Ordinary Kriging base on OLS-WLS Fitting Semivariogram: Case of Gold Vein Precipitation

Nur Ali Amri[a], Abdul Aziz Jemain[b] and Wan Fuad Wan Hassan[c]

[a]*Mining Engineering Department, Faculty of Mineral and Technology, UPN "Veteran" Yogyakarta, 55283 SWK 104 (North Ring Road) Condong Catur, Yogyakarta, Indonesia*
[a,b]*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia*
[c]*Geological Department, University of Malaya, 50603 Kuala Lumpur, Malaysia*

Abstract. The comparison of OLS fitting based on exponential semivariogram models, where the range is 811.76 meters, sill is 37.23 $(g/t)^2$ and the spherical model base, which the sill is 38.44 $(g/t)^2$ and range is 921.87 meters, used as a basis for OK prediction. Based on these two parameters, sill and range, obtained that result of kriging variances are almost the same, namely 4.96 g/t for the exponential models, and 4.90 g/t for the spherical models.

Keywords: Semivariogram, OLS and WLS fitting, Ordinary Kriging.

INTRODUCTION

In the gold mining region, something difficult to be achieved which is to find the vein gold deposits should be exploited immediately. The main consideration is the economical factor, since (and this is the general character of the mining industry) that the primary gold is almost always located in remote areas, even in places that are not close to the earth's surface. Therefore, we need sampling, (e.g., by drilling at certain points) to obtain presence information. Samples are expected to be important information for observation points around the site, which was not drilled. Here, then predictions should be implemented.

Geostatistic is a method that includes semivariogram and kriging. Semivariogram is a process that must be executed to obtain the semivariogram, which is the function describes the correlation between the values (or sample points) spatially separated. While kriging is a weighted method, which is used to predict the values ([1, 2, 3], and other authors include [4, 5], using a term estimates) at an unknown location by considering the points that have been known to them. As already known the spatial correlation (refere to the values that have been known about), the values in a location that has not been sampled will be predict. This set of values, which became the basis for the determination of reserves.

Reserve calculations in this paper are based on kriging predictions. While the results of kriging predictions are influenced, especially by the sill and range, which is a parameter obtained by fitting the experimental semivariogram (as a discrete function) to the theoretical semivariogram (continuous function). Sill and range, both of which are important parameters to be used as a stepping-stone when it will perform kriging predictions.

Fitting semivariogram, in this paper using two approaches, namely ordinary least squares (OLS) and weighted least squares (WLS), which is chosen semivariogram models with transition, i.e., spherical and exponential models. Spherical model, selected based on consideration of reliability, and is often used in the study of earth science [6]. While the exponential models used as a comparison. Moderate parameters of comparative results have been selected as the basis for the calculation of kriging. Meaning, that the sill value is not too large and the resulting range is not too long.

There are various types of kriging (disjunctive kriging, indicator kriging, multiple indicator kriging, etc.). But in this paper the prediction is using Ordinary Kriging (OK). Therefore the distribution of the data (both physically and disparity of grades) is uneven, then the variography based on robust semivariogram method [7]. Non-performance of the logarithm of the data (although the data are highly skewed), more considerate to determine the effect on the distribution of prediction results, which will be used as a reference to determine the priority of mine exploitation.

Proceedings of the 3rd International Conference on Mathematical Sciences
AIP Conf. Proc. 1602, 1039-1045 (2014); doi: 10.1063/1.4882612

MATERIALS AND METHODS

The research material includes of sampling data from exploration drilling results which is done by the company. Geostatistics, here is the method used for handling the data, starting from semivariogram modeling to fitting of theoritical semivariogram model. While kriging is used to predict the data points which are not sampled.

Data and Location

This paper examines the calculation of gold deposits obtained from the weighted averaging of 138 quartz vein drill samples, in Ciurug area have been assayed. Ciurug is an area owned of PT. Aneka Tambang, UBPE Pongkor, Indonesia which is at elevation from 1,110 m to 1,250 m above sea level, and administratively, included the sub-region of Bayah Lebak. Geographically, the area is bounded by lattitude of $106^0$24'00"E - $106^0$26'00"E and longitude $06^0$44'00"S - $06^0$46'00"S. Tectonically, this area is a part of the Indonesian Tertiary magmatic arc, namely the Sunda-Banda arc of Late Miocene – Pliocene age. Regionally, the area was included within the framework of regional geology of Bayah dome, South Banten [14].

