Bollettino di Geofisica Teorica ed Applicata Vol. 61, n. 2, pp. 177-198; June 2020
DOI 10.4430/bgta0306

Determining the optimum search range for 2D and 3D mapping based on kriging through quantitative analysis

O. Asghari, M. Safikhani and S. Talesh Hosseini

[1] *Simulation and Data Processing Lab, University of Tehran, Tehran, Iran*

(Received: 6 August 2019; accepted: 4 November 2019)

ABSTRACT As the best linear estimator, kriging is now a well-established method in all types of 3D geomodelling, including geochemical mapping, rock type modelling, 2D geophysical mapping, and resource estimation. In this context, investigating kriging performance has always been of interest to numerous researchers. Evaluating kriging implementation for different applications has been a growing field of study in the last few decades. Although many authors have discussed various kriging parameters, it seems necessary to conduct more detailed studies on range searching, high and low nugget effect, as well as 2D and 3D estimations. In this paper, an optimal search range was determined using Quantitative Kriging Neighbourhood Analysis (QKNA), and the utility of this search range was explored by assessing kriging efficiency. Because of the existence of different numerical measures of search ranges in each criterion, it is difficult to define the optimal search range of an estimation process. In this research, different Multiple Criteria Decision Making (MCDM) methods were employed to determine the optimal search range via QKNA and by considering criteria which were applied to different cases. Given the unique capacity of this method in meeting this challenge, the Fuzzy-TOPSIS method, a variant of MCDM, was used in this study.

Key words: kriging neighbourhood, quantitative analysis, kriging efficiency, optimal search range, multiple criteria decision making, Fuzzy-TOPSIS method.

1. Introduction

Geostatistical methods are now widespread in Earth science applications such as geochemical mapping, rock type modelling, and resource estimation. Among those methods, kriging, as a linear estimator, is popular in resource estimation (David, 1977), geochemical mapping (Changjiang *et al.*, 2009; Talesh Hosseini *et al.*, 2018), estimation of trace elements contamination (Tavares *et al.*, 2008), identification of geochemical anomalies (Jimenez-Espinosa *et al.*, 1993), prediction of flow-duration curves (Castellarin, 2014; Varouchakis *et al.*, 2016), spatial variability of aquifer level (Varouchakis *et al.*, 2016), and analysis of the relation between signal spectral range and noise (Jarmokowski, 2019). Kriging is also the base of most geostatistical simulation methods like sequential Gaussian simulation which is widespread in Earth science applications [forecasting the grade-tonnage curves and their uncertainty (Hosseini *et al.*, 2017), multivariate simulation of multi-element deposit (Mahlooji *et al.*, 2019)].

Kriging is an estimator that yields the Best Linear Unbiased Estimate (BLUE) of point values or of block averages (Matheron, 1973; Armstrong, 1998); but this is only true provided that its search neighbourhood is properly defined. The most important characteristic of kriging is its unbiasedness: blocks estimated to have a value of Z_v^* will on average have that value (David, 1977). This characteristic is called "conditional unbiasedness". Kriging neighbourhood parameters are very important in the process of estimation. These parameters including min/max number of used data, use of octant or quadrant search, and maximum search ranges in two or three dimensions are consequential inasmuch as they directly affect kriging weights (Vann *et al.*, 2003).

