Australian Journal of Basic and Applied Sciences, 7(10): 562-569, 2013
ISSN 1991-8178

# Robust Estimation for Two Different Sets of Spatial Data with Application

[1]Ghanim Mhmood Dhaher, [2]Muhammad Hisyam Lee

[1]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
[2]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.

**Abstract:** This paper deals with the application of spatial prediction in mining field. The motivation of this research is an estimate completion of kriging techniques performance. Kriging techniques and two methods of robust variogram estimator are applied in this work. It uses an experimental variogram function of regionalized variables and the application of this function in all directions of compass. The objectives of this research are; to determine the ideal method of parameters to obtain conciliation of the mathematical model; to predict the hypotheses in a random process of second order stationary and isotropy; to an estimate and determine the best performance of robust variogram estimator. The data adopted of this work include real spatial data consists of (200) samples of two metals, the first data set involve the chromium ore (Cr) which lies in a regular lattice and the second set is copper ore (Cu) lies in an regular lattice and each set contains (100) from the ore in Galang river Tomas Don. Through the behaviour of all the curves a clear correspondence between the results of experimental variogram functions and covariance models as compared with robust variogram estimator, which emphasizes the compliance and similarity between the results of the values of the kriging variance. In addition, the prediction process gives an idea of the error include some extent the accuracy of the results can be noted. The estimation by using kriging variance and kriging variance estimator, as well as the standard error show are small and few estimators which indicates and supports the error in prediction process. In conclusion, the model of chromium data is similarity to Gaussian model of covariance function while the model of zinc data resembles very much to the form of power model and this indicates that the nature of the data of the first metal (chromium) is different from the nature of the second metal data (copper).

**Key words:** spatial variable, kriging, stationary, isotropy, robust variogram estimator.

## INTRODUCTION

Geology began applying the theory of the regionalized variables to solve problems of prediction of geology and mining field. Applied statistics has many types: one of them is Geostatistics or called spatial statistics or mining engineering, which are used to solve many problems of the spatial prediction. The study of the spatial data includes the real values of spatial real data that is obtained from the areas of metal ores, underground water, or natural plants and their locations. Matheron (France) was the first scientist worked in the field of the spatial statistics based on the thesis of Krige (South African). Later, many scientists were interested in prediction of spatial random process. (Matheron,1963), (Journel and Huijbregts, 1978),(Christakos, 1992) (Zidek and Lee, 1992), (Cressie, 1993),(Goovaerts, 1997), (Stein, 1999). The first step of kriging technique is to compute the variogram function (Isaaks and Srivastava, 1989). (David and Genton, 2000). The method of kriging techniques are used in hydrology and meteorology, as well as environmental science and structural engineering. (Lark, 2009), (Marchant, *et al*, 2011), (Lark, and Lapworth, 2012), and many studies deals with the robust the variogram estimator such as: (Mingoti and Rosa, 2008), (Lark, 2008). The main purpose of this work is to obtain the best estimator and the best performance of robust variogram estimator in the field of study in the mining data.

***Methodology:***

The spatial data obtained from the application of environment, agriculture, earth sciences where the spatial variables includes the real value represents the phenomenon or observations (such as groundwater level, the degree of raw metal, or air pollution gases), with their locations (two dimensions or three dimensions) Suppose that $z(x)$ represents the spatial variable in the location $(x)$ within the region D, in the space of Euclid, where $x \in D \subseteq R^p$. Then the spatial variables are: $z(x_1), z(x_2), \ldots, z(x_n)$ in the sites $x_1, x_2, \ldots, x_n$, and let $z(x)$ and $z(x+h)$ are be two random variables at two points $(x)$ and $(x+h)$ separated by the vector $h$. The variability between these two quantities is characterized by the variogram function:

**Corresponding Author:** Ghanim Mhmood Dhaher, Department of Mathematical Sciences, Faculty of Science Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
E-mail: ghanim_hassod@yahoo.com

$$2\gamma(x,h) = E\left\{\left[z(x_i) - z(x_i + h)\right]^2\right\} \quad (1)$$

In practice, in mining applications only one such realization $z(x)$ and $z(x+h)$ is available

To overcome this problem, the intrinsic hypothesis is introduced that the variogram function $2\gamma(x,h)$ depends only on the separation vector $(h)$ and not on the location $(x)$. It is possible to estimate the variogram $2\gamma(x,h)$ from the available data: an estimator of $2\gamma(x,h)$ is the arithmetic mean of the squared differences between two experimental measures $z(x_i) - z(x_i + h)$ at any two points separated by the vector $(h)$ i.e.

