

## A Hybrid Strategy to Optimize the Search Ellipsoid Dimensions: Case Study from Anomaly No 12A Iron Deposit in Central Iran

*Örnek Tarama Elipsoid Boyutlarını Optimize eden Melez bir Stratejinin geliştirilmesi: Orta İran'da 12A Nolu Demir Yatağında Örnek bir Uygulama*

Marzieh SHADEMAN KHAKESTAR<sup>1</sup>, \*Hossein HASSANI<sup>1</sup>, Saeed ANGORANI<sup>2</sup>

<sup>1</sup>Department of Mining, Metallurgical and Petroleum Engineering, Amirkabir University of Technology, 424, Hafez Ave., P.O. Box 15875-4413, Tehran, 15914 Iran.

<sup>2</sup>School of Mining Engineering, University of Tehran, Tehran, Iran.

Geliş (received) : 25 Aralık (December) 2010

Kabul (accepted) : 19 Mart (March) 2011

### ABSTRACT

The definition of search volume or kriging neighborhood in kriging estimators is an exercise in compromises. Determination of this neighborhood significantly influences the outcome of the kriging estimate. The main criteria used in the evaluation of a particular kriging neighborhood include the kriging variance, the number of non-estimated blocks, the cumulative sum of the kriging negative weights, and the slope of the regression of the real block grade to the estimated block grade. The performance of the above methodology is noticeably influenced by the radius of the search volume. This paper presents a new strategy to find the optimum value of the search radius. Using experimental data, we develop a neural simulator that would predict, accurately enough, the values of kriging variance, the number of non-estimated blocks and cumulative sum of kriging negative weights for a given search radius. The simulator is then used as the objective-evaluator in a numerical optimization code based on the Complex direct search method which would find the search radius corresponding to the optimum values of the evaluation criteria. Having generated multiple solution sets in multiple runs of the algorithm, the slope of the regression is then used to prioritize the solutions and to spot the most viable choice. The applicability and efficiency of the proposed strategy is demonstrated using anomaly No.12A iron deposit, located in Bafgh in central Iran, as a case study.

**Key Words:** Complex method, kriging, neural networks, optimization, search ellipsoid dimensions.

### ÖZ

*Kriglemede kestirim komşuluğunun ya da örnek tarama bölgesinin uygun bir şekilde tanımlanması gerekir. Bu komşuluğun belirlenmesi kestirim değerini oldukça etkiler. Krigleme komşuluğunun belirlenmesinde kullanılan ana kriterler; krigleme varyansı, kestirilmeyen blokların sayısı, negatif ağırlıkların birikimli toplamı ve gerçek blok tenörünün kestirilen blok tenörüne karşı çizilen regresyonun eğimidir. Bu yaklaşımın performansı örnek tarama bölgesinin çapından büyük ölçüde etkilenir. Bu yazı, örnek tarama çapının optimum bir şekilde belirlenmesine yönelik olarak yeni bir yaklaşım geliştiriyor. Deneysel veriler kullanılarak, örnek tarama çapına ilişkin krigleme varyansını,*

---

H.Hassani

E-mail: hhassani@aut.ac.ir.

kestirilmeyen blokların sayısını ve krigleme negatif ağırlıklarının birikimli toplamını hassas bir şekilde kestiren bir sinir ağı simülatorü geliştirilmiştir. Simülator daha sonra sayısal optimizasyon kodunda objektif bir değerlendirici olarak kullanılmıştır. Bu kod, değerlendirme kriterlerinin optimum değerlerine karşılık gelen örnek tarama çapını bulan karmaşık doğrudan tarama yöntemine dayanmaktadır. Algoritmanın bir çok kez çalıştırılması suretiyle çok sayıda çözüm kümesi üretildikten sonra regresyonun eğimi en uygun çözümü bulmak amacıyla kullanılmıştır. Önerilen stratejinin uygulanabilirliği ve verimliliğini ortaya koymak amacıyla Orta İran'da Bafgh'de bulunan 12A nolu demir yatağında örnek bir inceleme sunulmuştur.

