

Testing the Exactitude of Estimation Methods in the Presence of Outliers: An accounting for Robust Kriging

Shahrokh Paravarzar^{*1}, Nasser Madani², Abbas Maghsoudi³ and Peyman Afzal⁴

1. Department of Mining and Metallurgical Engineering, Amirkabir University of Technology, 1359767191 Tehran, Iran

2. Advance Mining Technology Center (AMTC), Santiago, Chile

3. Department of Mining and Metallurgical Engineering, Amirkabir University of Technology, Tehran, Iran

4. Department of Mining Engineering, Faculty of Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran

Received 3 February 2014; accepted 24 March 2014

Abstract

The estimation of gold reserves and resources has been of interest to mining engineers and geologists for ages. The existence of outlier values shows the potential worth of the deposits and is not a result of human or technical error. The presence of these high values causes a pseudo dramatic increase in the variance estimation of the economic value when applying conventional methods such as kriging. Conventional approaches such as replacing the upper limit (capping), log-normal and median-indicator kriging have resolved some of the problems associated with estimation techniques but do, however, still pose numerous drawbacks. The Zarshuran Gold Deposit is one of the proposed gold deposits in Iran that offers a great opportunity for the use of estimation procedures. Robust kriging is a method that is able to reduce the allocated pseudo kriging weights of the blocks surrounded with high value data, and correct them in a manner that not only enhances the accuracy of the estimation but also reduces the estimation variance to a reasonable level. In this method, the estimation can be implemented in the presence of outliers and does not need to eliminate, reduce or transfer the data to another space (normalization). In this paper an attempt was made to estimate the amount of gold in ore using the robust kriging method compared with the capping high value method. Results showed that the robust kriging method has a high efficiency and is able to run estimations accurately and with high precision due to the high value reduction.

Keywords: Robust Kriging, Lognormal Kriging, Zarshuran, Gold.

1. Introduction

Mineral resource estimation has long been of interest to mining engineers and geologists. Geostatistical methods such as kriging are very applicable approaches that can be used in the field of grade estimation, including gold and precious metals. However, the presence of high values on one hand can indicate the economic value of the deposit and on the other hand lead to an overestimation of the block [1]. Conventional methods can provide a reliable estimation of deposits and are helpful tools for the mining industry. The kriging methods can be divided into two categories; first is the capping of high values [2]. In this method, a threshold is considered as a high value and the grades more than the threshold decline toward this specified threshold. Selection of the values is difficult and in some cases is experimental. There is no strong mathematical or geological concept behind the capping method [3]. An advantage of this method is a reduction in the variance of estimation, however, the most important problem is that miners are interested in the omitted information which is not made available [4].

The second part involves estimating an accurate value in the presence of high value data without reducing or eliminating their effects using, for example, log-normal kriging (LK) [5] and median indicator kriging (MIK) [6]. Each of these methods has its own pros and cons. Lognormal kriging has been shown to be extremely sensitive to the sill of the variogram and can overestimate values in the case of log normality assumption. Additionally, it is not perfectly verifiable if the variogram model used is incorrectly chosen when the real sill is unknown [4, 7]. IK appears to be more appropriate than Lk in dealing with outliers although final estimates are sensitive to the choice of thresholds and assumptions used to extrapolate the cumulative distribution at the upper tail. IK also requires a correction for change of support since IK provides the cumulative distribution function (CDF) of the grade at sample support; this cumulative density function needs to be corrected for the blocks. The final change of support can be an additional source of bias in the final estimation. LK merely mitigates the influence of the high values during the kriging steps. However, the entire problem remains and is left to be dealt with in other steps. LK also requires that it be converted back to the original scale, which brings back the issue of dealing with outliers. As an alternative approach to LK and MIK, robust kriging (RoK) is based on robust statistics that are able to reduce the high value effect in the presence of outliers and create

*Corresponding author.

E-mail address (es): Paravarzar.sh@aut.ac.ir

reliable estimations. Hawkins and Cressie, (1984) [8] presented this method for the first time and Costa (2003) [7] applied it for professional use thereafter. In this article, this method was incorporated to estimate the gold grade in the Zarshuran Gold Deposit and the results were compared with conventional methods including LK and capping.