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

Where $n(h)$ is the number of experimental pairs $z(x)$, $z(x+h)$ of data with distance $(h)$.

Note that the intrinsic hypothesis is simply the hypothesis of second-order stationarity of the differences $z(x), z(x+h)$

***Variogram Parameters:***

In geostatistics, to describe and distinguish the variogram function must knowledge the following parameters:

1. Negget effect: is defined as the discontinuity at the origin point and denoted $(\psi_o)$.
2. Sill: is defined as the limit for variogram function to infinity, and denoted $(\psi_o)$.
3. Range: is defined as the distance on the x-axis increasing from the start of the curve from the point of origin to a distance of approach when it the curve of variogram is stability, and coded $(a)$ .refer to Figure (1) .(Chiles and Delfiner, 1999)

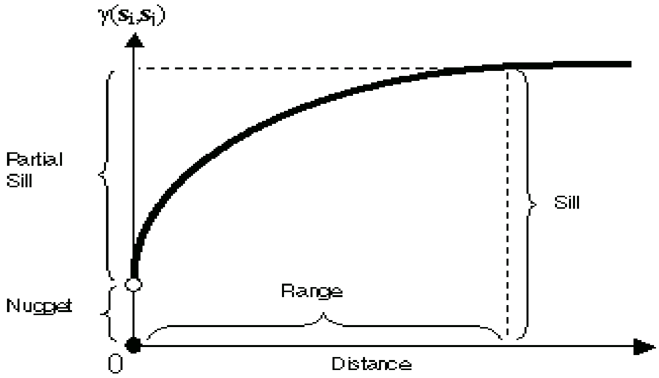


**Fig. 1:** variogram parameters.

***The Covariance:***

Let the variables $z(x)$ and $z(x+h)$ are be two random variables at two points $(x)$ and $(x+h)$ separated by the vector *h*, then they have Covariance relied up on the lag h and defined by:

$$\mathrm{cov}[z(x), z(x+h)] = E[\{z(x) - \mu\}\{z(x+h) - \mu\}] = C(h), \quad \forall x, x+h \in D \quad (3)$$

(Cressie, 1993), (Meyers, 1994)

***Moments and Stationary:***

***(a) First Order Moment:***

Let $z(x)$ be spatial random variable in location $(x)$. And if probability distribution functions for $z(x)$ has Expectation (that assumes its existence) .Then Expectation in general is a function and defined as:

$$E[z(x)] = \mu(x), \quad \forall x \in D \quad (4)$$

***(b) Second Order Stationary:***

Random function is called second order stationary if and only if:

$E[z(x)] = \mu(x),$

$$\text{cov}[z(x), z(x+h)] = E[\{z(x)-\mu\}\{z(x+h)-\mu\}] = C(h) \quad (5)$$

Defined for all values of $z(x)$ and, depends only on $(h)$ In this case

$\text{cov}[z(x), z(x+h)] = C(h)$

The relationship between covariance and variogram function and variance as in the following formula:

And

$$Var[z(x)] = C(0) = \sigma^2 \quad (6)$$

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

***Kriging:***

A random process $\{z(x), x \in D\}$ is a family of random variables where $D \subseteq R^p$ ,.p=1,2,3 Ordinary kriging is one of the most several types of kriging (Journel and Huijbregts, 1978), (Christakos, 1992) and, (Cressie, 1993). Ordinary kriging which makes two assumptions:

(i) The model assumptions $z(x) = \mu + e(x), \forall \text{ x} \in \text{D}, \mu \in \text{R}$ (7)

Where $\text{e}(x)$ error term.