**Anahtar Kelimeler:** Complex method, kriging, neutral networks, optimization, search ellipsoid dimensions.

## INTRODUCTION

This paper presents a methodology for optimizing search radius using criteria by which a particular kriging neighborhood is evaluated. These criteria involve kriging variance, the number of non-estimated blocks, cumulative sum of kriging negative weights, and the slope of the regression of the true block grade on the estimated block grade. All mentioned criteria except the number of non-estimated block were introduced by Vann et al. (2003). The authors introduce 'the number of non-estimated blocks' as a new criterion and consider it to define the best search radius. If the shape of kriging neighborhood is determined ellipsoid, the dimensions could be considered as search radius. De-Vitry (2003) used kriging variance, the slope of the regression and kriging negative weights and plotted their statistics found optimum search ellipse dimensions by determining where the dimensions of the search ellipse would not significantly improve the estimate. This was where increasing the dimensions of the search ellipse would not significantly increase the slope of the regression between 'actual' and 'estimated' grades and also decrease the kriging variance and increase the number of negative kriging weights.

In this paper the values of above-mention criteria were computed, for different dimensions of search ellipsoid. An Artificial neural network was used as a function approximation tool to find the relation between the ellipsoid dimensions and the criteria. A multiobjective function that involves comparing and making decisions about different objectives with different order of magnitude was obtained. The function was optimized by complex direct search method to find the optimum ellipsoid dimensions, by using weighted sum method on normalized objects.

## DEFINITION AND EVALUATION CRITERIA

The criteria to consider when evaluating a particular ellipsoid dimensions are:

1. The kriging variance;
2. The number of non-estimated blocks;
3. The cumulative sum of kriging negative weights; and
4. The slope of the regression of 'true' block grade on the 'estimated' block grade.

### Kriging Variance

Kriging is an estimation procedure that minimizes the estimation variance. The expression for the minimum estimation variance (Eq. 1) also called the kriging variance (KV), is (Vann et al., 2003):

$$\sigma_{ok}^2 = Var(z_v^* - z_v) = \sum_{i=1}^N \lambda_i \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) + \mu \quad (1)$$

Where  $\sigma_{ok}^2$  is kriging variance,  $\lambda_i$  is kriging weights,  $\bar{\gamma}(x_i, V)$  is the average values between a sample and the block to be estimated and  $\bar{\gamma}(V, V)$  is the average gamma value within the block to be estimated. Here, KV was computed for each ellipsoid dimensions.

### The Number of Non-Estimated Blocks

A neighborhood that is too restrictive can result in serious conditional biases and some areas can not be estimated. In mining block estimation it is important to estimate blocks as much as possible. Since different ellipsoid dimensions change the number of non-estimated blocks (NEB), this criterion is considered to determine the best dimensions of ellipsoid. Here the NEB was computed for each ellipsoid dimensions.

### Kriging Negative Weights

Negative weights are a peculiarity of certain data geometries of kriging systems combined with a high degree of continuity (including a low to negligible nugget effect) in the variogram model. In these circumstances, a 'screen effect' can be expected and at some distance negative weights will be observed (Sinclair and Blackwell (2002)). The distance we need to search before negative weights are encountered progressively increases as the effective nugget effect increases. In the case of 'pure nugget' every sample found gets equal weight (1/N) no matter how far we search (Vann et al., 2003).

Depending on the variogram and the amount of screening, the negative weights can be significant. There is nothing in the OK algorithm that alerts the kriging system about the zero thresholds for weights. Also, negative weights when applied to high data values may lead to negative and nonphysical estimates (Deutsch, 1996). Szidarovszky et al. (1987) considered an additional constrain in the kriging process and presented a numerical algorithm which generates optimal nonnegative weights from a set of sample points. Two other algorithms were proposed by Froidevaux (1993) and Journel and Rao (1996) for correcting negative weights. But Vann et al. (2003) advised against modified kriging algorithms that adjust negative weights or set them to zero. Since such approaches will assure conditional bias. If only a small proportion of total samples in any one kriging array get negative weights and outliers are absent, the effect of negative weights is negligible (Sinclair and Blackwell (2002)). Here the number of negative weights was computed and then the Cumulative sum of negative weights (CSNW) was obtained for each ellipsoid dimensions.

### Slope of the Regression

Considering the assumptions that the variogram is valid and the regression is linear, it is possible to compute the main parameters of the regression between estimated and true block grades. Because we don't know individual true block grades the covariance between estimated and true block grades can be computed. Equation 2 gives the slope in terms of this covariance and the variance of the estimated blocks:

$$a = \frac{Cov(z_V^*, z_V)}{Var(z_V^*)} \quad (2)$$

Where the slope of the regression is  $a$ ,  $z_V$  is the true block grade. In a perfect estimate the slope of the regression  $a$  should be very close to one. In these circumstances, the true grade of a set of blocks should be approximately equal to the grade obtained by the kriging estimation.