2. Methodology

2-1. Robust kriging

The methodology and fundamental of robust kriging are as follow:

Consider a stationary and ergodic random function (RF) $Z(x)$ defined as R^n where $Z(x_\alpha)$ exists for each set of random variables. For a random function $Z(x)$ mean and variance in a stationary position is defined respectively as [7]

$$E\{Z(x)\} = m(x) \quad (1)$$

$$Var\{Z(x)\} = E\{[Z(x) - m(x)]^2\} \quad (2)$$

Defining an estimator for $Z(x)$ over a domain $V(x_0)$ (point or block) is an objective that reduces the effect of high value data in random variable distribution $Z(x_\alpha)$; this estimator can be written as follows:

$$Z_V^* = \sum_{j=1}^{n-1} X_\alpha Z^{(e)}(x_\alpha) \quad (3)$$

Where X_α is a kriging estimation of weights and $Z^{(e)}(x_\alpha)$ is a RV defined by editing the original $Z(x_\alpha)$ to obtain the estimated value $Z_V^*(x_0)$ the computation starts with drawing the robust variogram and fitting the model. The next step is using the produced variogram for calculating the kriging weights by using all samples in each sampling location point according to:

$$Z_{(\alpha)}^* = \sum_{\beta \neq \alpha}^1 \lambda_{\alpha\beta} Z_\beta \quad (4)$$

Each $Z_{(\alpha)}$ sample will be replaced by winsorized values and equations:

$$Z_{\alpha}^{(e)} = \begin{cases} \tilde{Z}_{(\alpha)} + c\sigma_{K\alpha} & \text{if } Z_{\alpha} - \tilde{Z}_{(\alpha)} > c\sigma_{K\alpha} \\ Z_{\alpha} & \text{if } |Z_{\alpha} - \tilde{Z}_{(\alpha)}| \leq c\sigma_{K\alpha} \\ \tilde{Z}_{(\alpha)} - c\sigma_{K\alpha} & \text{if } Z_{\alpha} - \tilde{Z}_{(\alpha)} < -c\sigma_{K\alpha} \end{cases} \quad (5)$$

Cressie and Hawkins (1980) [9] suggested the use of c values between 1.5 to 2.5 which controls the high value of $\sigma_{K\alpha}$ that is the kriging standard deviation. Here attention must be given because the high value of c causes a lower impact on the distribution function of the data. Values lower than 1 cause a smoothing effect on the data distribution. In this stage, after editing the data, (Eq.5) the kriging method (simple or ordinary) can be applied using a robust variogram. As a whole, in this method after obtaining the specific edited kriging weights from Eq.5 (this editing is more for high value data) the conventional kriging method was applied (simple and ordinary) in the stationary estimation [1, 7].

2-2. Log-normal kriging

The theory of kriging for normal and log normally distributed data has been discussed in a variety of resources. Hence, the intention of this section is to review this kriging as a conventional geostatistical method [5, 10 and 11].

2-2-1. Ordinary lognormal kriging (OLK)

$Z(x)$ has a two parameter lognormal distribution (three parameters of lognormal is a simple extension of the two parameters case). Therefore

$$\beta(x) = \ln Z(x) \sim N(\mu, \sigma^2) \quad (6)$$

The ordinary lognormal kriging estimator

$$Z_{OLK}^*(x_0) = \exp\left[\sum \lambda_{OLK_i} \beta_i + \frac{\sigma_{OLK}^2}{2} + \tau\right] \quad (7)$$

Weighting factors λ_{OLK_i} obtained by solving the following kriging system

$$\sum_i \lambda_{OLK_i} \rho(x_i, x_j) + \tau = \rho(x_j, x_0) \quad j = 1, \dots, n \quad (8)$$

$$\sum_i \lambda_{OLK_i} = 1 \quad (9)$$

It is obvious that the value of σ_{OLK}^2 and τ in the system depend linearly on the value of the sill of variogram of the logs [$\rho(0) = \sigma^2$], however the weights λ_{OLK_i} do not.

