ISSN: 04532198
Volume 62, Issue 03, April, 2020

A Comparison of Spatial Interpolation Methods to Estimate Coal Thickness and Quality Based on The Value of Root Mean Square Error (RMSE)

Irfan Marwanza[1*)], Chairul Nas[2], Masagus Ahmad Azizi[3], Wiwik Dahani[4], Subandrio[5], Ericsson Nirwan[6], Riskaviana Kurniawati[7]

Mining Engineering Department, Trisakti University, Indonesia 11440[1,2,3,4,5,6,7]
[*)]Corresponding Author : irfanm@trisakti.ac.id

Abstract— Exploration is one of the activities of the mining process which aims to obtain information about the geological conditions of a deposit that is below the surface of the land. Exploration activities have risks and require a large cost. Therefore, a more accurate estimation method is needed in determining the value of areas that are not sampled in exploration based on the surrounding data. In coal deposits, there are two main output parameters which are coal thickness and quality. In this study, the coal quality parameters used are calorific and sulfur values. In this study, we will compare the Kriging method with Inverse Distance Weighting, in estimating coal thickness and quality. The aim is to find out the most accurate estimation of coal thickness and quality between the Kriging and IDW methods based on the Root Mean Square Error (RMSE) value. The method of testing the estimation results is cross validation. The choice of the variogram model is based on the lowest RMSE value. The research method used is a quantitative method. Coal exploration data in the form of bar data were analyzed with descriptive statistics with Minitab 17 software and continued with geostatistical analysis using GS + software. From the results of the calculation of the two estimation methods, it was found that the kriging method was more accurate than the IDW method based on the lower RMSE Kriging value on the thickness data, calorific value and coal sulfur worth 1,622 m, 71,504 Kcal / Kg, and 0.140% compared to the IDW method RMSE on coal thickness, calorific value, and sulfur data are 1,704 m, 74,731 Kcal / Kg, and 0.142%.

Keywords— Coal, Geostatistics, Kriging, Inverse Distance Weighting, Root Mean Square Error

1. Introduction

Exploration is one of the activities of the mining process that aims to obtain information about the geological conditions of a deposit that is below the surface of the land. Exploration activities are highly risk and require large cost. Therefore, a more accurate estimation method is needed in determining the value in areas that are not sampled based on the surrounding data. In this study, kriging method will be compared with Inverse Distance Weighting in estimating thickness and quality. Based on several previous studies comparing the kriging and IDW method shows that the kriging method is more accurate than the IDW method. [1], [2], [3].

In coal deposits, there are two main output parameters which are coal thickness and quality. In this study, the quality of coal to be estimated includes the value of calories and sulfur. In the actual situation in the field, the distribution of coal parameters has a heterogeneous distribution. This is influenced by changes in geological conditions during the coal deposition process.

Kriging is a geostatistical method that uses a basic device in the form of a variogram [4]. Variograms are used to measure spatial correlations between observation points. The variogram model is a mathematical function that has been matched to the experimental variogram. The model can be used to estimate values at points that are not sampled [5]. To provide a test method for the variogram model, cross validation is used. The choice of the variogram model is based on the lowest RMSE value. One kriging technique that is commonly used is ordinary kriging [6]. IDW method uses a number of data around it to predict the data sought. Each data has an influence on the prediction results according to their weight. Data weight is determined by the distance to the location sought.

As coal demand in the world increases, coal exploration and production activities must also increase. This requires the importance of knowing the spatial distribution pattern of coal thickness and calorific value to control the progress of the mine development plan so that it runs effectively and efficiently so that the production process to coal sales will be more economical.

2. Literature Review

2.1. Basic Statistic

Statistics is a set of ways or rules relating to the collecting, processing, and concluding, on data in the form of numbers by using certain assumptions [7]. Basic statistical analysis in geostatistics is used to find out that the data to be calculated experimental variogram comes from normal data, because geostatistics assumes data to be normal [8]. Statistical analysis used is descriptive statistical analysis.

Descriptive statistics are generally divided into 3 categories, which are location size, distribution size, and shape size [9]. Statistics in the first group provide information about where the various parts of the distribution are located. The mean, and median can provide some ideas where the distribution center is located. The second group includes variance, and standard deviation. This is used to describe variability in data values. The shape of the distribution is explained by the coefficient of skewness and the coefficient of variation [10].