(ii) Predictor for the point $(x_o)$ denoted by $\hat{z}(x_o)$

$$\hat{z}(x) = \sum_{i=1}^{n} [\lambda_i * z(x_i)] \quad (8)$$

$$E[\hat{z}(x)] = E[z(x)] = \mu(x), \ \forall x \in R$$

The ordinary kriging minimize the average square error prediction:

$$\sigma^2{}_k = E[z(x) - \hat{z}(x)]^2 \quad (9)$$

To minimize (10) and, let P be the Lagrangian multiplier then we can write the prediction error as:

$$\sigma_k{}^2 = E[z(x_0) - \hat{z}(x_0)]^2 - 2P(\sum_{i=1}^{n} \lambda_i - 1) \quad (10)$$

After many mathematical process then we get $2\gamma(\text{h}) = [e(x_0) - \sum_{i=1}^{n} \lambda_i e(x_i)]^2$

(Wim, and Jack, 2002), (Martina, 2009).

By matrices written

$$Q = -\lambda' \gamma \lambda + 2\lambda' \gamma_0 - 2p(\lambda' J - 1) \quad (11)$$

$J = (111...1)'$ ,$\gamma_0 = (\gamma_{01}\gamma_{02}...\gamma_{0n})'$ , $\gamma = (\gamma(x_i - (\text{x}_i + \text{h})))$ , $i = 1,2,3..., n$ , after derivative Eq. (11) by λ, P and equal to zero we get:

$$\lambda = \gamma^{-1}\gamma_0 - \gamma^{-1}(\frac{\gamma_0' \gamma^{-1} J - 1}{J' \gamma^{-1} J})' J \quad (12)$$

( Diggle, at al 2010)

***Kriging Variance :***

The optimal weights give the minimal mean squared kriging error, so becomes:

$$\sigma^2{}_k = E[z(x) - \hat{z}(x)]^2$$

$$= E[z(x_0) - \sum_{i=1}^{n} \lambda_i z(x_i)]^2 - \sum_{j=1}^{n} \sum_{i=1}^{n} \lambda_i \lambda_j \gamma(x_i - (\text{x}_i + \text{h})) + 2\sum_{i=1}^{n} \lambda_i \gamma(x_i - x_0)$$

By using the theorem of Best Linear Unbiased Estimator (BLUE), can get the following equations:

$$\hat{z}(x_o) = \lambda' z \tag{13}$$

$$\sigma_z^{\ 2} = \gamma_{00} - \gamma'_0 \gamma^{-1}\gamma_0 + [\frac{(1-\gamma_0 ' \gamma^{-1} J)^2}{J ' \gamma^{-1} J})]. \tag{14}$$

$$\sigma_{\hat{z}_o}^{\ 2} = \gamma'_0 \gamma^{-1}\gamma_0 + [\frac{(1-\gamma_0 ' \gamma^{-1} J)^2}{J ' \gamma^{-1} J})] \tag{15}$$

(Ripley, 1981), (Kumar and Remadevi, 2006)

***Variogram Estimators:***

For all the estimators in a sample of the spatial process $\{z(x), x \in D\}$, there are several types of robust estimators such as Cressie-hawkins robust, Haslett, Median, Genton,...etc, and two of methods were used.

***Cressie-Hawkins Robust Variogram Estimator:***

The estimator proposed by Cressie and Hawkins (1980) is given by

$$2\hat{\gamma}(h) = \frac{1}{\mathrm{B}}\left(\frac{1}{\mathrm{n}(h)}\sum_{\mathrm{i}=1}^{n(h)}\left|\mathrm{z}(\mathrm{x}_i) - \mathrm{z}(\mathrm{x}_\mathrm{i} + h)\right|^{0.5}\right)^4 \tag{16}$$

Where B= (0.457+0.457/n (h)) and B is a correction factor for bias.
(David and Genton, *2000)*.

***Data analysis:***

Data adopted in this research is the real spatial data and contains (200) samples for two sets of different data where (100) sample for chromium metal and (100) samples for copper metal from the ore in Galang river Tomas Don.(Rudolf, 2003) From this data the experimental variogram function is calculated according to the Equation (2) of the four tasks of the compass, and these directions are north - south, north east - south west, east - west, and north west - south east, through the application of experimental variogram function, The results of this function for chromium data with average of the directions that's have the same lag h=1,2,…,9 and the average of the same lag h=1.414, 2.828,…,12.727, these results reached in Table1.