2-2-2. Simple lognormal kriging (SLK)

μ is a known mean of $Y(x)$ and $\rho(h)$ is the covariance function. The simple lognormal kriging is:

$$Z_{SLK}^*(x_0) = \exp\left(\sum \lambda_{SLK_i} Y_i + \sigma_{SLK}^2/2\right) \quad (10)$$

Where kriging weights λ_{SLK_i} , and kriging variance σ_{SLK}^2 are obtained by solving the following equation:

$$\sum \lambda_{sk_i} \rho(x_i, x_j) = \rho(x_j, x_0) \quad j = 1, \dots, n \quad (11)$$

2-3. Capping

The high values which are usually interpreted as outliers are the intrinsic properties of deposits with long tailed distribution in terms of the geological setting and ore formation. This issue has been observed in gold deposits that cause a highly skewed distribution in the overestimation or underestimation of variograms [4]. Therefore, one of the conventional methods in the mining industry for dealing with highly skewed data is capping the data by choosing a suitable threshold. This procedure, however, is a trial and error process. In this method the outliers, the assumed capping value, the coefficient variation and variance are omitted. As a result high fluctuations in the variograms reach an acceptable result and make the process more efficient [3].

3. Case study

The Zarshuran Deposit is located at 36° 43.390' N, 47° 08.219' E, and is 42 km north of Takab in the West Azerbaijan Province. (Fig. 2)[12]. It is well known for arsenic mining, orpiment and realgar which have been mined for hundreds of years. It is now is a gold prospect with a reserve of 88t Au at a grade of 7.9 g/t [13]. The geology and genesis of the Zarshuran

deposits and mineralization were described by Mehrabi [14], Asadi et al. [15] and Mehrabi et al. [16].

3-1. Geological setting

According to Iranian geological and structural deviation [17], the Zarshuran Deposit is part of the Azerbaijan – Alborz Zone (fig.1). The main geological feature of the Zarshuran area is a thick Precambrian basement covered with Quaternary travertine rocks with a shallow dip across the lithological sequence. Gold is spread throughout the deposit along with high levels of arsenic and sulphide. The host rock is Chaldagh sandy marble. The carbonated organics make gold-organic carbonates which results in a high grade. There are numerous similarities between the Zarshuran and Carlian deposits that can be described as a Carlian type. The mineral assemblages include orpiment, realgar, stibnite, getchellite, cinnabar, As-Au-bearing pyrite, base metal sulphides and sulphosalts, thallium minerals and micron with angstrom-sized gold. An exploration program has indicated the presence of four types of mineralization. Steeply dipping mineralization (70 degree dip or greater) includes two ore types: a) black gouge with orpiment and jasperoid clasts and b) jasperoid and massive orpiment. Flat lying mineralization (45 degree or flatter) includes two ore types: a) mineralized breccias and tectonites and b) sandy or powdery Chaldagh marble. Most of the mineralization, approximately 60% of the volume, occurs in flat lying carbonaceous decalcified marble. The highest gold grade mineralization, approximately 40% of the volume, appears to occur in the steeply dipping black gouge zone [13].

3-2. Statistical Analysis

For the purpose of modeling the spatial structure (variogram model) and estimation, the data set consists of 27 diamond drilled holes from core cut samplings. The Anglo American Company analyzed the samples for assay quantification. The data is composed of 4,305 samples for statistical and geostatistical analysis. The results of the classical statistic measurements are presented in table 1 in ppm units for the Au grade. In the next step, the first and second order moment was quantified and the results showed a stationary domain. We then applied the geostatistical method for stationary spaces.

3-3. Variography

Investigation of the variability of the regional variables are considered to be the most essential geostatistical studies that can be carried out by the Eq. 7 [18]:

$$2\gamma(h) = \frac{1}{N(h)} \sum_{n=1}^h [z(u+h) - z(u)]^2 \quad (12)$$

The variogram of random variables has to be determined in order to obtain the major and minor anisotropy axis, therefore, the variogram is calculated

at 0, 45, 90, 135 and 180 degrees. The major anisotropy is on $Az=0^\circ$ and $dip=0^\circ$ with the minor axis at $Az=0^\circ$ and $dip=90^\circ$. Table No.2 shows the variogram parameters and the fitted model for different data sets. According to the variogram figures, the variograms with the original data set have an extraordinary fluctuation and are not robust. With the exception of the edited data set with $c=1.5$ not only do the variations in the variogram remain correct the data variance reduces properly as well.