In observing the normality of data, the coefficient of variance (CoV) and skewness can be used. If the CoV value is less than 0.5 then the data are considered to be normally distributed [11]. Normal data distribution has a skewness value of ± 1 [12].

2.2. Geostatistics

Geostatistics is a statistical method used to see the relationship between variables measured at a certain point with the same variables measured at a point with a certain distance from the first point of view (spatial data) and is used to estimate data in an unknown value [13]. Geostatistics has a theory known as "Regionalized Variables Theory" [14]. which contains:

a. Random data
b. Not independent data
c. Spatially adjacent data variations are smaller than spatially spaced data.

In its application, the geostatistical method can work optimally if the data is normally distributed and stationary (the mean and variance do not change too significantly). Data can be considered to be stationary if it does not have a tendency or the fluctuations in the data is close to the average value or the constant variance with space [6].

Geostatistical estimates are divided into 2 stages, modeling of spatial variability within the study area, and using this spatial model to provide appropriate estimation techniques. The first stage consists of the construction and interpretation of a semi-variogram graph, and the second is the development of a suitable kriging method [15].

2.3. Variogram

Variogram is a tool used in geostatistics in the form of graphs to show spatial variations between the measured data. According to the direction of data search, variograms are divided into two types which are omnidirectional and directional variograms. The omnidirectional variogram searches for data pairs in all directions horizontally, while the direction variogram requires a specific azimuth direction to search for the data pairs. Omnidirectional variograms are used when deposits have isotropic continuity, which is equal in all directions. By variogram, isotropic means having the same range for all directions of azimuth. Directional variogram is used on sediments that have anisotropic continuity, which means they have a different continuity distance for each azimuth.

The value of the experimental variogram for the separation distance from h (lag) is half of the average squared difference in the value of Z between pairs of samples separated by distance [10].

$$\gamma(h) = \sum_{i=1}^{N} \frac{\left(Z_i - Z_{(i+h)}\right)^2}{2N} \qquad (1)$$

Information:

$\gamma(h)$	= Variogram value at interval h
N	= Number of data pairs.
Zi	= Value at point i
$Z(i+h)$	= Value of a point as far as h from the i-th point

Variogram is the foundation of many geostatistical applications. The experimental variogram and any models that fit it must be accurate. Only then can the variogram model describe minimum error. Kriging requires a variogram and by ensuring its accuracy which will result in a minimum error with kriging. If the variogram represents a bad error, then kriging estimation tends to be bad as well [16]. There are many causes for bad variograms, for example insufficient data, mismatched models, and poor installation.

2.4. Cross Validation

Before the interpolation model is used, it is necessary to know in advance how accurate the model is used. One way to test the accuracy of a model is to use cross validation. The procedure of this method is to eliminate one data and use the remaining samples as data to predict the data that is removed with the model.
The index most often used to assess the accuracy of interpolation of a method is the root mean square error (RMSE) [11], which is defined by:

$$RMSE = \sqrt{\sum_{i}^{n}\left(Z_i - \check{Z}_0\right)^2} \qquad 2)$$

Information :

Ži	: Predicted value at the i-th location
Z_i	: The actual value at the i-th location.

The accuracy of the measurement error estimation method is indicated by the presence of a small RMSE value. An estimation method that has a smaller RMSE is said to be more accurate than an estimation method that has a larger RMSE [15].

2.5. Ordinary Kriging

Ordinary Kriging is one of the methods found in the kriging method used in geostatistics. This method is an interpolation of a variable value at a particular point by observing similar data at particular location [ ]. This assessment is widely used because its simplicity and easy to understand. This method is a triggering method that produces an estimator that is BLUE (Best Linear Unlimited Estimator). Unbiased Estimator is the best linear estimator. Ordinary kriging is linear because the estimators are influenced by a linear combination of data, not biased because it aims to get the mean error equal to zero, and is said to be the best because it aims to reduce the error variance [10].

The equation used by kriging to determine samples of unknown value is:

$$\check{Z}_0 = \sum_{i=1}^{n} W_i \,.\, Z_i \qquad (3)$$

Information :

Ž_0 = Estimated point value
W_i = Weighting factor from point i
Z_i = Value of the estimator point

2.6. Inverse Distance Weighting

Inverse Distance Weighting (IDW) is an interpolation technique that calculating the relationship of space location (distance), and is a linear combination or weighted average of the data points around it. The main factor influencing the accuracy of IDW interpolation is the value of the power parameter (p) [18]. Power is influential in determining the value of data samples in the interpolation calculation which serves to regulate the significance of the influence of the points around it. In this study, an estimation comparison of the IDW using the power parameters 1,2,3,4, and 5 is the most commonly used as literature [19].