High grade gold which is relatively homogeneous, located on the left (or west side) of research area. While residing on the east side, is a group of low-yield spread (also relatively homogeneous).

Experimental Semivariogram

Semivariogram is one step which is performed after statistical analysis of the behavior of data distribution. Normal distribution often used as an option before semivariogram calculation. In a particular case, the omission of the data distribution (without pruning the outliers, for example), can help to understand the possibility of spatial semivariogram behavior anomaly, after reaching sill on the practical range, for example changing in the behavior of rock structures, and also on other geological phenomena [8, 9].

Theoretically, suppose grade of an ore body at point **s** (here $\in \mathrm{R}^2$) was observed at certain points of the $\mathbf{s}_i$ (i=1, ..., n) is a realization of a random process {Z(**s**): $\mathbf{s} \in D$} of an ore body, while the intrinsic stationarity is determined through a first diferences process of $E(Z(\mathbf{s}+\mathbf{h})-Z(\mathbf{s}))=0,$ then the intrinsic stationarity is defined as [1],

$$\mathrm{var}(Z(\mathbf{s}+\mathbf{h})-Z(\mathbf{s}))=2\gamma(\mathbf{h})=E[\{Z(\mathbf{s}+\mathbf{h})-Z(\mathbf{s})\}^2].$$

Thus the semivariogram estimator is,

$$\hat{\gamma}(\mathbf{h})=\tfrac{1}{2}\,mean[\{Z(\mathbf{s}+\mathbf{h})-Z(\mathbf{s})\}^2].$$

If Z($\mathbf{s}_i$) is the value of a variable (say, the gold grade) located in $\mathbf{s}_i$, Z($\mathbf{s}_j$) is the value of variable (also grade gold) which is in $\mathbf{s}_j$, or located at a distance **h** from Z($\mathbf{s}_i$), while N(**h**) is the number of pairs of observation points at lag **h**, the experimental semivariogram estimator (known as classical semivariogram) is formulated as a function of [10]

$$\hat{\gamma}(\mathbf{h})=\frac{1}{2|N(\mathbf{h})|}\sum_{i=1}^{N(\mathbf{h})}[Z(\mathbf{s}_i)-Z(\mathbf{s}_j)]^2$$

Functions which is above, assessed by [11] as susceptible to the emergence of atypical data observations, so by transforming Box & Cox (1964) and assumes a principle to normality, he then fine-tune by presenting a more muscular semivariogram estimator, i.e.,

$$\bar{\gamma}(\mathbf{h})=\left(\frac{1}{2|N(\mathbf{h})|}\sum_{i=1}^{N(\mathbf{h})}[\mathrm{Z}(\mathbf{s}_i)-\mathrm{Z}(\mathbf{s}_j)]^{1/2}\right)^4\Big/\left(0.457+\frac{0.494}{|N(\mathbf{h})|}\right).$$

Fitting Semivariogram

Fitting experimental semivariogram with reference to the theoretical semivariogram is to obtain a trend line. The objective of this fitting is acquisition of sill and range (or other). Various models of the fitting can be carried out, including the eye observations. But this paper presents the automatic fitting with the two least squares methods, namely OLS and WLS where the objective were compared to obtain the most moderate model. In the least squares fitting, the estimated semivariogram constructed by minimizing the sum squared difference of [1, 12]

$$R(\theta)=\sum_{i=1}^{k} w_i^2[\hat{\gamma}_z(\mathbf{h}_i)-\gamma_z(\mathbf{h}_i;\theta)]^2 \text{ namely,}$$

$$\min \sum_{i=1}^{k} w_i^2[\hat{\gamma}_z(\mathbf{h}_i)-\gamma_z(\mathbf{h}_i;\theta)]^2.$$

Therefore, the weight OLS is $\sum_{i=1}^{k} w_i^2 = 1$ then the OLS fitted by taking

$$\min \sum_{i=1}^{k} [\hat{\gamma}_z(\mathbf{h}_i)-\gamma_z(\mathbf{h}_i;\theta)]^2.$$

The WLS weight is $\sum_{i=1}^{k} w_i^2 = 1/\mathrm{var}[\hat{\gamma}_z(\mathbf{h}_i)]$ then the fitting is [1, 12]

$$\min \frac{1}{2}\sum_{i=1}^{k} N(\mathbf{h}_i)\left[\frac{\hat{\gamma}_z(\mathbf{h}_i)}{\gamma_z(\mathbf{h}_i;\theta)}-1\right]^2$$

Fitting with the various stages yield a range (especially practical range) and sill. Semivariogram said to achieve of sill γ_∞, if [1, 13] $\lim_{|\mathbf{h}|\to\infty} \gamma(|\mathbf{h}|)=\gamma_\infty<\infty$.

