)^2}}\right) \quad (11)$$

Fig. 8 - Circular model of clay percentage for depth 0 to 2.5 meters.



$$\gamma(h) = 0.17221 + 172.21\left(1 - \exp\left(-\frac{h^2}{(147.71)^2}\right)\right) \quad (12)$$

Fig. 9 - Gaussian model of clay percentage for depth 0 to 2.5 meters.

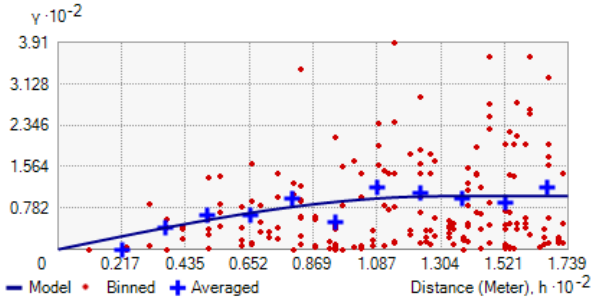
From Table 1, three out of four semivariogram models shared the similar value of nugget, which is zero, except gaussian. Sill and range for all models were different. However, the most significant value is RMSE, whereby model that produced the lowest RMSE value was exponential model.

The exponential model managed to produce the lowest RMSE value due to the fact that the semivariogram cloud was distributed along the distance, and have some inclinations form, which enable exponential model to reach most of the points in the semivariogram cloud. Hence, contribute to the higher accuracy that reflects the lowest RMSE value. And this model would be selected for the mapping purposes for this depth section.

Table 1 - Nugget, sill, range, and RMSE values for all semivariogram models for depth between 0 to 2.5 meters.

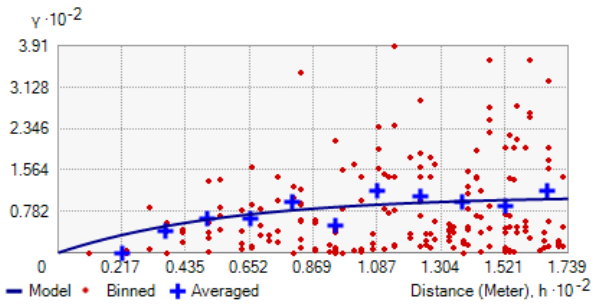
Semivariogram Model	Spherical	Exponential	Circular	Gaussian
Nugget (%)	0	0	0	0.17221
Sill (%)	160.03	152.17	160.05	172.21
Range (m)	181.17	199.36	158.45	147.71
RMSE	9.9399	8.5935	8.6688	9.6413

Fig. 10 to Fig. 13 are showing the semivariogram of clay percentage for the depth section of 2.5 to 5 meters. For this depth, the numbers of boreholes involved are 29 boreholes, which explained these semivariograms look more crowded than previous depth. Similar to the previous depth, the x-axis representing the distance in meters, while y-axis representing the average semivariance of clay percentage.



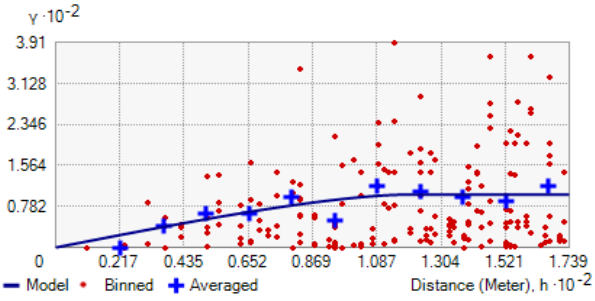
$$\gamma(h) = 0 + 100.59\left(\frac{3}{2}\frac{h}{(131.56)} + \frac{1}{2}\left(\frac{h}{131.56}\right)^3\right) \quad (13)$$

Fig. 10 - Spherical model of clay percentage for depth 2.5 to 5 meters.