3-4. Ordinary and simple robust kriging

After a spatial analysis of the original data set, in order to carry out comparisons of the different methods for estimating the data in a grid measuring $10m \times 10m \times 12m$, block size dimensions were considered over the domain. Editing the original data using different c values (1.5, 1.75, 2.25, and 2.5) is a critical step in robust kriging. Robust kriging was applied using simple and ordinary kriging. Then the estimated data obtained was compared to the original compositing data set using the cross-validation method for assessing the estimated validation at different levels of C . Table 3 shows the cross validation results.

For lognormal kriging it is necessary to transform the original data sets to normal score distribution by logarithmic transformation. The experimental and fitted models and their parameters are shown in Figure 2 and Table 2. Next, by using lognormal variogram ordinary and simple kriging was applied to the data sets.

Capping data is a conventional method for dealing with outliers and was used by replacing the data at a threshold of 30 ppm in the data set. The reason for choosing this number is purely experimental and was done in order to validate each method. The results of the above procedure were analyzed using cross validation technique and are shown on table 3.

4. Discussion

According to the results (see table 3) there are some significant differences between the correlation coefficient values for the three methods of estimation with the highest value belonging to robust kriging. The Mean Square Error (MSE) and Mean Absolute Error (MAE) are lower for robust kriging compared to other methods (table. 3). The lowest correlation coefficient can be seen in the case of ordinary kriging with capping data whereas the highest belongs to robust kriging with SK and OK where $c=1.5$ (see table 3, no.1 and no.5). Between these two statements no.5 with SK shows the lowest total error. Figure 3 illustrates the scatter plot for cross validation results of the estimation for RoK vs original data where $c=1.5$ using simple kriging, capping data vs original data.