A rewriting of the expression for the slope in terms of correlation coefficient  $\rho$  is possible (Eq. 3):

$$a = \rho \frac{\sigma_{z_V}}{\sigma_{z_V^*}} \quad (3)$$

Where  $a$  is the slope of the linear regression,  $\rho$  is the linear (Pearson) correlation coefficient,  $\sigma_{z_V}$  is the standard deviation of true block grades and  $\sigma_{z_V^*}$  is the standard deviation of estimated block grades.

From the above expression it can be seen that even for slope equaling one, the correlation may be less than one (because the smoothing effect of kriging necessitates that the variability of estimates is lower than that of true blocks) (Vann et al., 2003). The slope and its interpretation are discussed more completely by Krige (1994; 1996) and Rivoirard (1987).

Another rewriting of the expression for the slope which can be determined for each block estimate as follows in Eq.4 (Sinclair and Blackwell (2002); De-Vitry (2003)):

$$a = \frac{(\sigma_V^2 - \sigma_K^2 + \mu)}{(\sigma_V^2 - \sigma_K^2 + 2\mu)} \quad (4)$$

Where  $\sigma_V^2$  is block variance, i.e. the variance of actual block values,  $\sigma_K^2$  is kriging variance, i.e. the error variance of the block estimate and  $\mu$  is the absolute value of the Lagrange multiplier for each parent cell. Here, SREG was computed for each ellipsoid dimensions.

### ARTIFICIAL NEURAL NETWORKS

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological neural systems. As

in nature, the connections between elements largely determine the network function. A neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Many such input/target pairs are needed to train a network. Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification, classification, and speech, vision, and control systems. Neural networks can also be trained to solve problems that are difficult to approach by conventional computing or human beings (Demuth and Beale (2002)).

### Feed-Forward Neural Networks

Neural networks can be classified into dynamic and static. Static (feed-forward) networks have no feedback elements and contain no delays; the output is calculated directly from the input through feed-forward connections. In dynamic networks, the output depends not only on the current input to the network, but also on previous inputs, outputs, or states of the network. Linear dynamic networks are used as Linear Filters. Dynamic networks can also be divided into two categories: those that have only feedforward connections, and those that have feedback, or recurrent, connections (Demuth and Beale (2002)).

Feed-forward networks have no feedback elements and contain no delays; the output is calculated directly from the input through feed-forward connections. The most common learning algorithm for feed-forward networks is called Back-propagation. Standard back-propagation is a gradient descent algorithm, in which the network weights are moved along the negative of the gradient of the performance function. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function. Feed-forward networks often have one or more hidden layers of non-linear neurons followed by an output layer of linear neurons (Demuth and Beale (2002)).

In the current problem a multi-layer NN is employed to map the input vector of search ellipsoid dimensions onto the output vector of characteristic attributes of kriging, namely the kriging variance, the number of non-estimated blocks and the cumulative sum of negative weights. This is further elaborated in the following sections.

### THE COMPLEX (BOX) OPTIMIZATION METHOD

The Box (Complex) method is an algorithm used to determine a set of decision variables to optimize an objective function developed by Box (1965). A complex is a flexible mathematical figure made up of at least  $n+1$  point where  $n$  is the number of variables. The complex lies in  $n$  dimensional space. Each point consists of coordinates which corresponds to individual variables of the objective function (Box (1965)). The complex moves around the solution space by expanding in contracting in any direction as long as it is feasible.

The generation of the initial complex begins with determining a feasible initial point that satisfies both explicit and implicit constraints. Implicit constraints are those that limit the value of some group of variables (i.e.  $F(x) < 0$ ) and explicit constraints limit the values of an individual variable (i.e.  $0 < X_i < 100$ ). Once this initial feasible point has been determined, a random number generator is used to obtain the remaining points of the initial complex. The random number generator should be set up to return variables within the range of the explicit constraints. It is then necessary to check and see if the point satisfies the implicit constraints.

If an infeasible point is generated the following process will move it back towards feasibility. First, determine the centroid of the feasible points already determined (including the initial point). Move the infeasible point halfway towards this centroid. If the point is still infeasible continue moving it half the remaining distance towards the centroid until it becomes feasible. Continue this process until  $n+1$  feasible point have been generated to form the initial complex.