Just like methods developed for assessing the accuracy of simulation algorithms (Mantoglou and Wilson, 1982; Ripley, 1987; Tompson *et al.*, 1989; Omre *et al.*, 1993; Gotway and Rutherford, 1994; Tran, 1994; Deutsch, 1997; Emery, 2004, 2008; Emery and Peláez, 2011; Safikhani *et al.*, 2017), a common method of validating kriging results and determining optimal search ranges is the one developed by Rivoirard (1987). This method was, then, taken up by Armstrong (1998), Vann *et al.* (2003), Boyle (2010), and Shademan *et al.* (2013). It operates similar to the method proposed by Krige (1996a). Rivoirard (1987) deployed two parameters to determine appropriate search ranges: Weight of the Mean (WM) for Simple Kriging (SK) and Slope Of the linear Regression (SOR) of true values versus estimated ones. Armstrong (1998) modified Rivoirard's (1987) method and suggested the same parameter along with the proportion of nugget effect for optimising the search range. Vann *et al.* (2003) introduced a similar approach to determine the optimal estimation neighbourhood using a combination of WM of SK, SOR, distribution of kriging weight [including the proportion of Negative Weight (NW)], and Kriging Variance (KV). This method is also known as Quantitative Kriging Neighbourhood Analysis (QKNA). Emery (2009) refined kriging equations to enhance the kriging neighbourhood definition. The proposed equations were reported to be very useful in updating kriging weights and variances in a short time. Madani and Emery (2019) employed different strategies to determine the optimal cokriging neighbourhood by investigating five alternatives: single search, multiple search, strictly collocated search, multi-collocated search, and isotopic search. Boyle (2010) obtained the best search ranges using WM and SOR in a case study on the Jura data set. Khakestar *et al.* (2013) used three parameters of SOR, WM, and KV to establish the best search ranges in an estimation program and compared the results with those reported by different researchers. Coombes (2008) aimed at identifying a set of estimation parameters that could afford the highest Kriging Efficiency (KE) and SOR statistics to ensure the most reliable estimate of block grades. Coombes and Boamah (2015) suggested a Localised Kriging Neighbourhood Analysis (LKNA) approach which could be used to optimise different parameters of each block within each domain. Hundelshaussen *et al.* (2018) advanced a new approach known as Localised Kriging Parameter Optimisation (LKPO) which overcame the disadvantages of LKNA approach.

All of these works focus on some special parameter and try to show the importance of search range in estimation results. As a research gap, it can be argued that none of these studies addresses the type of estimation. According to Deutsch *et al.* (2014), there are three kinds of estimates, each requiring a different strategy and criteria for assessing results: 1) estimates for visualization and geological understanding, 2) estimates for interim planning, and 3) ultimate estimates for reserve classification. Moreover, the authors put forth a new definition for KE and introduced another kriging quality criterion which could be used to characterise estimation results.

In this study, different search ranges for 2D (geochemical anomaly detection) and 3D (resource estimation) cases are evaluated using QKNA (Vann *et al.*, 2003). Range searching is just one of the parameters discussed in this work. Min/max of data, using octane search, and max data from a single borehole are other parameters that are optimised here. The search range with the best QKNA criteria will be introduced as the optimal search range. Subsequently, the advantage of QKNA will be investigated by calculating KE. Cross validation analysis and swath plots are presented to show the effect of QKNA results on the estimation quality. Different synthetic and real cases with various statistical and geostatistical properties are presented to take account of all possible conditions. It should be clarified that the authors of the present research have focused on Ordinary Kriging (OK), and SK is used for some specific purpose [e.g. defining variance of global SK (σ^2_{GSK}) and calculating KE and WM for any search range]. Considering that different search ranges offer different values of QKNA criteria by means of the kriging method, the multiple criteria decision making (MCDM) approach is a highly useful choice for finding the best optimal search range. Among different methods of MCDM, TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution), thanks to its selective nature, is used for determining the optimal search range (Opricovic and Tzeng, 2004). First developed by Hwang and Yoon (1981), TOPSIS helps solve best choice problems. In order to establish the weights of QKNA criteria in this study, the relationship between the criteria can be defined using the linguistic variables, which includes triangular fuzzy numbers (Wang and Elhag, 2006; Chen and Tsao, 2008). Therefore, the triangular fuzzy number is combined with TOPSIS to select the best alternative (optimal range) among others.

By deploying QKNA for estimating different variables such as geochemical assays, resource grade, geophysical parameters (e.g. induced polarisation and resistivity), and geomechanical parameters, users could be sure that the estimation results are the best. In addition, QKNA could be an essential step in setting up any kriging estimate, including that used for conditioning a simulation (Vann *et al.*, 2003). In addition to finding the optimal search ellipsoid, this method provides better results in the abundance of nugget effect and sparse data. This could be seen in some case studies on geochemical anomaly detection, primary resource estimation, etc. To show the efficiency of the method, QKNA was here applied to different case studies through a combination of synthetic and real data sets. This paper initially explores the efficiency of different parameters under consideration in order to select the optimal estimation neighbourhood; next, a number of quantitative tools are employed to compare the results of each kriging neighbourhood.

2. How QKNA can help?

As discussed, kriging is the best linear estimator that provides unbiased results with minimum estimation variance. However, kriged values are highly dependent on the kriging neighbourhood. Neighbourhood parameters are often defined by users. For example, search ranges are commonly defined based on variogram ranges. Defining precise search ranges by means of variogram range could be very risky insofar as variogram ranges are not always the best search ranges. The slope of the variogram model at short lags and relative nugget effect (behaviour at origin) exerts a greater impact on search range selection than it does at larger increments (Vann *et al.*, 2003). Too restricted search ranges give rise to conditional bias (Krige, 1994, 1996a, 1996b), whereas wide ranges may lead to highly smoothed results and, consequently, loss of local accuracy.