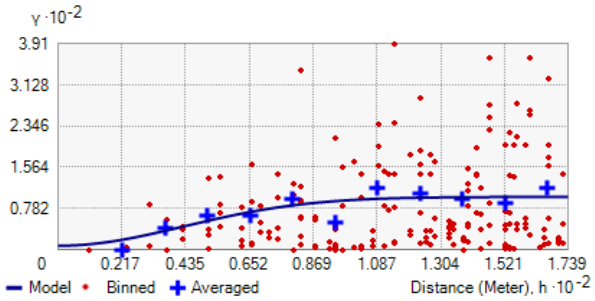
$$\gamma(h) = 0 + 107.4\left(1 - \exp\left(-\frac{h}{173.86}\right)\right) \quad (14)$$

Fig.11 - Exponential model of clay percentage for depth 2.5 to 5 meters.



$$\gamma(h) = 0 + 100.68\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{116.75}\right) + \sqrt{1 - \frac{h^2}{(116.75)^2}}\right) \quad (15)$$

Fig. 12 - Circular model of clay percentage for depth 2.5 to 5 meters.



$$\gamma(h) = 7.217 + 92.967\left(1 - \exp\left(-\frac{h^2}{(104.43)^2}\right)\right) \quad (16)$$

Fig.13 - Gaussian model of clay percentage for depth 2.5 to 5 meters.

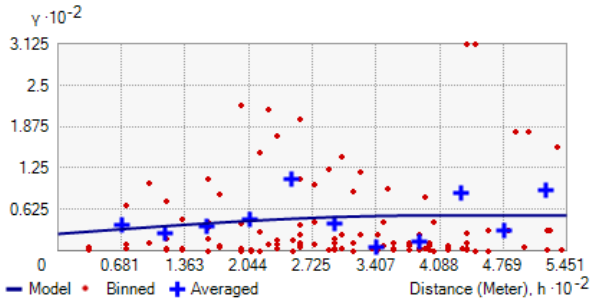
From Table 2, certain similarities could be observed, compared to the data from Table 1. The nugget values between spherical, exponential and circular model were similar, which was zero. Sill, range, and RMSE values were different for all models.

For this depth, spherical model was producing the lowest RMSE values, despite of shorter range values compared to the exponential model. The spherical model managed to reach more semivariogram cloud than any other model was due to the fact that the semivariogram cloud was located towards the beginning of the semivariogram, and met the conditions for the spherical model shape to reach most points.

Table 2 - Nugget, sill, range, and RMSE values for all semivariogram models for depth between 2.5 to 5 meters.

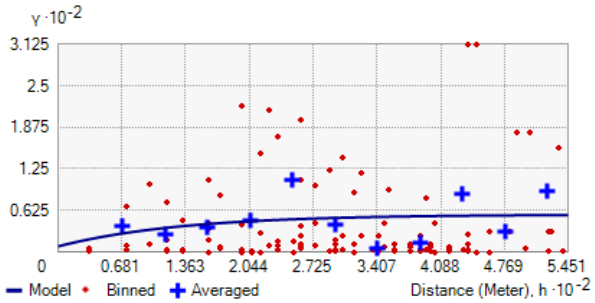
Semivariogram Model	Spherical	Exponential	Circular	Gaussian
Nugget (%)	0	0	0	7.217
Sill (%)	100.59	107.40	100.68	92.967
Range (m)	131.56	173.86	116.75	104.43
RMSE	9.067	9.1308	9.1005	9.6532

Fig. 14 to Fig. 17 are showing the semivariogram of clay percentage for the depth section of 5 to 7.5 meters. The numbers of boreholes involved are 11 boreholes, which explained these semivariograms look less crowded than previous two depth sections.



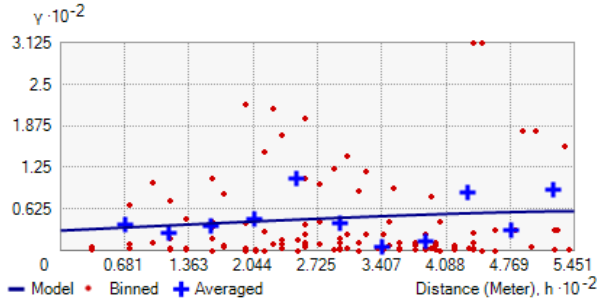
$$\gamma(h) = 25.443 + 27.802\left(\frac{3}{2}\frac{h}{(396.57)} + \frac{1}{2}\left(\frac{h}{396.57}\right)^3\right) \quad (17)$$

Fig. 14 - Spherical model of clay percentage for depth 5 to 7.5 meters.