It is important to be noted that there is a difference between factor range and practical range (or the range of spatial dependence), which in turn, in the paper is simplified as a range under the symbol a. Range (for exponential) of the lag $\mathbf{h}$, where $\gamma(\mathbf{h})=0.95\gamma(\infty)$ is the distance at which the semivariogram close to 95% of the sill, while the range factor is a condition before reaching the practical range [2].

Theoretical Semivariogram

The semivariogram theoretical (more noticeable in the visual graph), in general will behave on the increase, ranging from the nugget up to a maximum distance of influence or range (a). After that, because of the variability among the points which no longer has a spatial correlation, then the semivariogram curve will be constant. Various types of theoretical semivariogram presented, but often used in mining geology, as general is a transitional form that has [1],

a. Exponential model, $\gamma(\mathbf{h})=C_0+C\{1-\exp(-|\mathbf{h}|/a)\}$

b. Spherical model $\gamma(\mathbf{h})=\begin{cases} C_0+C[(3/2)(|\mathbf{h}|/a)-(1/2)(|\mathbf{h}|/a)^3], & 0<|\mathbf{h}|\le a \\ C_0+C, & |\mathbf{h}|\ne a. \end{cases}$

C_0 is nugget effect, C is sill and a is range.

Both theoretical semivariograms are selected for fitting the experimental semivariogram, i.e. using OLS and WLS.

Ordinary Kriging

Kriging, essentially is a weighted averages method of points on a location. This weight is directly related to the semivariogram. In ordinary kriging (OK), the predictive value at unknown point can be observed with similar data in other locations. This prediction was made by modeling the stationary random function of some random variable, namely $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$.

Suppose a value to be predicted at the point $\mathbf{s}_0$ is $\hat{Z}(\mathbf{s}_0)$ in which each random variable has the same probability distribution at all locations, while the expectation value of a random variable is $E(Z)$, then [1]

$$\hat{Z}(\mathbf{s}_0)=\sum_{i=1}^{n} w_i Z(\mathbf{s}_i).$$

$\hat{Z}(\mathbf{s}_0)$ is the predicted value of $Z(\mathbf{s}_i)$ which is based on data observation $z_1, \ldots, z_n$, while w_i ($i = 1, \ldots, n$) is the weight of sample points. Prediction error at the point $\mathbf{s}_0$ obtained by

$$R(\mathbf{s}_0) = \hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)$$

$$E(R(\mathbf{s}_0)) = E(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)) = E\left(\sum_{i=1}^{n} w_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0)\right) = \sum_{i=1}^{n} w_i E\left(Z(\mathbf{s}_i)\right) - E\left(Z(\mathbf{s}_0)\right).$$

Because of Z is stationary, then $E(R(\mathbf{s}_0)) = \sum_{i=1}^{n} w_i E(Z) - E(Z)$. Term of biasness can be achieved if $E(R(\mathbf{s}_0)) = 0$, so $\sum_{i=1}^{n} w_i = 1$.

The variance error is

$$\widetilde{\sigma}_R^2 = \widetilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \widetilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \widetilde{C}_{i0}.$$

So that $\widetilde{\sigma}_R^2$ minimum, the equation must be deducted by Lagrange multiplier, λ and

$$\widetilde{\sigma}_R^2 = \widetilde{\sigma}^2 + \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \widetilde{C}_{ij} - 2\sum_{i=1}^{n} w_i \widetilde{C}_{i0} - 2\lambda\left(\sum_{i=1}^{n} w_i - 1\right)$$

Minimum error of variance achieved under conditions where the first partial derivatives of the weights equated to zero,

$$\frac{\partial \widetilde{\sigma}_R^2}{\partial w_i} = 0 \quad \text{so that} \quad \sum_{j=1}^{n} w_j \widetilde{C}_{ij} + \lambda = \widetilde{C}_{i0}, \ \forall i = 1,\ldots,n.$$

