ARPN Journal of Engineering and Applied Sciences

CONSISTENCY OF THE SEMIVARIOGRAM-FITTING ERROR ON ORDINARY KRIGING PREDICTION

Nur Ali Amri[1,2], Abdul Aziz Jemain[2] and Ahmad Fudholi[3]
[1]Mining Engineering Department, Faculty of Mineral and Technology, UPN "Veteran" SWK Condong Catur, Yogyakarta, Indonesia
[2]School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia Bangi, Selangor, Malaysia
[3]Solar Energy Research Institute, Universiti Kebangsaan Malaysia, Bangi Selangor, Malaysia
E-Mail: a.fudholi@gmail.com

ABSTRACT

Kriging is a construction method that is primarily built based on the structure of experimental semivariograms and the power of fitting. The two functions, i.e., classical and robust semivariograms, are used in the current study. The emivariograms are fitted using two approaches, namely, ordinary least squares and weighted least squares, whereas the spherical and exponential functions are utilized for the theoretical model. The estimation precision is calculated using the root mean square error. The error use of the root mean squares for predictions was tested using the mean absolute deviation.

Keywords: semivariogram, OLS and WLS fitting, ordinary kriging.

INTRODUCTION

Gold mining is an activity that can provide economical returns based on the amount of ore produced. Aside from promoting the macro-economic growth, this activity provides employment to people. However, mining plans that are improperly designed will adversely affect the natural environment and result in changes in the landscape.

Initial studies on mining plans must be conducted to prevent the aforementioned adverse effects. Detailed and comprehensive studies must be carefully conducted. One aspect to consider is the amount of ore reserves that can be mined and the mineral distribution in an area. Thus, geological explorations require random samples at various points to conduct thorough analyses.

Before conducting research, information about an area that may contain minerals must be obtained. The current study was conducted in Ciurug, Pongkor Indonesia, which belongs to PT. Aneka Tambang UBPE Pongkor is an area with a vein system. The existence of veins in this area was caused by the subduction of the Ocean Indo-Australian plate under the Eurasian Plate [1]. A vein is an area where ore minerals, which are naturally distributed in any location of deposition, settle in a rock fracture [2]. 128 data samples were randomly obtained in this area to predict the quantity of ore reserves. The information of the sample data was obtained by using assays from core-bore drilling.

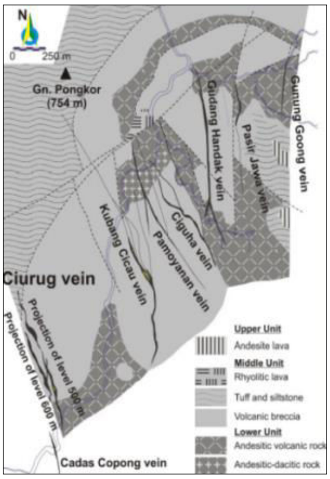

Figure-1. Geological map and vein system of Pongkor [1].

MATERIAL AND METHOD

Primary gold is a mineral that is found throughout the earth. Locations where ore gold settles in rock fractures or soils are known as gold mineralization. The mineralization process begins from a magmatic solution, which occurs because of the impulse of heat energy that breaks through a host rock and emerges to the surface of the earth through rock fractures. The solution is gradually frozen and settles [2, 3]. This process causes a link to exist between minerals in one place to those in other locations. This linkage is an event in geostatistics called, "regionalization." Therefore, the geostatistics method is used to obtain information about its content.

Geostatistics involves a statistical–mathematical modeling of trends based on regionalized variables. Geostatistics is a hybrid discipline from engineering mining, geology, mathematics, and statistics. This model is widely used to examine gold mining precipitation in veins [4]. This model is utilized to build semivariograms and kriging. Semivariograms are the main structure underlying kriging prediction. The experimental semivariogram comprises functions, generally in discrete forms, that are used to determine the behavior of spatial data with the precision of the parameters of the fitting result parameters. Nugget, sill, and range are the three parameters generated by fitting the semivariogram function. An estimation of the experimental semivariogram indicates the sample of the original data (original data) [5, 6].