**Table1:** Results variogram function for chromium (Cr)

| direction | $(h)$ | $\gamma(x)$ | direction | $(h)$ | $\gamma(x)$ |
|---|---|---|---|---|---|
| Average of $\theta=0^\circ \quad \theta=90^\circ$ | 1 | 0.0001 | Average of $\theta=45^\circ \quad \theta=135^o$ | 1.414 | 0.0001 |
| | 2 | 0.0002 | | 2.828 | 0.0002 |
| | 3 | 0.0002 | | 4.243 | 0.0003 |
| | 4 | 0.0002 | | 5.657 | 0.0004 |
| | 5 | 0.0003 | | 7.071 | 0.0005 |
| | 6 | 0.0004 | | 8.485 | 0.0007 |
| | 7 | 0.0006 | | 9.899 | 0.0011 |
| | 8 | 0.0007 | | 11.313 | 0.0013 |
| | 9 | 0.0011 | | 12.727 | 0.0013 |

In Figure (2) we note the curves of chromium (Cr) data based on the results of variogram functions in all directions. These curves between the lag (h) on the X-axis while the variogram function (gamma (h)) on the Y-axis. We show also the average of these functions (red curve of two thetas $\theta=0^\circ \quad \theta=90^\circ$ and black curve of two thetas $\theta=45^\circ \quad \theta=135^o$). We note all these curves have the same behavior in all directions.

The results of experimental variogram function according equation (2) for the second sample (copper data) and compute the average of the variogram function in two directions ($\theta=0^\circ \quad \theta=90^\circ$) that is have the same lag h=1, 2,…, 9 and the average of( $\theta=45^\circ \quad \theta=135^o$) have the same lag h=1.414, 2.828,…,12.727, these results reached in Table 4.

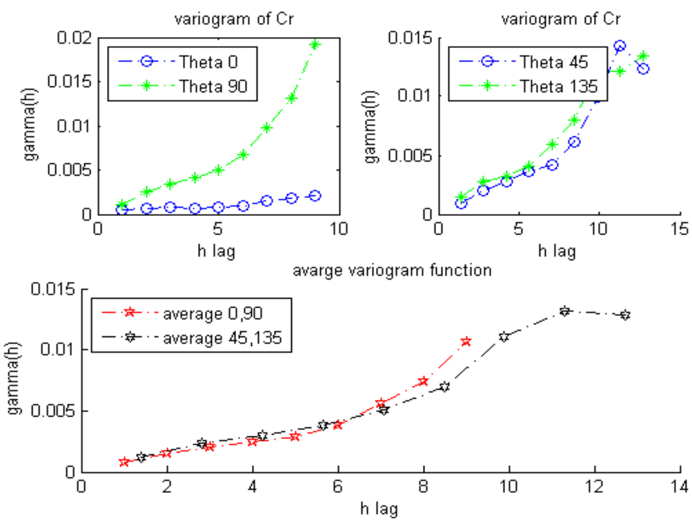


**Fig. 2:** varigram function for chromium data:

**Table 2:** Results variogram function for Copper (Cu)

| direction | $(h)$ | $\gamma(x)$ | direction | $(h)$ | $\gamma(x)$ |
|---|---|---|---|---|---|
| Average of $\theta = 0^{\circ}$ $\theta = 90^{\circ}$ | 1 | 3.6601 | Average of $\theta = 45^{\circ}$ $\theta = 135^{o}$ | 1.414 | 5.6877 |
| | 2 | 6.3479 | | 2.828 | 9.6639 |
| | 3 | 8.0186 | | 4.243 | 11.8746 |
| | 4 | 9.8119 | | 5.657 | 15.1691 |
| | 5 | 12.4275 | | 7.071 | 20.0407 |
| | 6 | 16.2544 | | 8.485 | 27.9453 |
| | 7 | 22.1718 | | 9.899 | 44.3293 |
| | 8 | 32.9978 | | 11.313 | 78.0667 |
| | 9 | 60.5198 | | 12.727 | 209.2687 |

Figure (3) below we show the curves of copper (Cu) data using the results of variogram functions in Table (4) in all directions. These curves on the X-axis the lag (h) while on the Y-axis the variogram function (gamma (h)), and we compute the average of these functions (red curve of two thetas $\theta = 0^{\circ}$ $\theta = 90^{\circ}$ and black curve of two thetas $\theta = 45^{\circ}$ $\theta = 135^{o}$). Also curves of variogram functions in all directions of compass have the same behaviour.