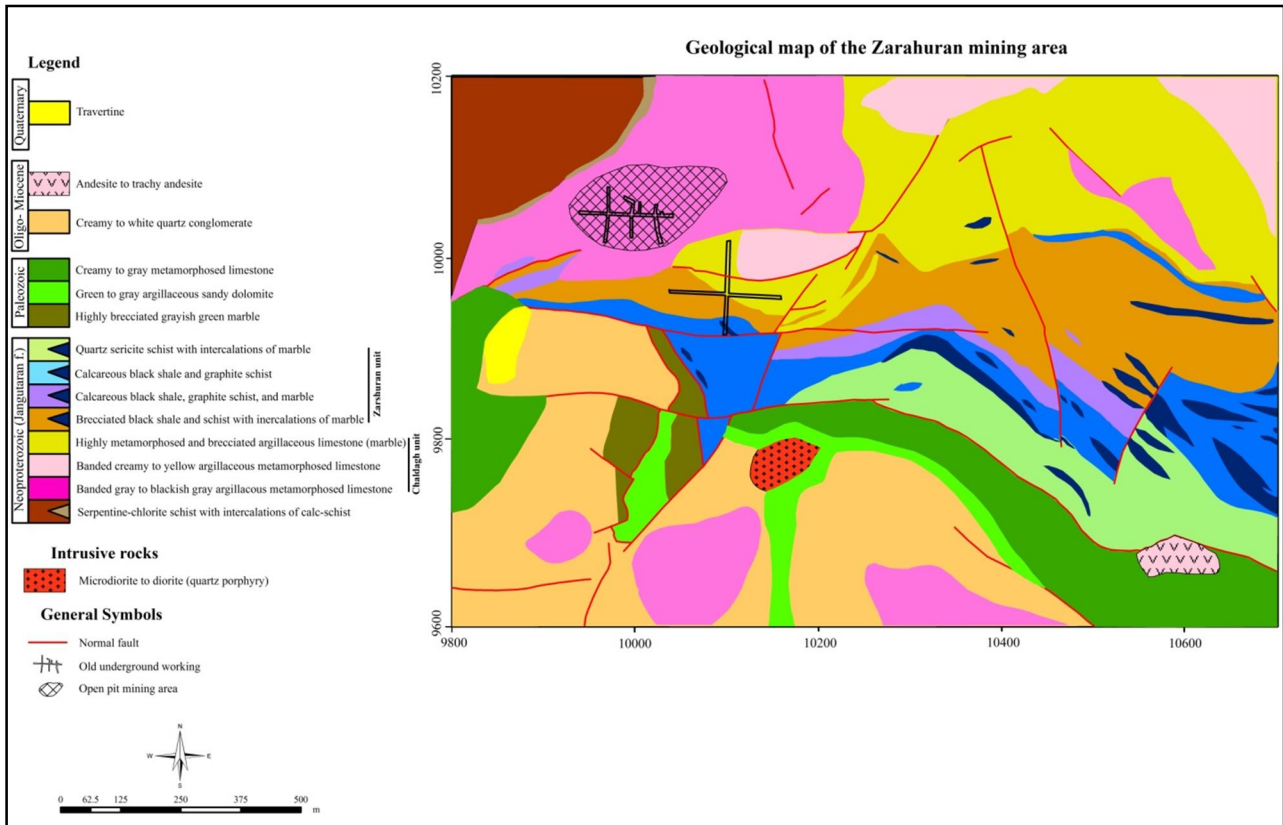


Fig. 1. Geological map (Scale 1:25,000) of the Zarshuran Gold Deposit.

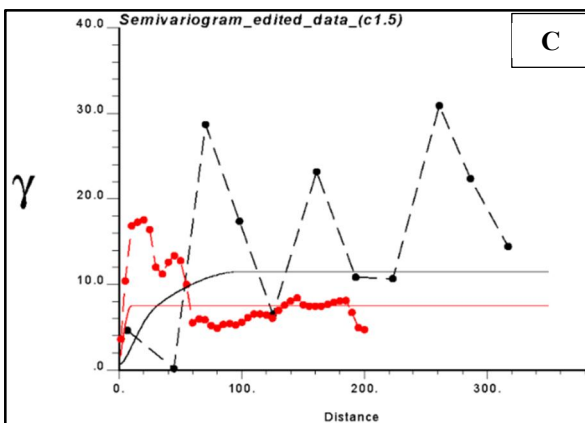
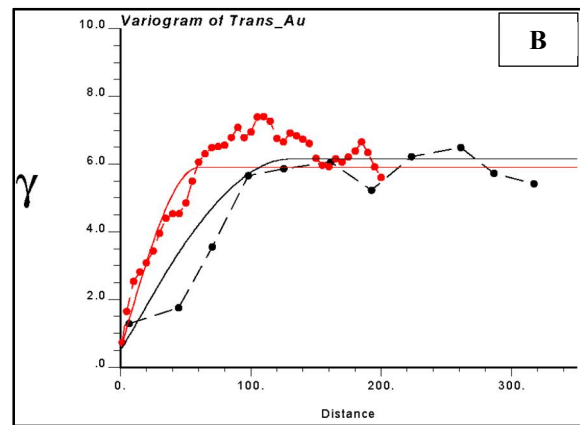
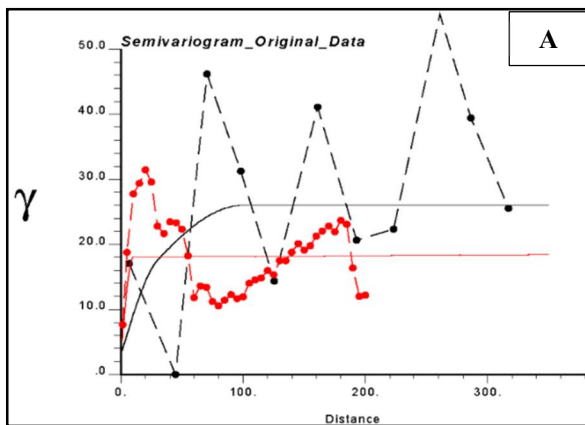


Fig. 1-A) Variogram showing original data B) Variogram of lognormal data C) Variogram of corrected and weighted data for $c=1.5$ in robust kriging case. Experimental variogram values are illustrated by dotted lines; variogram model illustrated by solid line(s) (black solid line(s) for major and red solid line(s) for minor anisotropy) for assayed total gold grades, for the entire deposit.