Two experimental functions, namely, the models of Matheron [7] and the Cressie and Hawkins model [8], were used in the estimation of semivariograms. The fitting of semivariograms is calculated using spherical and exponential functions. However, weighted least squares (WLS) and ordinary least squares (OLS) are used as a model approach. Semivariograms are estimated and predicted in kriging execution using the open sources R package [9].

Semivariogram

The locus of drill holes points in the region is signified by $\mathcal{D}$. Grade quantity at point $\mathbf{s}_i$, which is randomly obtained, is denoted by $Z(\mathbf{s}_i)$, i=1, 2, …, n. Suppose that $Z(\mathbf{s}_i)$ is omnidirectional and stationary [6], $E[Z(\mathbf{s}_i)] = \mu$ and $var[Z(\mathbf{s}_i)] = \sigma^2$ for all i = 1, 2, ..., n. Consider the value of minerals in any two locations, namely, $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_j)$. The expectation is the squared difference $E[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 = 2\gamma(|\mathbf{s}_i - \mathbf{s}_j|)$. The function $\gamma(|\mathbf{s}_i - \mathbf{s}_j|)$, which is known as a semivariogram, is identical with

$$\frac{1}{2}E[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]^2 = \frac{1}{2}\mathrm{var}(Z(\mathbf{s}_i) - Z(\mathbf{s}_j)) = \gamma(|\mathbf{s}_i - \mathbf{s}_j|)$$

The expected value of Z ($\mathbf{s}_i$) is zero. Thus, $E[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)]$ is also equal to zero, and the estimated semivariogram ultimately depends on $E\left[Z(\mathbf{s}_i) - Z(\mathbf{s}_j)^2\right]$. Matheron [7] presents the equation known as the classical model, which is shown below:

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

The semivariogram function by Cressie and Hawkins [8], which is known as the robust semivariogram function, is

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

Fitting semivariogram

The two mathematical equations that are represented as semivariogram lines in the current study are the spherical and exponential models. The spherical model can be given as [6, 10]

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h}=0 \\ C_0 + C_1\left[\frac{3}{2}\left(\frac{|\mathbf{h}|}{a}\right) - \frac{1}{2}\left(\frac{|\mathbf{h}|}{a}\right)^3\right], & 0<\mathbf{h}\leq a \\ C_0 + C_1, & |\mathbf{h}|>0. \end{cases} \quad (3)$$

The mathematical expression for the exponential function is as follows:

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \mathbf{h}=0 \\ C_0 + C\left[1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right)\right], & \mathbf{h}\neq 0. \end{cases} \quad (4)$$

Two rules are used to build three semivariogram parameters, i.e., nugget, sill, and range. Let $\hat{\gamma}(\mathbf{h}_j)$ be an estimate semivariogram in distance $\mathbf{h}_j$, $\gamma(\mathbf{h}_j;\theta)$ be the value of the semivariogram line that is equal with the distance, and parameter θ be the value to be estimated. Cressie [6] provides the minimum of OLS method in generating parameterθ.

$$\theta = \text{minimum}\sum_{j=1}^{J}\left[\hat{\gamma}(\mathbf{h}_j) - \gamma(\mathbf{h}_j;\theta)\right]^2. \quad (5)$$

The second approach is WLS method where $N(\mathbf{h}_j)$ is the number of pair points in distance $\mathbf{h}_j$. This rule was chosen because the weight values for all couples of grade points at distance $\mathbf{h}_j$, which is as defined as

$$\theta = \text{minimum}\sum_{j=1}^{J}\left(\frac{\hat{\gamma}(\mathbf{h}_j)}{\gamma(\mathbf{h}_j;\theta)} - 1\right)^2 N(\mathbf{h}_j). \quad (6)$$