Boundary condition, generated by searching a first partial derivative of the equation with a Lagrange multiplier and equating to zero

$$\frac{\partial \widetilde{\sigma}_R^2}{\partial \lambda} = 0 \ \text{then} \ \sum_{i=1}^{n} w_i = 1.$$

Equation system of OK can be presented in matrix form as follows,

$$\underbrace{\begin{bmatrix} \widetilde{C}_{11} & .. & \widetilde{C}_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \widetilde{C}_{n1} & .. & \widetilde{C}_{nn} & 1 \\ 1 & .. & 1 & 0 \end{bmatrix}}_{(n+1)\mathsf{x}(n+1)} \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \lambda \end{bmatrix}}_{(n+1)\mathsf{x}1} = \underbrace{\begin{bmatrix} \widetilde{C}_{10} \\ \vdots \\ \widetilde{C}_{n0} \\ 1 \end{bmatrix}}_{(n+1)\mathsf{x}1}$$

Within a vector form can be written as follows, $C\vec{W} = \vec{D}$. C is the covariance matrix of the observed data on the location, and usually a positive definite matrix. To gain weight $\vec{W}$, on the left-hand side of the equation is multiplied by the matrix C^{-1}, so that $\vec{W} = C^{-1}\vec{D}$.

DISCUSSION

The discussion includes the execution of semivariogram (experimental and theoretical) to ordinary kriging predictions which are based on a grid block size of 5×5. Fitting comparison is done using OLS and WLS method, based on the exponential and spherical models.

Semivariogram

Semivariogram is a function (and tool) that are used to describe or determine the variability (level of similarity) of regionalized data separated by **h**, up to a certain distance (called, the range of influence) in which the data is no longer spatially correlated. Although after that, anomaly events can occur where semivariogram fluctuated, (can continue ascending, or even a drastic decline).

Semivariogram calculation, in this case, based on the assumption of omnidirectional, meaning, that the semivariogram being equal, for a variety of directions. Each lag distance is 100 meters, which is the maximum distance of 1500 meters (longitude axis is East-West). Longitude starts from 7435.47 to 9069.75, while the latitude is 343.55 to 752.08, and the whole area is 545.606 m2. Execution of the data calculation (i.e. semivariogram and kriging) are using the open source program, R package [15].

Although the distribution is skewed (skewness=2.499) so that its distribution is still awake, then the experimental semivariogram calculation is using robust models (data need not be transferred to a logarithmic form). **TABLE 1** presents the values of the semivariogram lag distances (h), regularly every 100 m with abscissa between 7500 to 9000. Therefore, there are 138 data, so, in the semivariogram calculation, there are $n \times (n-1)/2$ or $(138 \times 137)/2 = 9543$ unique pairs of observations.

Maximum pair of points (named, pairs) is 1589, while the minimum is 2. Disparities (colored very sharp) occurs due to the left (Southwest, relatively) is a rich zone (≥15 g/t). More towards the east, where the number of samples to be less and less, grade is being smaller, and this is called as a poor zone (≤1.5 g/t).

In general, it appears that the trip of robust semivariogram (or, modulus) is fluctuating, though at a certain lag, it was tended to increase. Rise significantly, occurs on the ninth lag, which is at a distance of 900 where the increase is almost double from 36.83 towards 63.73, both in (g/t) Au^2. After that, is decreased again until the eleventh lag.

TABLE (1). Omnidirectional Robust Semivariogram

Lag Distance (m)	Semivariogram*)	Pairs
100	17.51	1589
200	21.48	1540
300	19.88	1268
400	21.06	1114
500	29.51	981
600	30.62	840
700	23.88	629
800	36.83	514
900	63.73	370
1000	51.17	280
1100	44.17	170
1200	55.83	94
1300	44.26	39
1400	15.87	23
1500	1.51	2

*)in $(g/t)^2$ Au.

Fitting Semivariogram

Fitting the experimental semivariogram, especially performed to obtain sill and range (*a*), which on inquiry was conducted against the exponential and spherical models. Based on **TABLE 2**, robust semivariogram fitting OLS models yield a smaller sill (or might say better) than the WLS. Distance between the influence of sample points is also larger (or longer) than the range of the sample points on the overall classical semivariogram.