Table 1. Statistics for Gold Grade (ppm) from Dataset

Data set	C	No.ofData	Mean(ppm)	St. Dev	Min(ppm)	Max(ppm)	C.V
Original	-	4035	1.20	4.78	0.00	101.16	3.98
Edited	1.5		0.96	3.41	0.00	84.43	3.53
	1.75		0.98	3.44	0.00	84.89	3.52
	2.25		0.98	3.44	0.00	84.89	3.52
	2.5		1.01	3.54	0.00	86.27	3.51
Lognormal	-		-2.65	2.38	-5.30	4.62	Undefined
Capping	-	1.09	3.47	0.00	30.00	3.18	

Table 2. Model Parameters for Different Dataset

Data type	C value	Var.Direction	Dip	Azimuth	Var.type	Rang(m)	Nugget(% ²)	Sill(% ²)
Original	-	Horizontal	0	0	Spherical	100	0.5	20
		Vertical	-90	0	Gaussian	20	0.5	7
Edited	1.5	Horizontal	0	0	Spherical	100	0.5	7
		Vertical	-90	0	Gaussian	30	0.5	4
	1.75	Horizontal	0	0	Spherical	100	0.5	8
		Vertical	-90	0	Gaussian	30	0.5	5
	2.25	Horizontal	0	0	Spherical	100	0.5	7.5
		Vertical	-90	0	Gaussian	30	0.5	4.5
	2.5	Horizontal	0	0	Spherical	100	0.5	7.5
		Vertical	-90	0	Gaussian	30	0.5	4.5
Lognormal	-	Horizontal	0	0	Spherical	130	0.5	5.4
		Vertical	-90	0	Gaussian	60	0.5	0.25
Capping	-	Horizontal	0	0	Spherical	110	0.5	12
		Vertical	-90	0	Gaussian	60	0.5	2

Table 3. Cross-validation results and errors for different methods

No	Estimation method		C value	Correlation	Error				
					SDE	MED	ME	MAE	MSE
1	RoK	OK	1.5	0.8731	2.5488	-0.0010	0.2390	0.5613	6.5520
2			1.75	0.8727	2.5389	-0.0010	0.2267	0.5631	6.4959
3			2.25	0.8727	2.5389	-0.0010	0.2267	0.5631	6.4959
4			2.5	0.8715	2.5144	-0.0010	0.1974	0.5676	6.3599
5		SK	1.5	0.8731	2.5517	-0.0005	0.2490	0.5611	6.5716
6			1.75	0.8727	2.5389	-0.0010	0.2267	0.5631	6.4959
7			2.25	0.8728	2.5418	-0.0010	0.2368	0.5630	6.5151
8			2.5	0.8715	2.5171	-0.0010	0.2078	0.5637	6.3777
9	LK	OK	-	0.8294	2.6942	-0.0059	0.0033	0.7021	7.2568
10		SK	-	0.8296	2.6935	-0.0164	-0.0016	0.7051	7.2532
11	Capping	OK	-	0.7926	3.0590	-0.0022	0.1126	0.6508	9.3678
12		SK	-	0.7927	3.0585	-0.0039	0.1126	0.6523	9.3647

According to the results, significant fluctuation is not clear in the RoK method using different coefficient c values. However, it is better to consider the lower value for the c coefficient value for pseudo modification prevention [7, 9]. The c value depends on the user application, purpose and data description. It is obvious that the estimation variance is much lower using the capping method and LK compared to RoK. As a result improper estimation from the economic sector can cause irreparable compensation from an economic point of view. Figure 3 demonstrates the estimation maps resulting from the three methods. The LK method has an additional interstitial result in comparison to the other estimations.

5. Conclusion

In this paper the robust kriging method is applied to the estimation of the Zarshuran Gold Deposit. The method was compared with other conventional methods of estimation such as lognormal kriging (LK), kriging with capping data and median indicator kriging (MIK). Results obtained from robust kriging are more reliable after considering the cross validation for each method in the case of SK and OK. In the estimation process, the accuracy and optimized variance of estimation are two significant parameters that should be considered simultaneously. For further steps it is suggested that the robust kriging equations be in modifying order to correct the variance estimation in an appropriate manner and reduce the variance to an optimized level as much as possible.

Acknowledgement

The authors wish to acknowledge Mr.Qavanloo, Executive Manager and Mr.Yeganlie, Mine Supervisor of IMIDRO (Zarshuran Gold Mine) for authorizing the use of logging data for the Zarshuran Gold Mine.