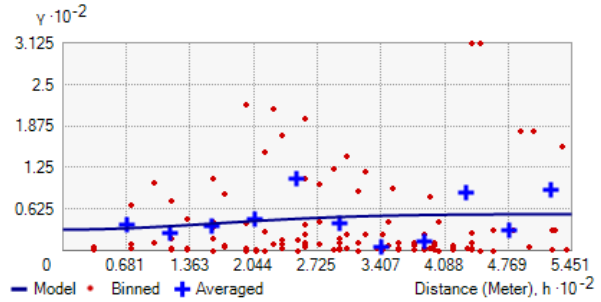
$$\gamma(h) = 8.4293 + 47.82\left(1 - \exp\left(-\frac{h}{396.57}\right)\right) \quad (18)$$

Fig. 15 - Exponential model of clay percentage for depth 5 to 7.5 meters



$$\gamma(h) = 29.97 + 29.235\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{545.07}\right) + \sqrt{1 - \frac{h^2}{(545.07)^2}}\right) \quad (19)$$

Fig. 16 - Circular model of clay percentage for depth 5 to 7.5 meters.



$$\gamma(h) = 31.635 + 23.294\left(1 - \exp\left(-\frac{h^2}{(396.57)^2}\right)\right) \quad (20)$$

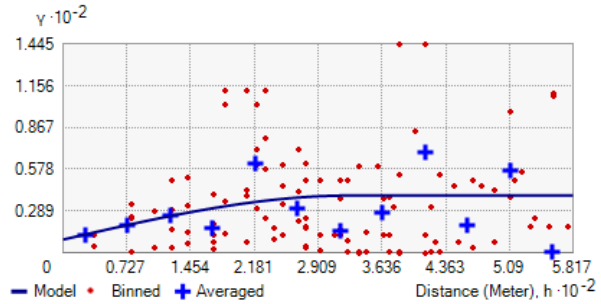
Fig. 17 - Gaussian model of clay percentage for depth 2.5 to 5 meters.

From Table 3, similarities in the range values could be observed for all models except circular, while nugget, sill, and RMSE values remain different for all models. For this depth section, circular model produced the lowest RMSE value, compared to other models. Circular model managed to achieve highest accuracy semivariogram due to the semivariogram cloud that located in the middle of the semivariogram, and allows circular model to reach most points towards the whole distance range.

Table 3 - Nugget, sill, range, and RMSE values for all semivariogram models for depth between 5 to 7.5 meters

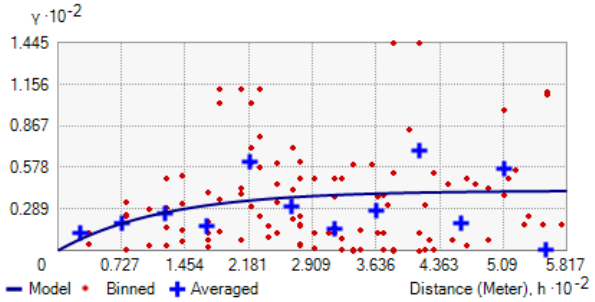
Semivariogram Model	Spherical	Exponential	Circular	Gaussian
Nugget (%)	25.443	8.4293	29.97	31.635
Sill (%)	27.802	47.82	29.235	23.294
Range (m)	396.57	396.57	545.07	396.57
RMSE	6.4369	6.5812	6.4173	6.4223

Fig. 18 to Fig. 21 are showing the semivariogram of clay percentage for the depth section of 5 to 7.5 meters. For this depth, the numbers of boreholes involved are 12 boreholes.



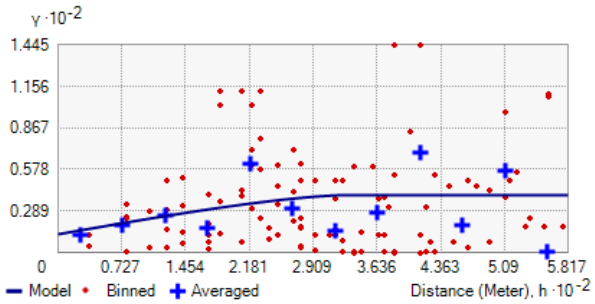
$$\gamma(h) = 3.7809 + 36.766\left(\frac{3}{2}\frac{h}{(327.75)} + \frac{1}{2}\left(\frac{h}{327.75}\right)^3\right) \quad (21)$$

Fig.18 - Spherical model of clay percentage for depth 7.5 to 10 meters.