Ordinary kriging

Various kriging methods were developed, but ordinary kriging was used in the current study. Kriging is a method of extrapolating data, which is believed to provide the best linear unbiased estimator and generate good minimum variance estimators. However, kringing only applies when environmental studies [11–13], particularly the ore or gold grade precipitates in veins, were chosen correctly [14, 15]. Let $Z(\mathbf{s}_i) \in \mathcal{D}$ (i=1, …, n) be n gold grade numbers of a value that exist in several locations of $\mathbf{s}_i$ (values $\mathbf{s}_i \in \mathcal{D}$ in dimension 2). Points have properties as a regionalized variable. Thus, these points can be used as a basis to predict the $Z(\mathbf{s}_0)$ value situated in an un-sampled point $\mathbf{s}_0$ ($\mathbf{s}_0 \in \mathcal{D}$). If the predicted value is depicted as the objective function, $\hat{Z}(\mathbf{s}_0)$ [7]; this function can be written as

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} w_i Z(\mathbf{s}_i). \quad (7)$$

Variance as the minimum condition is

$$\sigma^2(\mathbf{s}_0) = E\left[Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\right]^2. \quad (8)$$

Eight parameters are used as the basic components of a prediction model, i.e., four models employ a classical approach, and the other four models use a robust approach. Models 1 and 2 are based on classical experiments where the theoretical semivariogram is a spherical function. These models employ both the WLS and OLS approaches. Models 3 and 4 are based on classical experiments modeled by exponential function

ARPN Journal of Engineering and Applied Sciences

theory. These models follow the WLS and OLS approaches. Models 5, 6, 7, and 8 are all based on robust experimental semivariograms. Models 5 and 6 are based on spherical function theory, whereas Models 7 and 8 are based on the theoretical function of exponential semivariograms. These models employ both the WLS and OLS approaches.

RESULTS AND DISCUSSIONS

The experimental semivariogram values generated using Equations (1) and (2) are presented in Table-1. At least 69 pair points are valid for the first lag with a distance of 17.306. Classical semivariograms produce an initial semivariogram value greater than that of robust semivariograms.

Table-1. Classical and robust semivariogram estimation at **h** distance.

Distance, h	***np***	$\hat{\gamma}$(h)	$\bar{\gamma}$(h)
17.298	68	4.617	2.529
51.894	289	7.340	7.085
86.490	385	8.806	8.203
121.086	417	10.460	10.012
155.683	451	10.741	10.227
190.279	491	11.737	10.970
224.876	437	11.733	10.722
259.472	417	11.990	12.233
294.068	358	11.720	12.264
328.664	375	10.724	9.654
363.260	347	11.132	10.037
397.857	343	12.389	12.581
432.453	318	10.921	11.147

The equation model for continuous line of the semivariogram function is based on Equations (3)–(6) for each model, as shown in the second column of Table-2. The minimum number of pair points is 68, which meets the recommended number suggested by Journel and Huijbregts [5]. The theoretical semivariogram equation in the second column indicates the effect of nuggets from classical models, which are generally greater than the fitting produced by the semivariogram robust equation. The level of precision error in the WLS base is always smaller than that of the OLS base. The error prediction (RMSE and MAD) was obtained based on the parameters in Table-2, as shown in Table-3.

VOL. 12, NO. 4, FEBRUARY 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Table-2. Semivariogram fitting and RMSE for each model.