TABLE (2). Robust Semivariogram Parameters

	OLS	WLS
Exponential		
Sill $(g/t)^2$	37.23	50.72
Range (m)	811.76	1608.82
Spherical		
Sill $(g/t)^2$	38.44	47.80
Range (m)	921.87	1139.96

The range with WLS fitting for exponential models is 1608.82 m, almost twice of range using OLS, which is 811.76 m (797 m differences). This distance is too far from the maximum distance (1500 m). In the spherical case, the difference between the OLS model of sill and WLS is only 9.36 $(g/t)^2$, whereas the difference in range is 218 m.

Based on these results, obtained the general form of the two models OLS fitting estimation of semivariogram basis, which exponential models is

$$\hat{\gamma}(\mathbf{h}) = 37.23\{1 - \exp(-|\mathbf{h}|/811.76)\}.$$

As for the spherical models is

$$\hat{\gamma}(\mathbf{h}) = \begin{cases} 38.44[(1.5)(|\mathbf{h}|/921.87) - (0.5)(|\mathbf{h}|/921.87)^3], & 0 < |\mathbf{h}| \leq 921.87 \\ 38.44, & |\mathbf{h}| \neq 921.87. \end{cases}$$

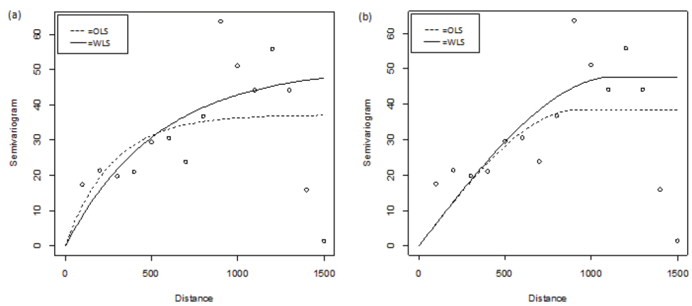

FIGURE 1. Fitting base on OLS and WLS Robust Semivariogram of (a) Exponential; and (b) Spherical

Ordinary Kriging

Journal (1983) states, that the effectiveness indication of the spatial data completion using kriging method can be carried out by looking at two things, namely the skewness and the coefficient of variation (CV). The data, which have a very uneven distribution or skewed (where the skewness is greater than 1.5), more suited solved by kriging method. More specifically, linear kriging will be effective to solve the data that has CV is less than one. In this case, the value of CV is 0.883 (CV<1), while the skewness is 2.499 (>1.5). Meaning, these data will be effectively solved by the kriging method, in particularly is a linear kriging (OK).

The prediction of OK performed on the area of 375,000 m2, where the size of each block is (5×5) m^2, using two models, exponential and spherical. Using OLS fitting exponential models, the value of sill (C) is 37.23 (g/t)^2, at a distance of influence (a) along the 811.76 m. Whereas at the spherical models, sill value (C) is 38.44 (g/t)^2 which the distance between the influence of spatial points (a) is 921.87 m. Based on **FIGURE 2**, it appears that the rich zone (≥15 g/t) spreads near longitude of 7650-7850, with a yellow to green colored (bright), is in a position of relatively to the Southwest area of about 2,500 m^2. Poor zone almost spread in most of the middle to eastern sampling locations, light blue to purple (dark).

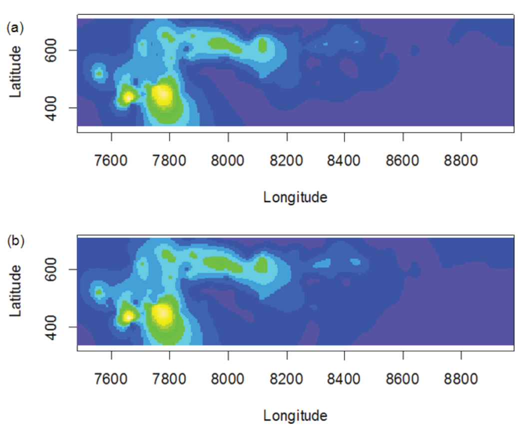

FIGURE 2. Ordinary Kriging Estimation base on Fitting OLS model (a) Exponential (C=37.23 & a=811.76) dan; (b) Spherical (C=38.44 & a=921.87)

Fitting of spherical semivariogram (**FIGURE 2 b**), produces the distribution of grade, medium to high which is slightly wider than the exponential models. In contrast, the exponential models produces of low grade distribution (purple color) and it is wider, almost breaking into high grade (**FIGURE 2 a**). Mean prediction, both as **TABLE 3** is almost same, namely 4.96 g/t for exponential and 4.90 g/t for spherical. Nonetheless, the OK standard deviation (square of variance) of the prediction results (exponential models) is smaller (ie, 1.75 g/t) than the OK standard deviation of spherical models, namely 2.08 g/t.