Optimal power is determined by minimizing the root mean square error (RMSE). RMSE is a statistic calculated from cross validation. In cross validation, each point measured is deleted and compared with the predicted value for that location. RMSE is summary statistics that measure prediction errors [20]. The geostatistical analysis attempts several different powers on IDW to identify the power that produces the minimum RMSE. The diagram below shows how geostatistical analysts calculate optimal power. RMSE is plotted for several different strengths for the same dataset. The curve matches points (quadratic Local Polynomial equation), and from the power curve that provides the smallest RMSE determined as optimal power.

Weighting values in the IDW technique are generally calculated by the following equation : [10]

$$W_i = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \qquad (4)$$

Information :

di : Distance between the i-th observation point and the alleged point
P : power (integer).

The general formula for Inverse Distance Weighting is [10]:

$$\check{Z}_0 = \sum_{i=1}^{n} W_i \, . \, Z_i \qquad (5)$$

Information :
Ž_0 = Estimated point value
W_i = Weighting factor from point i
Z_i = Value of the estimator point

3. Research Methodology
The method used in this research is quantitative method. Quantitative methods are used to process secondary data, which are coal exploration data in the form of drill data. Exploration drilling data in the form of easting, northing coordinates, thickness of calorific value, and coal sulfur will be carried out descriptive statistical analysis with Minitab software 17. Continued geostatistical analysis using GS + software.

3.1. Data
This study used secondary data from coal exploration results located in Sangatta field, East Kalimantan, Indonesia [ ]. The data used include:
1. The name of the drill hole (DDH)
2. Borehole coordinates (easting and northing), easting starting from 96074.91 ° E to 99145.32 ° E and northing starting from 194243.5 ° N to 199768.7 ° N.
3. The calorie value of coal in each drill hole in units of Kcal / Kg.
4. Coal sulfur in units of %.
5. The thickness of the coal in each drill hole in meters.
6. The value of coal density is assumed to be 1.3 kg / m^3.
7. Drill holes in the study amounted to 142 points.

4. Results and Discussion

4.1. Descriptive Statistics Analysis
Based on the theory, if the CoV value is less than 0.5 then the data is considered normal distribution [11]. Based on the results of descriptive statistical analysis, thickness data, calorific value, and coal sulfur, the CoV values obtained were 0.27, 0.01 and 0.5, so the thickness, caloric and sulfur value of coal are normally distributed and can be used in geostatistical analysis. A summary of the statistics from the data analysis is presented in Table 1. Basic Statistics on Coal Thickness and Quality below.

Table 1. Basic statistics of coal thickness and quality

Descriptive statistics	Thickness	Calorie value	Sulfur
Mean	6.94	7509	0.3953
StDev	1.903	102.2	0.2
Variance	3.622	10440.2	0.041
CoV	0.27	0.01	0.5
Minimum	2.4	7236.4	0.15
Median	6.95	7520.8	0.33
Maximum	11.46	7771.8	0.97
Skewness	0.03	-0.18	0.94
Kurtosis	0.05	-0.24	0.15

4.2. Variogram

Variogram model and variogram parameters of coal thickness, calorific value and sulfur can be seen in Table 2. Thickness and Quality Variogram of Coal. Seen the type of model, nugget effect, sill and range for each parameter.

Table 2. Coal thickness and quality variogram

	Model	Nugget Effect	Sill	Range (m)
Thickness	Spherical	0.6	4.5	600
Calorie value	Exponential	3500	10000	566.67
Sulphur	Exponential	0.009	0.036	1800

The data shown in Table 2 was obtained based on the analysis of experimental variograms and the variogram modeling of each parameter using simulations in GS + software, seen in Figures 1.2 and 3, where Figure 1. shows the Thickness Variogram Model, Figure 2 Variogram Model Calorie Value and Figure 3. Sulfur Variogram Model.