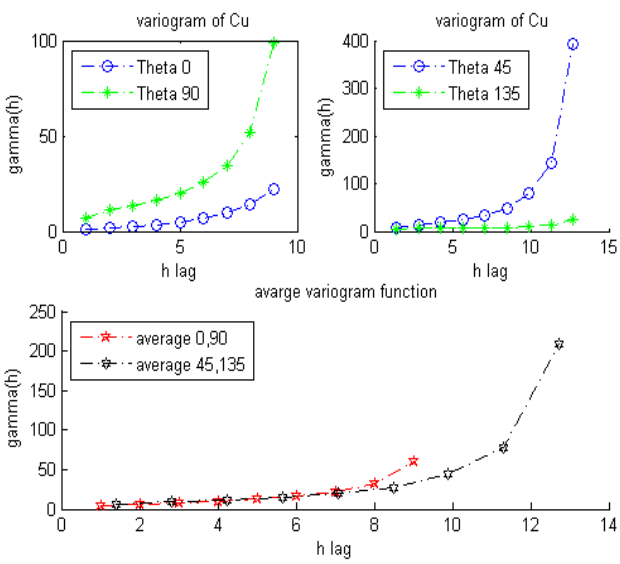


**Fig. 3:** variogram function for Copper metal.

Table (3) contains the data statistics for chromium (Cr) rely on the Figure (2). These data describe the features or parameters the variogram functions such as: Nugget effect, Sill and Range. In other words, we can clearly identify of nugget effect is =0.0007, Range on X-axis is equal (8) and the Sill is =0.0029 for chromium in the two angles (0, 90) while the nugget effect is =0.001, the Range on X- axis is equal (11.31) and Sill is =0.0056 in the two angles (45,135), because it have the same lag of (h). while the data statistics for copper data based on the Figure (3). These data show the parameters of the variogram functions, we can clearly identify the average of variogram for copper data in the two angles (0, 90) have the same lag (h), nugget effect is =3.66, Range is equal (8) on X-axis and the Sill is =12.43, while in the two angles (45,135), have the same lag of (h), the nugget effect is =5.688, the Range on X- axis is equal (11.31) and Sill is =20.04

**Table 3:** Results data statistics for chromium (Cr) and copper (Cu)

| Sample set | Trend (angles) | min | max | mean | median | std | range |
|---|---|---|---|---|---|---|---|
| Chromium (Cr) | (0, 90) | 0.0007 | 0.010 | 0.0041 | 0.0029 | 0.0031 | 8 |
| | (45,135) | 0.001 | 0.013 | 0.0064 | 0.0056 | 0.0046 | 11.31 |
| Copper (Cu) | (0, 90) | 3.66 | 60.52 | 19.13 | 12.43 | 17.96 | 8 |
| | (45,135) | 5.688 | 209.3 | 46.89 | 20.04 | 64.94 | 11.31 |

***Variogram Estimators:***

In variogram estimator, we choose the proposal submitted by the Cressie Hawkins (1980) defined as the equation (16). It is used to obtain the accuracy of the results for both samples data metals in this work. In the first sample data (Cr), it was noted that the curves of experimental variogram function is similarly to the proposed robust matches were found that are similar function in all directions, and when studying the model is nearest to the Gaussian model and the features of variogram estimator are close to features of experimental and the model of covariance function,( nugget effect =0.00266, range = 8, and Sill=0.0207), see Figure (4). While the result of variogram estimator with Cressie Hawkins for the second sample (copper data) seem to be very sensitive to departures from curves of experimental variogram function. And the parameters of variogram are (nugget effect =2.242, range = 8, and Sill=31.45), see Figure (5).