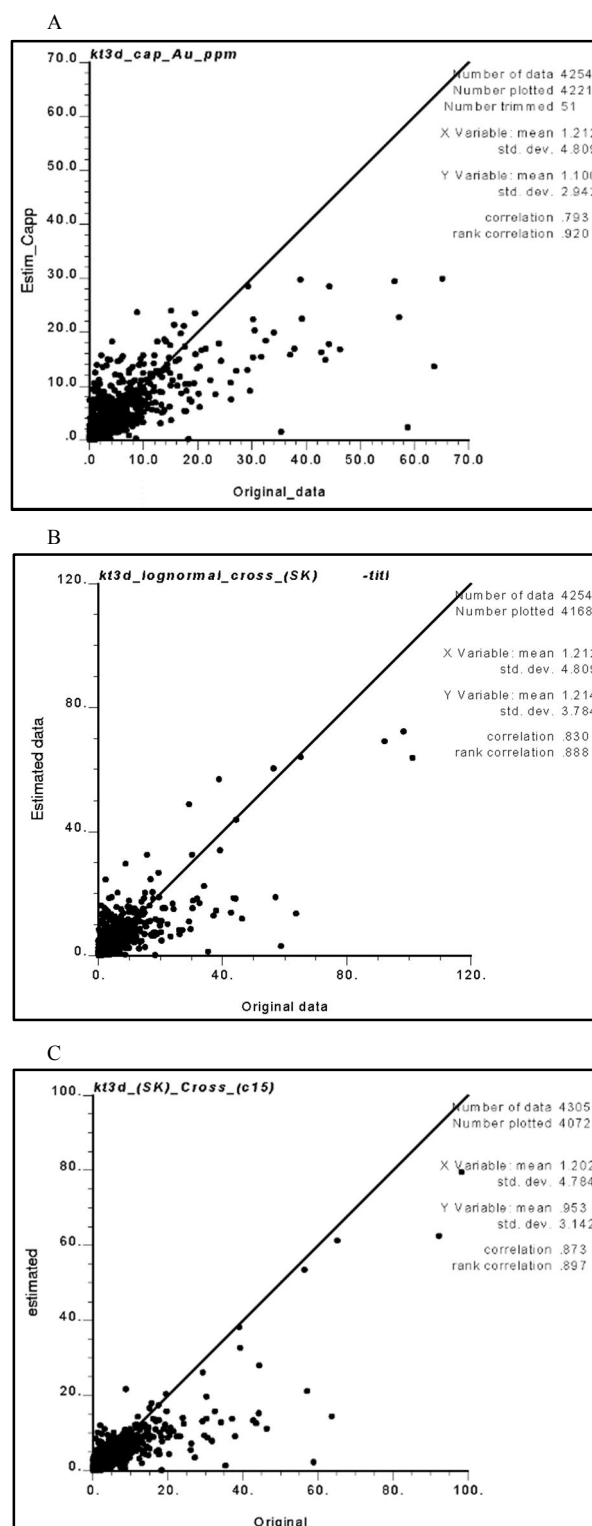


Fig. 2. Scatter and cross validation plot for different methods
A) Capping data B) Lognormal C) RoK ($c=1.5$)