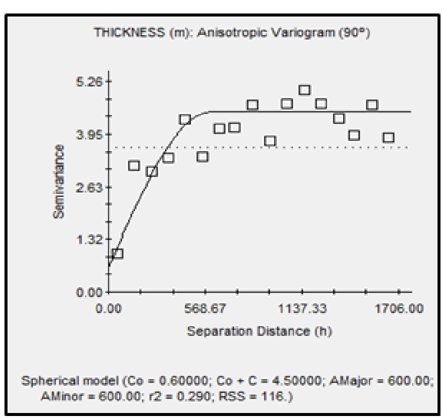

Figure 1. Thickness variogram model

ISSN: 04532198
Volume 62, Issue 03, April, 2020

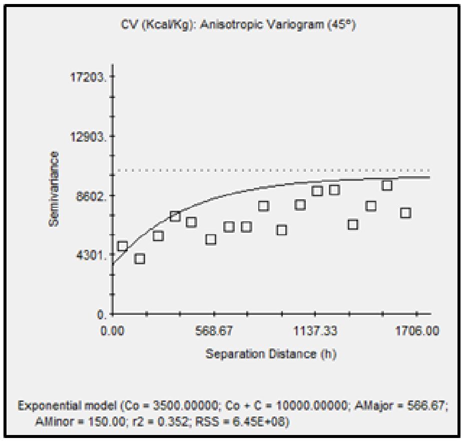

Figure 2. Calorie value variogram model

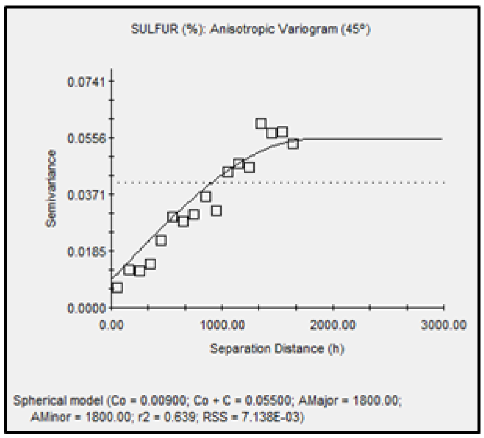

Figure 3. Sulfur variogram model

4.3. Kriging

Kriging estimation in research uses GS + software with Ordinary Kriging method. Figures 4, 5 and 6 show the results of the Kriging analysis for each parameter. In Figure 4. Kriging Coal Thickness, the thickness of the coal have variations, the thick coal is seen in the west and south of the map, while the thin section is in the middle of the map location.

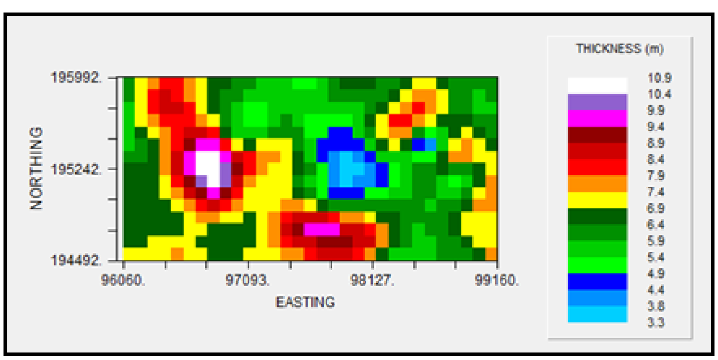

Figure 4. Coal thickness kriging

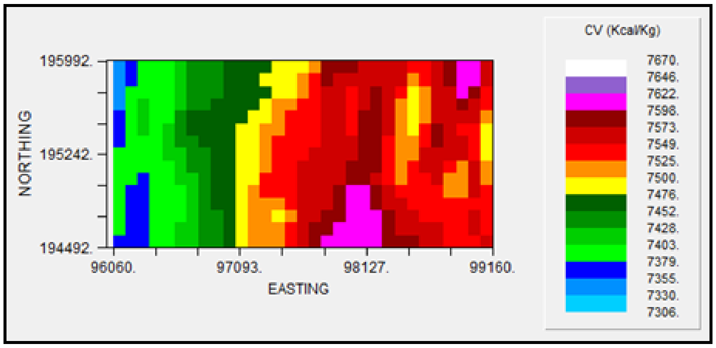

Figure 5. Coal calorie value kriging

Figure 5. Coal Calorie Value Kriging, the coal calorie value is low in the West and increasing in the East. In Figure 6. Kriging Sulphur values tend to shrink to the North.