In this regard, QKNA adjusts the neighbourhood to obtain the best regression statistics in order to reduce conditional bias. Meanwhile, smoothing could happen in the process of mitigating conditional bias and it is not possible to have a conditionally unbiased estimation with no smoothing effect. In fact, as Journel and Huijbregts (1978) articulated, the necessity for smoothing is a consequence of the 'information effect', meaning that non-exhaustive information may cause smoothing in the estimation result.

When using QKNA, it is desirable to determine an optimal combination of search neighbourhood which entails the least conditional unbiasedness (Vann *et al.*, 2003). The priority of criteria used to assess kriging performance is as follows (Vann *et al.*, 2003):

- the SOR of the true value on the estimated value;
- the SK WM;
- the distribution of kriging weight (including the proportion of NWs);
- KV.

3. How to calculate required parameters?

All of the parameters introduced above should be calculated in an estimation process. Some of them, like KV and SOR, are easy to be obtained, but others, such as NW and WM, need some post-processing.

3.1. Slope Of linear Regression (SOR)

It has been established that OK is an estimator which can minimise conditional bias (Matheron, 1963; David, 1977; Rivoirard, 1987; Ravenscroft and Armstrong, 1990; Krige, 1994, 1996a, 1996b). A perfect estimator Z_v^* (estimated values) equals Z_v (true values) but this does not happen in reality and some fluctuations will always occur. Even so, if the regression of Z_v^* on Z_v remains linear with a slope of 1.0, the result is conditionally unbiased and the obtained estimation could be acceptable. Armstrong (1998) defines the slope (P) of linear regression thus:

$$P = \frac{Cov(Z_v, Z_v^*)}{Var(Z_v^*)} \tag{1}$$

or equally:

$$P = \rho \frac{\sigma_{Z_v}}{\sigma_{Z_v^*}} \tag{2}$$

where σ_{Z_v} is the standard deviation of true values, $\sigma_{Z_v^*}$ is the standard deviation of estimated values, ρ is the linear correlation coefficient.

Moreover, for each position estimation, P can be calculated as follows (Sinclair and Blackwell, 2002; De-Vitry, 2003):

$$P = \frac{(\sigma_v^2 - \sigma_k^2 + \mu)}{(\sigma_v^2 - \sigma_k^2 + 2\mu)} \tag{3}$$

where σ_v^2 and σ_k^2 are respectively the variance of actual block values and kriging variance, and μ

is the absolute value of the Lagrange multiplier for each estimation position. For more details on SOR calculation, the reader is referred to the studies by Rivoirard (1987), Krige (1994, 1996a), Armstrong (1998), Sinclair and Blackwell (2002), and De-Vitry (2003).

If $P < 1$, then one could infer that the estimation result is under-smoothed. This means that the true grade of areas predicted to have high values is indeed lower than expected; on the other hand, areas predicted to have low values are likely to show higher grades (Rivoirard, 1987; Pan, 1998). In such circumstances, one can argue that the neighbourhood is too restricted (Boyle, 2010). This problem is illustrated in Fig. 1, in which a departure could be seen between the scatter plot of true values on the estimated ones and $X = Y$, which represents the conditionally unbiased estimation.

If $P > 1$, then the estimation result is over-smoothed. In other words, the variance of the estimated vector is less than the actual variance of the sample vector. Smoothing is somehow unavoidable in making an estimation. As Krige (1996a) observed, one cannot both avoid smoothing and achieve conditional bias; nevertheless, it is still possible to modify smoothing via post-processing the result (Rossi and Parker, 1994; Krige *et al.*, 2005).

Fig. 1 shows different SOR situations. It can be seen how a conditionally unbiased estimation can cause extra error in classifying the deposit. In the biased result, areas I and III are big, which means that a large part of the deposit predicted to be ore is not really so; conversely, a part of the deposit predicted to be waste is ore actually. Such predictions should not be used in the final estimation.