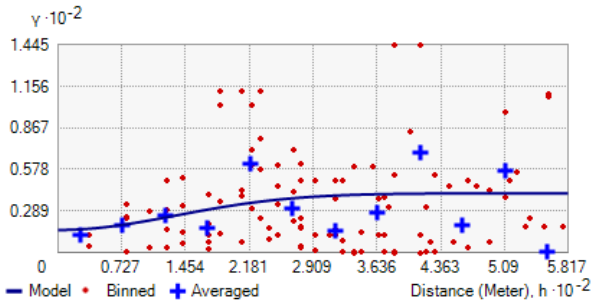
$$\gamma(h) = 0 + 46.536\left(1 - \exp\left(-\frac{h}{509.99}\right)\right) \quad (22)$$

Fig.19 - Exponential model of clay percentage for depth 7.5 to 10 meters



$$\gamma(h) = 7.7519 + 33.258\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{327.75}\right) + \sqrt{1 - \frac{h^2}{(327.75)^2}}\right) \quad (23)$$

Fig.20 - Circular model of clay percentage for depth 7.5 to 10 meters.



$$\gamma(h) = 13.026 + 28.996\left(1 - \exp\left(-\frac{h^2}{(327.75)^2}\right)\right) \quad (24)$$

Fig.21 - Gaussian model of clay percentage for depth 7.5 to 10 meters.

Table 4 is showing the similar range value for spherical, circular and Gaussian models. However, other values such as nugget, sill and RMSE were different for all models. For this depth section, exponential model managed to produce the lowest RMSE value, compared to other models. This condition is could be found quite similar to the data in the Table 1, whereby exponential model reached the longest range, and manage to reach most points in the semivariogram cloud due to the similar reason for the depth section of 0 to 2.5 meters.

Table 4 - Nugget, sill, range, and RMSE values for all semivariogram models for depth between 7.5 to 10 meters.

Semivariogram Model	Spherical	Exponential	Circular	Gaussian
Nugget (%)	3.7809	0	7.7519	13.026
Sill (%)	36.766	46.536	33.258	28.996
Range (m)	327.75	509.99	327.75	327.75
RMSE	5.0690	5.0340	5.3191	5.2533

Table 5 is showing the semivariogram model, which produced the lowest RMSE values for each depth sections, to be selected for the mapping purposes. For depth sections of 0 to 2.5 meters and 7.5 to 10 meters, the selected semivariogram model would be exponential model. For 2.5 to 5 meters and 5 to 7.5 meters depth sections, spherical and circular model were selected. And the mapping developed for each depth sections were shown in the Fig. 22 until Fig. 25. From Fig. 22 to Fig. 25, the observation could be made, whereby the map with higher number of data could expand towards all range displayed in the legend. However, the range also depends on the semivariogram shape. In all

depth sections, it could be seen that all semivariograms were showing excellent shape, according to the theoretical form [8].

With these maps, the location of the highest and lowest percentage of clay could be seen significantly for each depth sections. And the unknown values between known values could also be predicted, which could help the researchers to reduce the uncertainties of the unknown values in the known location.

Table 5 - Semivariogram models that will be used for mapping purposes.

Depth Sections (m)	Semivariogram Model
0 – 2.5	Exponential
2.5 – 5	Spherical
5 – 7.5	Circular
7.5 – 10	Exponential

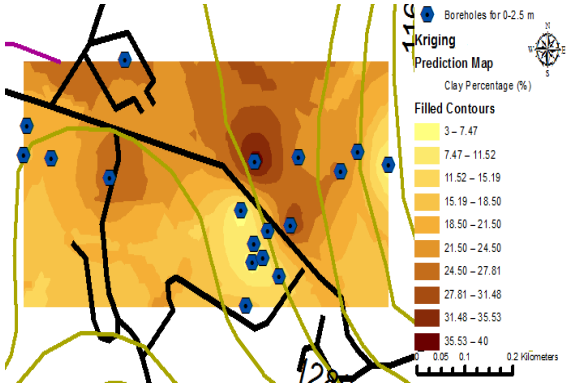


Fig. 22 - Clay percentage mapping for depth 0 to 2.5 meters.