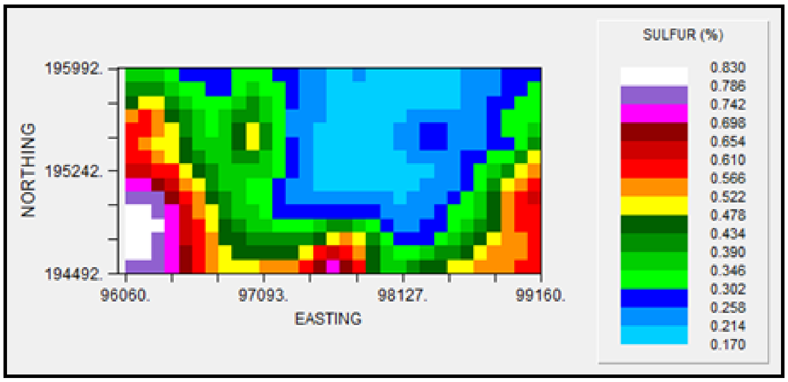

Figure 6. coal sulphur kriging

4.4. Inverse Distance Weighting

Inverse Distance Weighting (IDW) estimation is done using GS + software, the following is the result of estimated caloric, sulphur and coal thickness values using the IDW method, shown in Figure 7. Inverse Distance Weighting of Coal Thickness, Figure 8. Inverse Distance Weighting of Coal Calorie Value and Figure 9. Inverse Distance Weighting Sulphur.

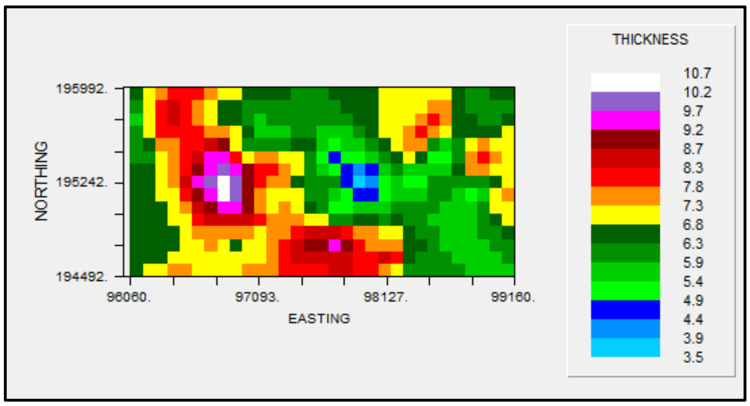

Figure 7. Inverse distance weighting coal thickness

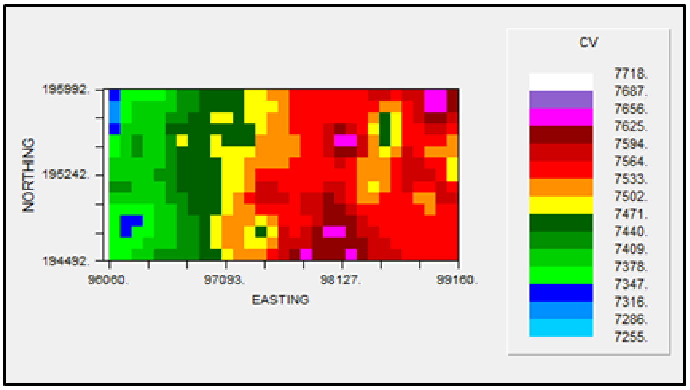

Figure 8. Inverse distance weighting of coal calorie value

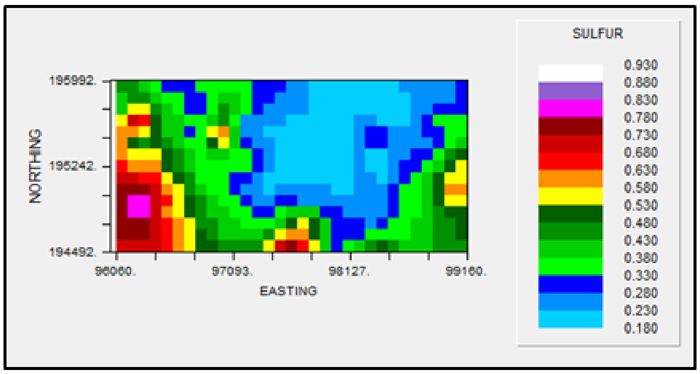

Figure 9. Inverse distance weighting sulphur

In Figures 7, 8 and 9, the results of the Inverse Distance Weighting (IDW) analysis are visually seen to be visibly the same as the results of Kriging, but the values are different.