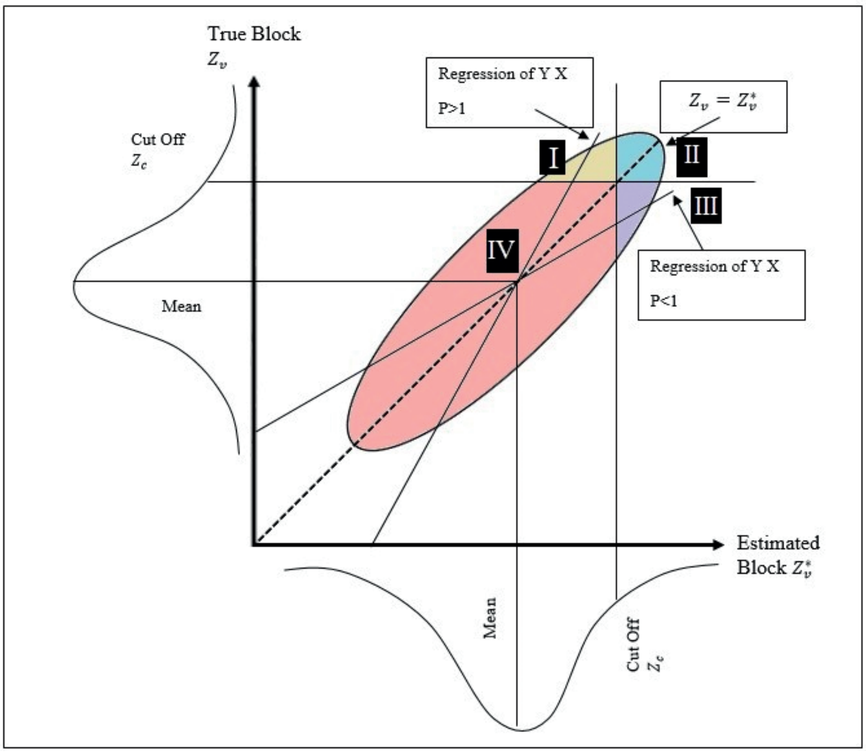

Fig. 1 - Different situations of P that could take place in an estimation process. Quadrants II and IV represent the correct classification of waste and ore, and quadrants I and III are their corresponding incorrect classifications (modified from Journel and Huijbregts, 1978).

3.2. Weight of the Mean (WM) in SK

When using SK, the sum of weights is not constrained to add to one. Hence, the remaining weight is allocated to the mean, which is assumed to be known. In SK, WM indicates the degree to which kriging depends on local samples, and it is an inversely proportional index of the so-called 'screen effect'. Rivoirard (1987) used WM to determine the minimal neighbourhood (the neighbour with maximum 20% weight allocated to the mean is introduced as the minimal search range). Though easy to be measured, WM is not a parameter in many geostatistical packages (Armstrong, 1998; Chilès and Delfiner, 2012). The related formula is as follows:

$$Z_v^{*sk} - m = \sum \lambda_i^{*sk} (Z(x_i) - m) \tag{4}$$

$$\lambda_m = 1 - \sum \lambda_i^{sk} \tag{5}$$

in which $\sum \lambda_i^{sk}$ is the sum of SK weights, and λ_m is WM. Using QKNA, one can obtain both the best SOR and the minimum weight allocated to the mean (Vann *et al.*, 2003).

3.3. Negative Weight (NW)

If the search is too restrictive, the meaningful positive weight will be assigned to the additional samples when the search range is expanded. In fact, marginal samples should get very small or even NWs, and modifying weights (Deutsch, 1996) or setting them to zero will not be helpful because it will bring about conditional bias (Vann *et al.*, 2003). Actually, if an NW represents less than 5% of the total weight, no problem occurs. Reporting weights in a kriging process is not part of conventional software applications. In the present study, the kriging functions in the 'mGstat' toolbox in MATLAB, 'krig.m' have been modified to maintain kriging weights.

3.4. Kriging Variance (KV)

Kriging is an estimator with minimum variance. Specifically, global SK is known to have minimal estimation variance. KV provides a measure of the error associated with the kriging estimator; consequently, it can be a criterion to evaluate the quality of data, density, and geometry. The less the KV is, the better the estimation quality will be. However, this criterion is not as important as the slope of the regression, but it can still be used as a yardstick for kriging evaluation. This property is highly dependent on the data involved in estimation and is usually reported from the kriging process by the software. Even so, its calculation is not very difficult, and one might do it using the following expression:

$$var\left(Z_v^*\right) = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j C(x_i, x_j) \tag{6}$$

in which $C(x_i, x_j)$ is the value of covariance function between sample locations x_i, x_j, and λ_i and λ_j represents the weight assigned to these samples.