Expansion and contraction of the complex may now take place. Compute the value of the

objective function at each point in the complex. Determine the point that produces the worst results ( $P_{worst}$ ) (worst is defined as opposite the goal of the objective function). A new point ( $P_{new}$ ) is then determined (Eq. 5) by going a specific distance away from  $P_{worst}$  in the direction of the centroid of the remaining feasible points,  $P_{centroid}$ .

$$P_{new} = (1 + \alpha)P_{centroid} - \alpha P_{worst} \quad (5)$$

The value  $\alpha$  is an expansion coefficient and Box recommended a value of 1.3. Evaluate the objective function at  $P_{new}$  and determine if it is better than  $P_{worst}$ . If  $P_{new}$  is better,  $P_{worst}$  is disregarded and  $P_{new}$  becomes part of the complex. If  $P_{new}$  is worse than  $P_{worst}$ , then a new point  $P_{new2}$  is contracted back towards the centroid at another specified distance based on the contraction coefficient (Eq. 6).

$$P_{new2} = \omega P_{new} + (1 + \omega)P_{centroid} \quad (6)$$

A value of 0.5 is recommended as this contraction coefficient (Tufail and Ormsbee (2007)). This continues until a  $P_{new}$  is obtained that produces a better value of the objective function than  $P_{worst}$ . This process shifts the complex towards better values of the objective function.

Eventually this process of expansion and contraction will shrink the complex near the optimal values of the objective function. It will terminate after consecutive objective functions give the same result, indicating that the complex has converged on the centroid (Ormsbee (1981)). For a more in depth description of the Box-Complex method, please see Box (1965), Tufail and Ormsbee (2007).

Complex is method in which only one objective function needed to be optimized, in this study we wanted to optimize three objective functions simultaneously. Here we used the weighted sum method to optimize a multiobjective function with complex method.

### Treating Multiple Objectives

The most common way to transform a multiobjective problem to a single-objective one is the well-known Weighted Sum method which

uses the following transformation (Eq. 7):

$$U = \sum_{i=1}^k w_i f_i(x) \quad (7)$$

Here,  $w$  is a vector of weights typically set by the decision maker such that  $\sum_{i=1}^k w_i = 1$  and  $w > 0$ . If objectives are not normalized,  $w_i$ 's need not add to 1. As with most methods that involve objective function weights, setting one or more of the weights to zero can result in weakly Pareto optimal points. The relative value of the weights generally reflects the relative importance of the objectives. This is another common characteristic of weighted methods. We wanted all objectives to be treated equally; so all the weights were set to 1.0. However, since the objective values have different units and different orders of magnitude, making comparisons is somewhat difficult. Therefore we normalized the objective functions such that they all have similar orders of magnitude (Arora (2004)). The most reliable approach to do this is to use the Eq. 8:

$$f_i^{norm} = \frac{f_i(x) - f_i^{min}}{f_i^{max} - f_i^{min}} \quad (8)$$

### APPLICATION FOR ANOMALY NO.12A IRON DEPOSIT

The study was performed on anomaly No.12A iron deposit located in Bafgh block in Central Iran. The study area was approximately 500\*200 in plan. In order to estimate data obtained from 19 boreholes with 60 and 100 m spacing were studied. The block size used for the model built for the deposit was 25\*25\*15 m that resulting in a model with 616 blocks.

### Generation of Variograms and Variogram Fitting

Two experimental variograms of percent iron grade in horizontal and vertical directions (variograms are calculated as half the average squared difference between the paired data values) were calculated with a 50 m lag using all the 60\*100 m drillholes data. These experimental variograms

were fitted using the spherical models and it was found that deposit has geometric anisotropy. The optimum sill and range were chosen for variograms by cross validation method. Table1 illustrates the variogram models.

### Data Acquisition

In order to optimize the ellipsoid dimensions for 155 ellipsoids with different dimensions, the value of kriging variance, the slope of the regression, Cumulative sum of kriging negative weights and the number of non-estimated blocks were computed. Domain variation of each criterion and the dimensions of ellipsoids are listed in Table 2. Since the deposit is isotropic in horizontal direction (Table 1) the dimensions of ellipsoid was considered the same in this direction.

### Network Architecture

Data vectors were divided into three sets using random indices, 60% for training, 20% for validation, and 20% for testing. The validation set is used to prevent networks over fit. At the end, the best network architecture was found. The network has three hidden layers with 9, 6, and 4 neurons, as shown in Figure 1. In order to build an ANN one needs to specify the number of processing units, the activation function used by these units and a training algorithm to find the synaptic weights (weights of the links that connect the neurons in various network layers.)