Model	Semivariogram function	RMSE
Model-1	$\begin{cases} 4.412\,; \ \mathbf{h} = 0 \\ 4.412 + 7.141\left[1.5\left(\frac{\lvert\mathbf{h}\rvert}{189.050}\right) - 0.5\left(\frac{\lvert\mathbf{h}\rvert}{189.050}\right)^3\right]; \ 0 < \lvert\mathbf{h}\rvert \leq 189.050 \\ 11.553; \ \lvert\mathbf{h}\rvert > 189.050 \end{cases}$	0.241
Model-2	$\begin{cases} 3.941\,; \ \mathbf{h} = 0 \\ 3.941 + 7.567\left[1.5\left(\frac{\lvert\mathbf{h}\rvert}{182.269}\right) - 0.5\left(\frac{\lvert\mathbf{h}\rvert}{182.269}\right)^3\right]; \ 0 < \lvert\mathbf{h}\rvert \leq 182.269 \\ 11.508; \ \lvert\mathbf{h}\rvert > 182.269 \end{cases}$	0.215
Model-3	$\begin{cases} 2.365; \ \lvert\mathbf{h}\rvert = 0 \\ 2.365 + 9.332\left[1 - \exp\left(-\frac{\lvert\mathbf{h}\rvert}{63.214}\right)\right]; \ 0 < \lvert\mathbf{h}\rvert \leq 189.372 \\ 11.697; \ \lvert\mathbf{h}\rvert > 189.372 \end{cases}$	0.216
Model-4	$\begin{cases} 2.274; \ \lvert\mathbf{h}\rvert = 0 \\ 2.274 + 9.341\left[1 - \exp\left(-\frac{\lvert\mathbf{h}\rvert}{61.200}\right)\right]; 0 \leq \lvert\mathbf{h}\rvert \leq 183.340 \\ 11.616; \ \lvert\mathbf{h}\rvert > 183.340 \end{cases}$	0.207
Model-5	$\begin{cases} 1.408\,; \ \mathbf{h} = 0 \\ 1.408 + 9.694\left[1.5\left(\frac{\lvert\mathbf{h}\rvert}{149.543}\right) - 0.5\left(\frac{\lvert\mathbf{h}\rvert}{149.543}\right)^3\right]; \ 0 < \lvert\mathbf{h}\rvert \leq 149.543 \\ 11.102; \ \lvert\mathbf{h}\rvert > 149.543 \end{cases}$	0.830
Model-6	$\begin{cases} 1.574\,; \ \mathbf{h} = 0 \\ 1.574 + 9.518\left[1.5\left(\frac{\lvert\mathbf{h}\rvert}{161.546}\right) - 0.5\left(\frac{\lvert\mathbf{h}\rvert}{161.546}\right)^3\right]; \ 0 < \lvert\mathbf{h}\rvert \leq 161.546 \\ 11.092; \ \lvert\mathbf{h}\rvert > 161.546 \end{cases}$	0.865
Model-7	$\begin{cases} 0.000 + 11.391\left[1 - \exp\left(-\frac{\lvert\mathbf{h}\rvert}{60.047}\right)\right]; \ 0 \leq \lvert\mathbf{h}\rvert \leq 179.885 \\ 11.391; \ \lvert\mathbf{h}\rvert > 179.885 \end{cases}$	0.702
Model-8	$\begin{cases} 0.000 + 11.302\left[1 - \exp\left(-\frac{\lvert\mathbf{h}\rvert}{59.429}\right)\right]; 0 \leq \lvert\mathbf{h}\rvert \leq 178.033 \\ 11.302; \ \lvert\mathbf{h}\rvert > 178.033 \end{cases}$	0.683

ARPN Journal of Engineering and Applied Sciences

Table-3. RMSE and MAD based on cross validation (C.Val) and 10%, 20%, 30%, 40%, and 50% losing data prediction.

Method	Model	Error	C.Val	10%	20%	30%	40%	50%	Rank
Classical	**Model-1**	**RMSE**	**2.142**	**1.279**	**2.365**	**1.936**	**1.940**	**2.275**	**1**
		MAD	1.205	1.308	1.297	1.930	1.772	1.855	
	Model-2	RMSE	2.121	1.286	2.385	1.945	1.947	2.281	3
		MAD	1.223	1.385	1.302	1.930	1.749	1.991	
	Model-3	RMSE	2.108	1.183	2.287	1.922	1.887	2.272	2
		MAD	1.198	1.124	1.186	1.966	1.971	2.124	
	Model-4	RMSE	2.102	1.171	2.262	1.903	1.868	2.249	1
		MAD	1.200	1.120	1.170	1.957	1.873	2.116	
Robust	Model-5	RMSE	2.131	1.323	2.408	2.011	2.040	2.424	3
		MAD	1.204	1.397	1.397	2.625	2.341	2.306	
	Model-6	RMSE	2.123	1.340	2.513	2.070	2.073	2.453	2
		MAD	1.199	1.434	1.378	2.638	2.163	2.309	
	Model-7	RMSE	2.106	1.258	2.426	2.069	2.062	2.459	1
		MAD	1.228	1.409	1.308	2.203	2.891	2.499	
	Model-8	RMSE	2.106	1.253	2.415	2.061	2.053	2.448	1
		MAD	1.227	1.409	1.299	2.195	2.867	2.491	

One of the important values generated by OK is the mean kriging prediction. Column 4 of Table-3 indicates the error (RMSE and MAD) of the mean kriging prediction based on cross validation. Columns 5 to 9 show the error of both RMSE and MAD of the losing data prediction. The removed data were used to predict the remaining data in the current study. Column 10 is the decision rank, which is calculated based on multi-criteria decision making rules of TOPSIS. Information can be obtained where the principle classical Models 1 and 4 occupy the first position. The first order of the robust models is generated by predicting Models 7 and 8.