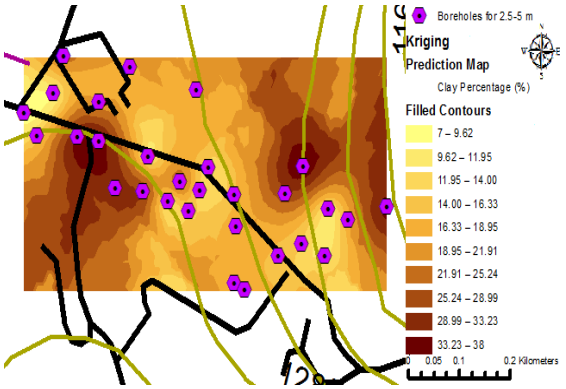


Fig. 23 - Clay percentage mapping for depth 2.5 to 5 meters.

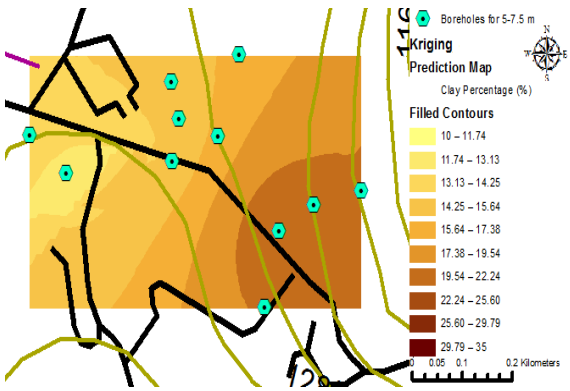


Fig. 24 - Clay percentage mapping for depth 5 to 7.5 meters.

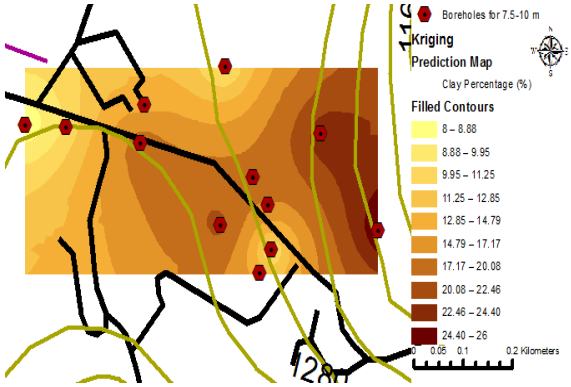


Fig. 25 - Clay percentage mapping for depth 7.5 to 10 meters.

4. Conclusion

OK and semivariogram models could produce some wonderful output in terms of mapping that related to the soil, should the opportunity rise and given. Given the digital satellite mapping and soil properties, OK managed to produce clay percentage distribution map at the Kundasang town, Sabah. With more soil properties or parameters, there are more possibilities to explore, and OK might be one of the methods that could help in geohazard mapping of this area. Through this study, OK showed the promising sign of mapping. This approach could help personnel that involved in the mapping study to expand the knowledge, whether researchers, design engineers, and constructors. Even the agriculturist could also apply this method to improvise their crop production.

References

[1] Brevik, Eric C., Hartemink, Alfred E. (2013). A brief history of soil mapping and classification in the USA. Soil Science Society of America Journal, pp. 1117-1132.
[2] Baum, Rex L., Godt, Jonathan W. (2009). Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides, pp.259-272.
[3] Chalkias, Christos., Ferentinou, Maria., Polykretis, Christos. (2014). GIS-based landslide susceptibility mapping on the Peloponnese Peninsula, Greece. Geosciences, pp. 176-190.
[4] Al-Ani, Haider., Oh, Erwin., Andargoli, Leila Eslami., Chai, Gary. (2013). Subsurface visualization of peat and soil by using GIS in Surfers Paradise, Southeast Queensland, Australia. Electronic Journal of Geotechnical Engineering, Research Article, pp. 1761-1774.
[5] Roslee, Rodeano., Tahir, Sanudin., Omang, Abd Kadir S. (2006). Engineering geological investigations on slope failure in the Kundasang Town, Sabah, Malaysia. Southeast Asian Natural Resources and Environmental Management, pp. 24-33.
[6] Minnit, R.C.A., Assibey-Bonsu, W. (2003). A tribute to Prof. D.G. Krige for his contributions over a period of more than half a century. South African Institue of Mining and Metallurgy, Keynote Address.
[7] Oliver, M.A., Webster, R. Kriging: (1990). A method of interpolation for Geographical Information System. International Journal of Geographical Information System, pp. 313-332.
[8] Cressie, Noel. (1990). The Origins of Kriging. Mathematical Geology, pp. 239-252.