4.5. Kriging and IDW Estimation Results

Table 3. Estimated Kriging and IDW methods

Parameter	Kriging	IDW
Coal volume (ton)	44889000	45693960
Coal thickness mean (m)	6.744	6.865
Calorie value mean (Kcal/Kg)	7496.95	7497.41
Sulphur mean (%)	0.396	0.390

According to the calculation results obtained in table 3. Estimation of the Kriging and IDW Method, the kriging method and the IDW method produce slightly different values in the four parameters, which are coal volume, coal thickness means, calorific value mean, and sulphur mean.

The volume of kriging coal produced a higher volume of 44889000 tons. This result was certainly due to the results of the average coal thickness obtained from the kriging which is higher than the IDW method of 6,744

m. This is different from the estimated coal volume of IDW which results in a lower coal volume of 45693960 tons. The lower coal volume is due to the lower average coal thickness of IDW, which is also 6,865 m. For the average calorific value, kriging estimation produces an average calorific value which is lower with a value of 7496.95 Kcal / Kg compared to the IDW estimate which produces an average calorific value of 7497.41Kcal / kg. On average sulphur, kriging estimation yields lower average sulphur which is 0.394% while IDW estimation produces sulphur 0.39% on average.

4.6. Discussion

4.6.1. The accuracy of the Kriging Method with IDW

The measure that can be used to compare the accuracy of an estimate is the Root Mean Square Error (RMSE). The smaller the estimated RMSE value indicates the more accurate the estimation results.

Table 4. RMSE Kriging and IDW methods

Root Mean Square Error	**Kriging**	**IDW**
Coal thickness (m)	1.694	1.704
Coal calorie value (Kcal/Kg)	74.539	74.73
Coal sulphur (%)	0.141	0.142

The kriging method produces a more accurate estimate than the IDW method. Based on the theory that has been explained, the kriging method is a method that produces an estimator that is BLUE (Best Linear Unlimited Estimator). Unbiased Estimator is the best linear estimator. Ordinary kriging is linear because the estimators are influenced by a linear combination of data, not biased because it aims to get the mean error equal to zero, and is said to be the best because it aims to reduce the error variance [10].

The smaller RMSE value of the kriging method is caused by the selection of an appropriate variogram model. To provide a test method for the variogram model, it is enough to represent the estimation results using cross validation. Actual values and estimated values are then compared in such a way that the model can be accepted or not [22]. Therefore, making and selecting a variogram model must be in accordance with the experimental variogram so that the estimated kriging results in the lowest error value [23].

5. Conclusions

Based on the estimated thickness, caloric and sulphur values of coal in the field, several conclusions can be drawn, including:

1. Descriptive statistical analysis results show the thickness, calorie value, and sulphur values of coal are normally distributed with a CoV value of less than 0.5 and a skewness value of ± 1.
2. The variogram model for coal thickness and sulphur data is a spherical model. The variogram model for coal calorie value data is an exponential model.
3. The RMSE value in the thickness, calorific value, and sulphur of the coal kriging method is lower than the IDW method. This means that the kriging method is more accurate than the IDW method.

6. Acknowledgments

This research was supported by the Department of Mining Engineering, Faculty of Earth and Energy Technology, Trisakti University, Jakarta, Indonesia. Gratitude delivered to PT. Kaltim Prima Coal, Sangatta, East Kalimantan and all those who contributed to the success of this research for their support and assistance. Gratitude is also expressed to ICSTEEEM IJAST.

7. References

[1] Bargawa, W.S., and Hendro Purnomo,"Performance Evaluation of Ordinary Kriging and Inverse Distance Weighting Methods for Nickel Laterite Resources Estimation", International Conference on Science and Technology 2nd Geomaritime Symposium 2016 Yogyakarta, 2016.

[2] Qiao P1,2, Lei M3, Yang S2, Yang J1, Guo G4, Zhou X1., "Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing", Environ Sci Pollut Res Int. 2018, June, Volume 25, Number 16, pp. 15597-15608, Doi: 10.1007/s11356-018-1552-y. Epub, 2018.

[3] Agung Setianto and Tamia Triandini, "Comparison Of Kriging And Inverse Distance Weighted (IDW) Interpolation Methods in Lineament Extraction and Analysis", J. SE Asian Appl. Geol., Volume 5, Number 1, pp. 21–29, 2013.