CONCLUSIONS

The following conclusions can be obtained based on the results of the current study.

- The smallest RMSE value of classical fitting semivariogram groups was observed through OLS based on the exponential function, which is depicted as Model 4, namely, 0207. The robust models apply to Model 8, namely, 0683.
- Model 4 also occupies the first rank based on TOPSIS model for jackknifing cross validation and kriging predictions using data reduction. This model prediction is based on classical parameters (in addition to Model 1). The robust parameter base applies to Model 8 (in addition to Model 7).
- A relationship exists between the precision in fitting semivariogram and the "degree of precision" in the "ordinary kriging prediction."

ACKNOWLEDGEMENTS

The authors would like to thank PT. Aneka Tambang UBPE Pongkor, especially Ir. Kristiawan who provided the data for the study and to UPN Yogyakarta, also FST and SERI UKM for the financial support (DLP-2015-005).

REFERENCES

[1] Syafrizal, Indriati T. and Valentin K. 2009. Studi distribusi ukuran butir elektrum dan asosiasi mineralisasi emas pada urat Ciurug, Pongkor, Indonesia. Journal Teknologi Mineral. 16(2): 99-109.

[2] Barnes J.W. 1988. Ore and minerals: Introducing economic geology. Milton Keynes, Philadelphia: Open University Press.

[3] Annels, A.E. 2011. Mineral Resources Evaluation A Practical Approach, 2nd Edition. London: John Davis.

[4] Bargawa W.S., Rauf A. and Amri N.A. 2015. Gold resource modeling using pod indicator kriging. AIP Proceeding of Progress in Applied Mathematics in Science and Engineering. Bali, 29 September - 1 October.

[5] Journel A.G. and Huijbregts C. J. 1978. Mining geostatistics. New York: Academic Press.

[6] Cressie N. 1993. Statistics for Spatial Data, New York: John Wiley & Sons, Inc.

[7] Matheron G. 1963. Principles of geostatistics. Journal of Economic Geology 58: 1246-1266.

[8] Cressie N and Hawkins D.M. 1988. Robust estimation of the variogram. Journal of the International Association for Mathematical Geology. 12(2): 115-125.

[9] Bivand R.S., Pebesma E.J. and Rubio V.G. 2008. Applied Spatial Data Analysis with R, New York: Springer Science+Business Media, LLC.

[10] Huang C., Zhang H. and Robeson S.M. 2011. On the validity of commonly used covariance and variogram functions on the sphere. Journal of Mathematical Geosciences. 43: 721-733.

[11] Webster R. and Oliver M.A. 2007. Geostatistics for environmental scientists. Second Edition. England: John Wiley & Sons Ltd.

[12] Giraldo R., Delicado P. and Mateu J. 2011. Ordinary kriging for function-valued spatial data. Journal of Environmental and Ecological Statistics. 18(3): 411-426.

[13] Khakestar M.S., Madani H., Hassani H. and Moarefvand P. 2013. Determining the best search neighbourhood in reserve estimation, using geostatistical method: A case study anomaly No 12A iron deposit in central Iran. Journal of the Geological Society of India. 81(4): 581-585.

[14] Amri N.A., Abdul Aziz J. and Wan Fuad Wan Hassan. 2014. Ordinary kriging based on OLS-WLS fitting semivariogram: Case of gold vein precipitation, Proceedings of the 3rd International Conference on Mathematical Sciences. AIPConf. Proc. 1602, hlm. 1039-1045.

[15] Ilyas A. and Koike K. 2012. Geostatistical Modeling of Ore Grade Distribution from Geomorphic Characterization in a Laterite Nickel Deposit. Journal of Natural Resources Research. 21(2): 177-191.