We need a set of input-output data pairs, called training set, and an optimization algorithm to fit the outputs to the given inputs by minimizing the deviation of the approximated outputs from the ideal ones. The deviation is usually represented by the mean square error of the output neurons over the entire training set. Various optimization algorithms have been employed in the training of the network. Our extensive experiments with a number of widely used algorithms revealed that in this case the Conjugate Gradients Method (CGM) has the best convergence rate, especially when it is augmented with scaled restarts.

### OPTIMIZATION RESULTS

The multiobjective function obtained from the neural network using weighted sum method was optimized. 1000 efficient point has been calculated from restarting mathematics program. The results are listed in Table 3.

The results of this method showed that 10 optimum points could be found with minimum kriging variance (KV), the number of non-estimated blocks (NEB) and the cumulative sum of negative weights (CSNW). Then slope of the regression was used as a factor to validate the performance of the system and also select the best dimensions among the others. Applying the slope of the regression (SREG) and according to table 3 in which ellipsoid No.7 has the most number of reiteration (92%), this ellipsoid was selected as the best one with optimum dimensions.

Table 1. Parameters for the spherical variogram models.

	Variogram model	Azimuth	Dip	Range(m)	Sill(% <sup>2</sup> )	Nugget(% <sup>2</sup> )
1	Spherical	-	90	30	171	46
2	Spherical	-	0	250	171	46

Table 2. Domain variation of criteria used for selection the best dimensions of ellipsoid.

criteria	Domain variation
Kriging variance	26.77-51.53
The number of non-estimated blocks	0-582
Cumulative sum of negative weights	0-1.787
Slope of the regression	0.712-0.9
The dimensions of ellipsoid in horizontal direction	3-45 (m)
The dimensions of ellipsoid in vertical direction	20-360 (m)

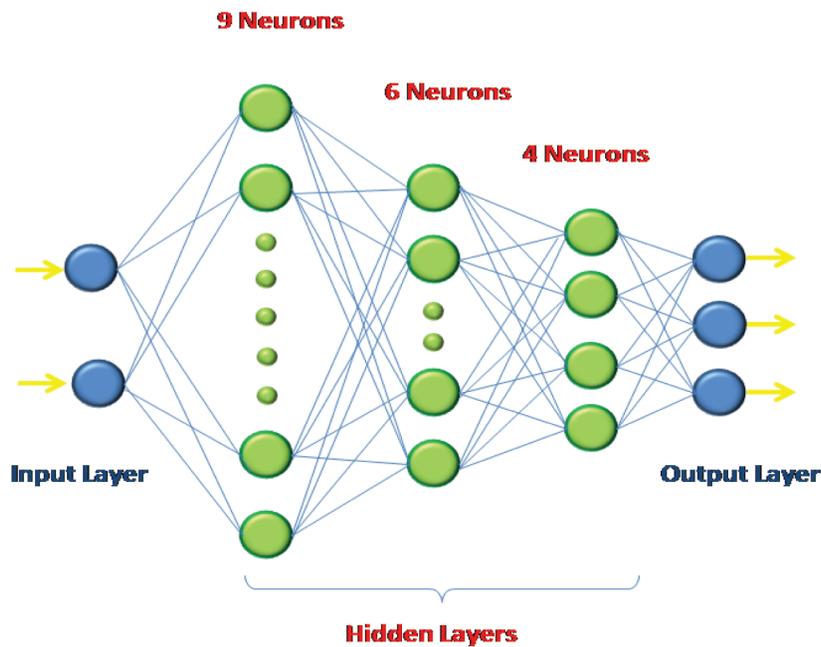


Figure 1. Our proposed network with two hidden layers. The input layer has 3 nodes, the next three hidden layers (intermediate layers) have 9, 6, and 4 nodes respectively, and the output layer have 3 node.

Table 3. The results obtained from restarting mathematics program of optimizing ellipsoid dimensions

No.	Horizontal dimensions	Vertical dimensions	KV	NEB	CSNW	SREG	Percentage of reiteration
1	130	24	35.35	0	0.87	0.871	1.4
2	142	23	34.99	0	1.19	0.875	1.1
3	155	26	34.55	0	1.29	0.882	0.85
4	164	23	34.5	0	1.36	0.883	0.71
5	178	22	34.42	0	1.49	0.884	1.3
6	182	23	34.36	0	1.5	0.885	0.85
7	190	24	34.25	0	1.48	0.887	92
8	200	26	34.16	0	1.53	0.889	0.42
9	220	24	34.10	0	1.59	0.89	0.71
10	250	22	34.14	0	1.66	0.89	0.71

## SUMMARY AND CONCLUSIONS

The definition of the search volume or kriging neighborhood significantly influences the outcome of the kriging estimate. Of the multiple criteria used to define the search volume, perhaps the search radius could be considered the most important one. A new hybrid strategy was proposed to find the optimal value of the search radius.