[4] Irfan Marwanza, Masagus Ahmad Azizi, Chairul Nas, Claudia Intan Permatasari, Wiwik Dahani, Subandrio, "The Use of Geostatistics in Setting the Spread of Claystone in Coal Mines", International Journal Of Scientific & Technology Research. Volume 8, Issue 10, 2019, ISSN 2277-8616 3607 IJSTR©2019 www.ijstr.org.

[5] Marwanza, I., Nas, C., Badaruddin, S.P., Azizi, M.A.," Copper resource estimation in PT X Batu Hijau, Regency of West Sumbawa, West Nusa Tenggara Province using geostatistical method". IOP Conference Series: Earth and Environmental Science, 2018, 012024. IOP Publishing. doi:10.1088/1755-1315/212/1/012024

[6] Bohling, G. "Introduction to Geostatistics and Variogram Analysis", Kansas, 2005.

[7] Shumway. R. H., "Statistics and Data Analysis in Geology", 2012, DOI10.1080/00401706.1987.10488290

[8] Oliver, Schabenberger and Carol A. Gotway. "Statistical Methods for Spatial Data Analysis", Boca Raton : Chapman & Hall/CRC, 2005.

[9] Cressie, N.A.C., "Statistics for Spatial Data, Revised Edition", Published by Wiley Classics Library, John Wiley, 2015, Total number of pages: 928. ISBN: 978-1-119-11518-2. DOI: 10.1111/jtsa.12168

[10] Isaaks, E.H and R.Mohan Srivastava. "An Introduction to Applied Geostatistics", 1989, New York : Oxford University Press.

[11] Koch,G.S and Link, R.F. "Statistical Analysis of Geological Data".(2002). New York : Wiley & Sons.

[12] Oliver, M. & Webster,R., "Basic Steps in Geostatistics: The Variogram and Kriging", 2015, Springer : 10.1007/978-3-319-15865-5.

[13] I. Marwanza, C. Nas, M. A. Azizi and J. H. Simamora, "Comparison between moving windows statistical method and kriging method in coal resource estimation", 4th Annual Applied Science and Engineering Conference Journal of Physics: Conference Series 1402, 2019, 033016, IOP Publishing doi:10.1088/1742-6596/1402/3/033016.

[14] Matheron, G. "The Theory of Regionalized Variables and Its Application", 1971, Les Cahiers du Centre de Morphologie Mathematique, Fas. 5, C. G.Fontainebleau

[15] Clark, I. "The Art of Cross Validation in Geostatistical Applications", 1986, Colorado Society of Mining Engineers., pp.211- 220

[16] Webster, R. and Oliver M.A., "Geostatistic for Enviromental Scientists", 2001, John Wiley & Sons Brisbane, Australia, 281 PP.

[17] Stefanie Jachner, K. Gerald van den Boogaart and Thomas Petzoldt, ”Statistical Methods for the Qualitative Assessment of Dynamic Models with Time Delay (R Package qualV)”, Journal of Statistical Software, Volume 22, Issue 8, 2007, http://www.jstatsoft.org/

[18] Deutsch, J.L., Szymanski, J. and Deutsch, C.V., “Checks and measures of performance for kriging estimates”, Journal of the Southern African Institute of Mining and Metallurgy, 2014, 114:3, 223-230.

[19] G. Garnero, and D. Godone,”Comparisons Between Different Interpolation Techniques”, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W3, 2013, The Role of Geomatics in Hydrogeological Risk, 27 – 28 February 2013, Padua, Italy.

[20] Alexandra Kravchenko and G. Bullock,” A Comparative Study of Interpolation Methods for Mapping Soil Properties”, Agronomy Journal, Volume 91, Number 3, 1999, DOI: 10.2134/agronj1999.00021962009100030007x

[21] Nas, Chairul, “Spatial Variations In the Thickness and Coal Quality of the Sangatta Seam, Kutei Basin, Kalimantan, Indonesia”, 1994, University of Wollonggong Thesis Collection.

[22] Tom Mueller , B. Pusuluri. K. Mathias, A. Shearer Scott, ”Map Quality for Ordinary Kriging and Inverse Distance Weighted Interpolation”, 2004, Soil Science Society of America Journal, Volume 68, pp. 2042-2047 DOI: 10.2136/sssaj2004.

[23] Fakhire Akbar and Mahmood Najafi Zilaie,”Comparison of Different Kriging Methods to Estimate the Tree Density, A Case Study: West of Karkheh in Southwest of Iran”, Usamvcluj Journal. Promediu. ProEnvironment Volume 7, 2014, pp. 204 – 212.