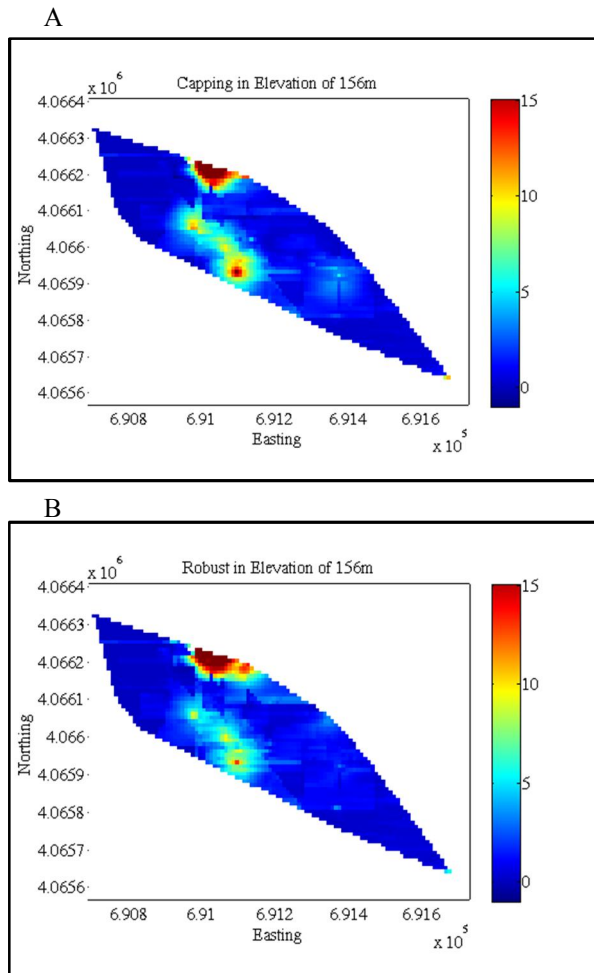


Fig. 3. (A) Kriging maps for capping (B) robust kriging

References

- [1] Chiles, J.P., Delfiner, P., 2012: Geostatistics: Modelling Spatial Uncertainty. Wiley, New York, PP. 695-697.
- [2] David, M., 1988: Handbook of applied advance geostatistical ore reserve estimation. Elsevier Science, Netherlands, P.5
- [3] Sinclair, A.J. and Blackwell, G., 2002, Applied Mineral Inventory Estimation: The University of British Columbia, Queen's University, Cambridge University Press, The Edinburgh Building, and Cambridge CB2 2RU, UK.
- [4] Rivoirard, J., Demange, C., Freulon, X., Lecureuil, A., and Bellot, N., 2012, A Top-Cut Model for Deposits with Heavy-Tailed Grade Distribution: International Association for Mathematical Geosciences
- [5] Journel, A.G., 1980, Non-parametric estimation of spatial distributions: Journal of Mathematical Geology Vol.15, No.3, PP. 445-468.
- [6] Journel, A.G., 1982, The lognormal approach to predicting local distribution of selective mining unit grades: Mathematical geology. 12, no.4, PP. 285-303.
- [7] Costa, J.F., 2003, Reducing the Impact of Outliers in Ore Reserves Estimation: Mathematical geology, Vol.35, no.3, PP. 323-345.
- [8] Hawkins, D. M., and Cressie, N., 1984, Robust kriging – A proposal: Mathematical geology, Vol.57, no.1, PP. 97-109.
- [9] Cressie, N., and Hawkins, D. M., 1980, Robust estimation of the variogram: I: Mathematical geology, Vol.57, no.1, PP. 97-109.
- [10] Dowd, P., 1982, Lognormal kriging-The General Case.: Mathematical geology, Vol.14, no.5, PP. 474-500.
- [11] krige, D. K., and Margi, E.J., 1982, Geostatistical case-studies of the Advantages of lognormal De Wijnan Kriging with Known Mean for a base metal Mine and a Gold Mine: Mathematical geology, Vol.14, no.6, PP. 547-556.
- [12] Asadiharoni, H., Sides, E., Ngonzi, K., 2000, Modeling of the Zarshuran Gold Deposit, IRAN: International Archives of Photogrammetry and Remote Sensing. Amsterdam, Vol. 53, Part B7. PP. 82-89.
- [13] Anglo exploration, 2001, Zarshuran gold deposit, final report, Ministry of Industry and Mine.
- [14] Mehrabi B., 1997, Genesis of the Zarshuran gold deposit, NW Iran. Unpublished Ph.D. Thesis, University of Leeds.
- [15] Asadi H.H., Voncken J.H.L., Kuhnel R.A., and Hale M., 2000, Petrology, mineralogy and geochemistry of the Zarshuran Carlin-like gold deposit.: Ibid, Vol.35, PP. 656-671
- [16] Mehrabi B., Yardley B.W.D, and Cann J.R., 1999, Sediment hosted disseminated gold mineralization at Zarshuran. NW Iran. Mineralum Deposita, Vol.34, PP. 673-969.
- [17] Nabavi, M. H., 1355, An introduction to Iran geology, G.S.I.
- [18] Deutsch, C.V., Journel, A.G., 1992, GSLIB: Geostatistical Software Library and User's Guide." Oxford University Press, New York, P.340.