The proposed strategy uses experimental data to develop a neural simulator that would predict

the values of kriging variance, the number of non-estimated blocks and cumulative sum of kriging negative weights for a given search radius. The neural simulator is then used to predict the value of the objective function in a Box-Complex optimization algorithm which would find the search radius corresponding to the optimum values of the evaluated criteria.

Application of this strategy to a real-world problem, the case of anomaly No.12A iron deposit in Bafgh in central Iran, resulted in 10

optimum points with minimum kriging variance, number of non-estimated blocks and the cumulative sum of negative weights. The slope of the regression was then used as a measure to validate the performance of the system and to choose the best dimensions from multiple choices. An ellipsoid of dimensions 190×190×24m was identified as the best solution for the case study through numerous runs of the computer code that was generated to implement the proposed strategy.

Compared to kriging estimation techniques, the hybrid strategy presented here could produce similar, if not more accurate, results much faster and at a considerably lower computational cost.

**ACKNOWLEDGMENTS** Authors would like to express their appreciation and gratitude to Dr. Isobel Clark, for her help in reviewing the paper.

## REFERENCE

- Arora, J., 2004. Introduction to Optimum Design, 2<sup>nd</sup> ed., London: Elsevier Academic Press.
- Box, M.J., 1965. A new method of constrained optimization and a comparison with other methods. *The Computer Journal* 8(1): 42-52.
- Demuth, H., Beale, M., 2002. Neural Network Toolbox for use with MATLAB, Version 4.0, User's Guide, MathWorks.
- Deutsch, C., 1995. Correcting for Negative Weights in Ordinary Kriging. *J of Computers & Geosciences* 22(7): 765-773.
- De-Vitry, C., 2003. Resource Classification – a Case Study From the Joffre-Hosted Iron Ore of BHP Billiton's Mount Whaleback Operations. *J of Mining Technology (Trans. Inst. Min. Metall. A)* 112:185-196.
- Froidevaux, R., 1993. Constrained kriging as an estimator of local distribution functions, in Capasso, V., Girone, G., and Posa, D., eds., Proceedings of the International Workshop on Statistics of Spatial Processes: Theory and Applications. Bari, Italy, p. 106–118.
- Hagan, M.T., Demuth, H.B., Beale, M., 1996. *Neural Network Design*, Boston: PWS Publishing Company.
- Journel, A.G., Rao, S.E., 1996. Deriving conditional distributions from ordinary kriging: Stanford Center for Reservoir Forecasting (Report No. 9), Stanford, 25 p.
- Krige, D.G., 1994. An analysis of some essential basic tenets of geostatistics not always practised in ore valuations, in Proceedings Regional APCOM: Computer Applications and Operations Research in the Minerals Industries, Slovenia, pp 15-18.
- Krige, D.G., 1996a. A basic perspective on the roles of classical statistics, data search routines, conditional biases and information and smoothing effects in ore block valuations, in Proceedings Conference on Mining Geostatistics, Kruger National Park, South Africa, pp 1-10 (Geostatistical Association of South Africa).
- Ormsbee, L., 1981. Optimization of Hydraulic Networks. In: Proceedings International Symposium on Urban Hydrology, Hydraulics, and Sediment Control, Lexington, KY.: July 27-30.
- Rivoirard, J., 1987. Two key parameters when choosing the kriging neighbourhood. *J Math Geol*, 19:851-856
- Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation. pp. 224-226.
- Szidarovszky, F., Baafi, E.Y., Kirn, Y.C., 1987. Kriging Without Negative Weights. *J of Mathematical Geology*. 19(6): 549-559.
- Tufail, M., Ormsbee, L.E., 2007. A Shuffled Box Complex-Based Optimization Model for Watershed Management, In: Proceedings of the World Environmental and Water Resources Congress, Tampa, Florida, USA: May 15-19.
- Vann, J., Jakson, S., Bertoli, O., 2003. Quantitative kriging neighbourhood analysis for the mining geologist – A description of the method with worked case examples, 5th International Mining Geology Conference. Pp:1-10.