TABLE (3). Ordinary Kriging Parameters (g/t)

Kriging Parameters	Exponential	Spherical
Mean Estimation	4.96	4.90
Median Estimation	3.29	3.30
3^{rd} Quantile Estimation	6.83	6.82
Kriging Std. Dev.	1.75	2.08

CONCLUSION

Based on the calculation of robust experimental semivariogram (both theoretical semivariogram fitting exponential models and models of spherical) obtained information that OLS fitting yields range is smaller than the WLS fitting. Even on exponential models, WLS fitting yields range of 1608 meters, and is beyond of maximum distance (i.e., 1500 m). Meaning, spatial effect between sample points is much greater than half of the longitude length, which is a gathering sites of grade rich.

OK prediction produces mean, median and 3rd quantile relatively the same, although the kriging variance (expressed in standard deviation) are slightly different. Predictions also show, that the rich zone located in the Southwest region, relatively.

The information of data distributions in the mining industry, serve as a very useful reference in calculating of reserves (or resources), and also, in determining the priority scale of mining operation (i.e., in mining exploitation).

ACKNOWLEDGEMENTS

Thanks to PT. Aneka Tambang UBPE Pongkor management, especially to Ir. Rustaman, Ir. Joseph and Ir. Kristiawan who has provided facilitation to the author to obtain the data as a part of this study.

REFERENCES

1. N. Cressie, *Statistics for Spatial Data*, New York: John Wiley & Sons, Inc., 1993.
2. T. Hengl, *A Practical Guide to Geostatistical Mapping of Environmental Variables*, Luxembourg: Office for Official Publications of the European Communities, 2007.
3. A. G. Journel and Ch. J. Huijbregts, *Mining Geostatistics*, London: Academic Press, 1974, Third Printing, 1984.
4. D. D. Sarma, *Geostatistics with Applications in Earth Sciences*, Dordrecht, Netherlands: Spriger, Second edition, 2009.
5. M. Armstrong, *Basic Linear Geostatistics*, Varlag Berlin Heidelberg New York: Springer, 1998.
6. N. Cressie and D. M. Hawkins, "Robust Estimation of the Variogram", *Journal of the International Association for Mathematical Geology* **12**(2), 1988, pp. 115-125.
7. M. Bachmaier and M. Backes, *Variogram or Semivariogram? Understanding the Variances in a Variogram*, New York: Springer Series in Statistics, Springer Science + Business Media, LLC. 9, 2008, pp. 173-175.
8. J. Coombes, "Mentoring in Mining: A Handbook for Geoscientists Managing Teams", *Coombes Capability*, 2009, 80 pages.
9. N.A. Amri, Abdul Aziz J., Wan Fuad Wan Hassan and Jatmika S., "*Geostatistical Analysis on Distribution of Gold Veins*" in Proceeding of 1st Earth Science International Seminar, Yogyakarta, 2012, pp. 19-27.
10. G. Matheron, "Principles of Geostatistics", *Journal of Economic Geology* **58**, 1963, pp. 1246-1266.
11. N. Cressie and D. M. Hawkins, "Robust Estimation of the Variogram", *Journal of International Association for Mathematical Geology* **12**(2), 1980, pp. 115-125.
12. S. N. Lahiri, Y. Lee and N. Cressie, "On Asymptotic Distribution and Asymptotic Efficiency of Least Squares Estimators of Spatial Variogram Parameters", *Journal of Statistical Planning and Inference* **103**, 2002, pp. 65–85.
13. L. Budrikaitė, K. Dučinskas, "Forms of Anisotropy for Spatial Variograms", *Journal of Liet. matem. rink.* **44**, 2004 pp. 542–546.
14. A. Basuki, D.A. Sumanagara and D. Sinambela, "The Gunung Pongkor Gold-silver Deposit, West Java, Indonesia", *Journal of Geochemical Exploration* **50**, 1994, pp. 371 – 391.
15. R.S. Bivand, E.J. Pebesma and V.G. Rubio, *Applied Spatial Data Analysis with R*, New York: Springer Science+Business Media, LLC, 2008.