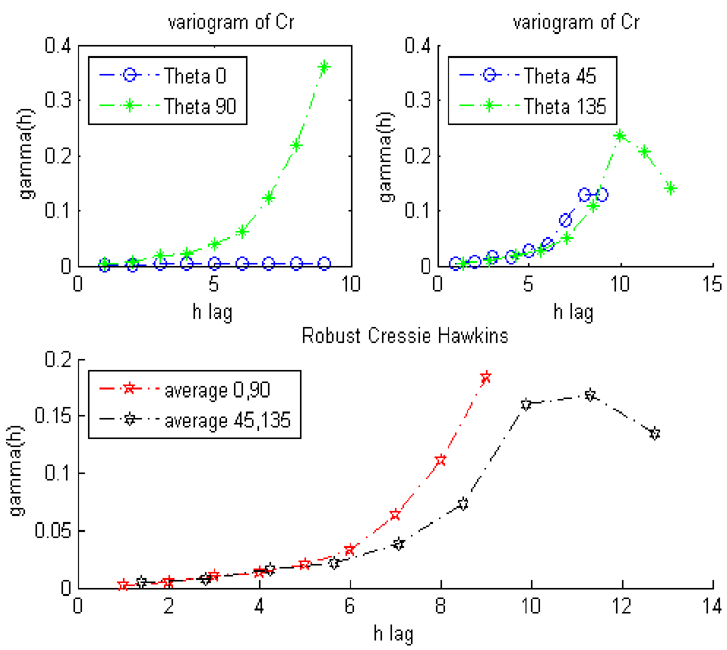


**Fig. 4:** Robust Cressie Hawkins for Chromium data.

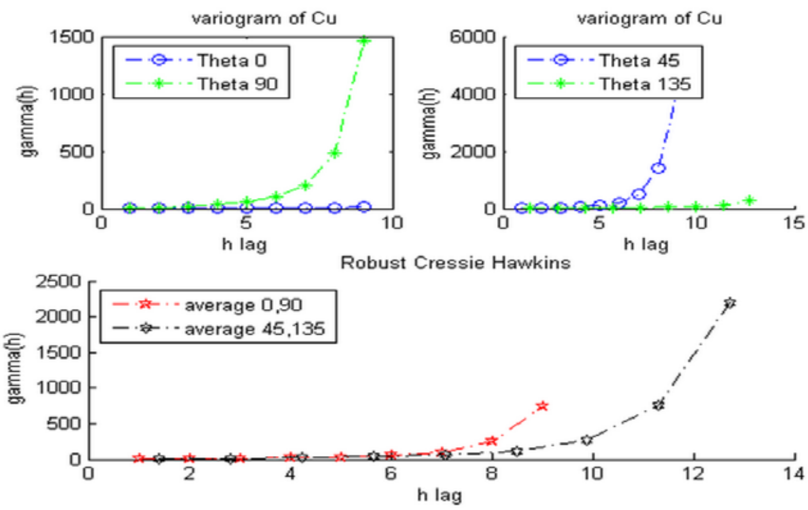


**Fig. 5:** Robust Cressie Hawkins for Copper data.

***Prediction Points:***

The curves of chromium (Cr) data are nearest to the Gaussian model refer to the equation $\gamma(h) = \psi_o \left[1 - \exp\left(\frac{-h^2}{a}\right)\right], h \succ 0$ , the results are used to calculate $\hat{z}(x_o)$ the prediction of variables of five points for chromium metal (4.8336, 4.5558, 4.5073, 4.6796, 5.0011) and calculate the weights $\left(\sum_{i=1}^{n} \lambda_i = 1\right)$ for all five points of prediction and the errors variance are (0.0280, -1.128, 0.2280, 0.1380, 0.1140). The curves of the second sample data of copper data are the best fitted to the Power model that is defined as: $\gamma(h) = \psi_o h^a, 0 \prec \text{h} \prec 2$ , the results are used to predict five points for copper metal and compute $\hat{z}(x_o)$ and we get these predictions (5.6267e+005, 5.8052e+005, 5.8562e+005, 5.7542e+005, 5.6643e+005) with the constraint $\left(\sum_{i=1}^{n} \lambda_i = 1\right)$ for all prediction process. Also the error variance (-0.5000, -1.5000, -1.5000, -0.5000, -0.5000). We note the point's predictions are very small for each sample data.