3.5. Kriging Efficiency (KE)

The term KE was first used by Yfantis *et al.* (1987) for evaluating square, triangular, and hexagonal sampling grids. However, it was Krige (1996a) who introduced it as a measure of the efficiency of block estimates. In fact, it is another definition of KV which is normalised by the true block variance. KE can be used as a measure of estimation quality. Coombes (2008) tried to identify a set of estimation parameters that result in the highest KE and SOR statistics.

Let σ^2 and σ_k^2 be the sample and kriging variance; then (Krige, 1996a):

$$KE_{DK}(\%) = \frac{\sigma^2 - \sigma_k^2}{\sigma^2} = \frac{\bar{C}(V, V) - \sigma_k^2}{\bar{C}(V, V)} \tag{7}$$

where KE_{DK} is KE and $\bar{C}(V, V)$ is the average covariance of points within the blocks.

In theory, for a perfect estimation algorithm it is expected to be 1 (100%), which necessitates KV to be zero. This situation can only happen for SK to have no estimation variance. In real cases, the value of KE is less than 1. Besides, in some special cases where KV is greater than the true block variance, KE is negative (Deutsch *et al.*, 2014).

Deutsch *et al.* (2014) restricted the application of KE to linear estimation and proposed another expression. Considering that global SK is associated with the linear minimum estimation variance, then:

$$KE(u) = \frac{\sigma_{GSK}^2(u)}{\sigma_k^2(u)} \tag{8}$$

in which $KE(u)$ is SK and $\sigma_{GSK}^2(u)$ is the global SK variance.

4. Comments

Before going through samples, several comments need to be made:

1) QKNA is a methodology for evaluating parameters of a model and a good way to determine the optimal search range;
2) through QKNA, it is possible to evaluate estimation quality and assess the accuracy of kriging;
3) using QKNA for establishing search ranges in two or three dimensions can assure us about the accuracy of the selected model;
4) when using QKNA, it should be considered that the most important parameters are SOR and WM. Selecting the optimal search range must be conducted with an eye primarily to these two parameters and then the other two parameters of NW and KV;
5) in this paper, the authors have not rigorously followed Rivoirard (1987) and Vann *et al.* (2003) in their approach to finding the optimal search range. Meanwhile, as Rivoirard (1987) recommended, SK was performed to calculate WM and determine the minimal search range, and the resulting WM was obtained accordingly. The minimal search range is that with at most 20% WM proportion.

5. Fuzzy-TOPSIS method

Being based on the concept of similarity to the ideal solution, TOPSIS is a compromise technique for finding the best alternative (optimal range) among all candidates (Wang and Elhag, 2006; Boran *et al.*, 2009). In fact, compromise techniques function based on a reference point (ideal solution). In the TOPSIS method, the distance of alternatives from both the positive and negative ideal points must be considered in choosing the best alternative (Chen and Tsao, 2008). In fact, the best alternative (optimal range) is obtained by calculating the minimum distance from the positive ideal solution and the maximum distance from the negative ideal solution. In order to consider the linguistic variables in the decision-making process, the fuzzy set theory is used in the TOPSIS method. The main purpose of Fuzzy-TOPSIS method is to rank different alternatives based on a fuzzy environment. Among different types of fuzzy sets, the triangular fuzzy number was applied in this research before applying the TOPSIS method. This type of fuzzy number can be defined by a (l, m, u) number (Eq. 9 and Fig. 2) (Shemshadi *et al.*, 2011; Abedi *et al.*, 2013):

$$\mu_A(x) = \begin{cases} (x-l)/(m-l), & l \leq x \leq m \\ (u-x)/(u-m), & m \leq x \leq u \\ 0 & otherwise \end{cases} \quad (9)$$

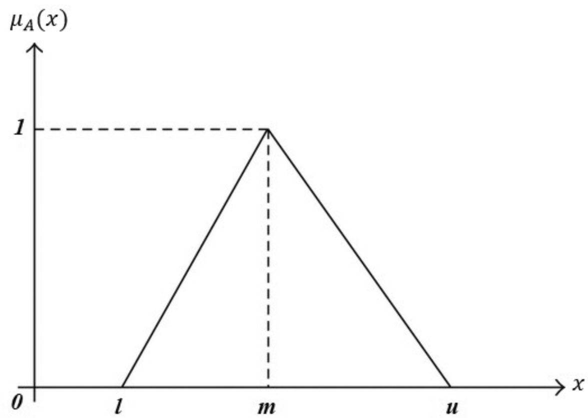

Fig. 2 - Triangular fuzzy number.