**Table 4:** Results of prediction for chromium (Cr) and copper (Cu)

| Sample set | Method | $z_1$ | $z_2$ | $z_3$ | $z_4$ | $z_5$ |
|---|---|---|---|---|---|---|
| Chromium (Cr) | Kriging | 4.8336 | 4.5558 | 4.5073 | 4.6796 | 5.0011 |
| | Robust cressie- Hawkins | 0.015 | 0.123 | 0.0356 | 0.1644 | 0.013 |
| Copper (Cu) *e+005 | Kriging | 5.6267 | 5.8052 | 5.8562 | 5.7542 | 5.6643 |
| | Robust cressie- Hawkins | 5.7787 | 6.9006 | 5.5582 | 5.8854 | 5.9347 |

***Conclusion:***

It is clearly seen that there is a big difference between the first data (chromium metal) lies in a regular lattice which is a Gaussian model, and the second set (copper metal) lies in an irregular lattice resembles very much the form of Power model and this indicates that the nature of the data of the metal first (chromium) is different from the nature of the second data (copper), It is by observing the weights, that it was predicted the total weights for each point is intended to equal to 1, this achieves the constraint and, the values of variation in prediction Process it is very small, as well as the values of kriging variance estimator which shows the error in the prediction is very small, the results of robust variogram estimation are very clear through the behaviour of all the curves correspondence between the results and all graphs. In conclusion, we note all these curves for each set of data have the same behavior in all directions, and then the variogram function does not rely on the trend that is mean Isotropy property in the prediction.

## REFERENCES

Cressie, N.A., 1993. ( 2nd Edition). *Statistics for Spatial Data*, John Wiley and Sons, ,New York.

Chiles, J.P. and P. Delfiner, 1999. *Geostatististics,* (1st Edition). (Wiley Series in Probability Publisher*:* Wiley- Interscience*.*

David, J. and M. Genton, 2000. *Variogram Model Selection via Nonparametric Derivative Estimation*. Department of Mathematics, Massachusett Institute of Technology,Cambridge,Massachusetts Mathematical Geology, 32: 3.

Diggle, P.J., R. Menezes and T.l. Su, 2010. *Geostatistical inference under preferential sampling* Journal of the Royal Statistical Society: Series C (Applied Statistics) Article first published online: 9 FEB 2010.

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*. Oxford University Press, New York.

Isaaks E.H, and R.M. Srivastava, 1989. *An introduction to appliedgeostatistics*. Oxford University Press, 561 P. N.Y.

Journel, A.G. And C.J. Huijbregts, 1978. *Mining Geostatistics*.Academic press, London.

Kumar, V. and Remadevi, 2006. *Kriging of Groundwater Levels – A Case Study* Journal of Spatial Hydrology 6(1), Indian Institute of Technolog, Hauz Khas, New Delhi – 110 016, India.

Lark, R.M., 2008. Some results on the spatial breakdown point of robust point estimates of the variogram. Mathematical Geosciences, 40: 729-751.

Lark, R.M., 2009. Kriging a soil variable with a simple non-stationary variance model. Journal of Agricultural Biological and Environmental Statistics, 14: 301-321.

Lark, R.M., D.J. Lapworth, 2012. Quality measures for soil surveys by lognormal kriging. Geoderma, 173-174: 231-240.

Marchant, B.P., N.P.A. Saby, C.C. Jolivet, D. Arrouays, R.M. Lark, 2011. Spatial prediction of soil properties with copulas. Geoderma, 162: 327-334.

Matheron, G., 1963.. *Principles of Geostatistics,.Economic Geology,* 58: 1246-1266.

Meyers, D.E., 1994. *Spatial interpolation: an overview*. Geoderma, 62: 17-28.

Mingoti,S.A. and G. Rosa, 2008. *A note on robust and non-robust variogram estimators*. Departamento de Estatística da Universidade Federal de Minas Gerais-UFM, Rev. Esc. Minas vol.61 no.1 Ouro Preto Jan./Mar. 2008, Print version *ISSN 0370-4467.*

Ripley, B.D., 1981 .*Spatial statistics* , New York, John Wiley.

Rudolf, D., 2003. *Introduction to geostatistics andvariogram analysis.* Vienna University of Technology.

Stein, M.L., 1999. *Interpolation of Spatial Data*: *Some Theory for Kgrigin,* Springer. NewYork.

Wim C.M. and P.C. Jack, 2002. *Kriging for Interpolation in Random Simulation* Department of Information Management, Tilburg University (KUB), Postbox 90153, 5000 LE Tilburg, The Netherlands Written: 28 June 2002 .Printed: 7 November 2003.

Zidek, J.V. and N.D. Lee, 1992. *Interpolation with Uncertain spatial covariance : A Bayesian alterative to kriging* . journal of multivariate analysis, 43(2): 351-374.