Let us assume that A is a decision matrix and x_{ij} represents the score of alternative A_i (a fuzzy number) in the criterion C_j. Also, l and m represent the lower and upper value of the support of the fuzzy number A and m denotes the maximum grade of $\mu_A(x)$. The Fuzzy-TOPSIS algorithm has the following steps (Wang and Elhag, 2006; Chen and Tsao, 2008; Boran *et al.*, 2009; Junior *et al.*, 2014):

Step 1. Calculate the normalised decision matrix using Euclidean normalisation method:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}}; \forall i, j \quad (10)$$

where r_{ij} gives the normalised values of the membership matrix A, and m represents the number of the alternatives (search range).

Step 2. Compute the weight values of QKNA criteria from the fuzzy set by observing the following stages:

- define the importance of QKNA criteria using the linguistic variables presented in Fig. 3;

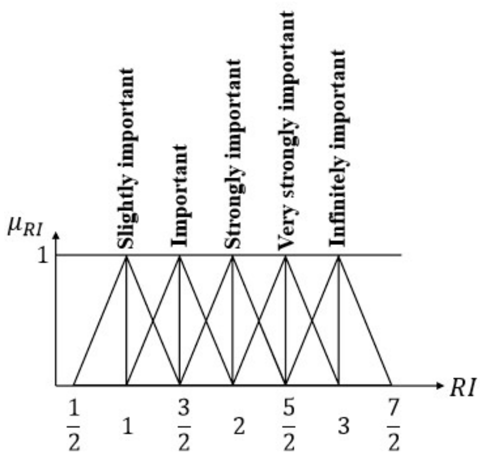

Fig. 3 - Linguistic scales for determining the importance of QKNA criteria.

- determine the pairwise comparison matrix from decision-maker's comments such that:

$$\widetilde{A} = (\widetilde{a}_{ij})_{n\times n} = \begin{bmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \cdots & (l_{1n}, m_{1n}, u_{1n}) \\ (l_{21}, m_{21}, u_{21}) & (1,1,1) & & (l_{2n}, m_{2n}, u_{2n}) \\ \vdots & & \ddots & \vdots \\ (l_{n1}, m_{n1}, u_{n1}) & (l_{n2}, m_{n2}, u_{n2}) & \cdots & (1,1,1) \end{bmatrix}$$

where $\tilde{a}_{ij} = (l_{ij}, m_{ij}, u_{ij}) = \tilde{a}_{ij}^{-1} = (\frac{1}{u_{ji}}, \frac{1}{m_{ji}}, \frac{1}{l_{ji}})$, for $i, j, = 1,..., n$ and $i \neq j$. Also, $\tilde{a}_{ij}$ represents the l, m and u values of alternative A_i in the criterion C_j;

- calculate the normalised values of the pairwise comparison matrix such that:

$$\widetilde{S}_i = \sum_{j=1}^{n} \widetilde{a}_{ij} \otimes \left[\sum_{k=1}^{n} \sum_{j=1}^{n} \widetilde{a}_{kj} \right]^{-1} \tag{11}$$

where the numerator is a n×3 matrix with entries given by $\begin{cases} \sum_{i=1}^{n} l_i & \text{first row} \\ \sum_{i=1}^{n} m_i & \text{second row for } n \text{ criteria} \\ \sum_{i=1}^{n} u_i & \text{third row.} \end{cases}$

Also, the denominator is obtained by summing the columns of the previous matrix;

- compute the greatest degree of S_i, compared to that of all other fuzzy numbers, based on the following equation:

$$V\left(\tilde{S}_i \geq \tilde{S}_j\right) = \sup\left[min\left(\tilde{S}_j(x), \tilde{S}_i(y)\right)\right],\ y \geq x\,. \tag{12}$$

In other words:

$$V\left(\tilde{S}_i \geq \tilde{S}_j\right) = \begin{cases} 1 & m_i \geq m_j \\ \dfrac{u_i - l_j}{(u_i - m_i) + (m_j - l_j)} & l_j \leq u_i \quad i, j = 1, \ldots, n;\, j \neq i \\ 0 & otherwise \end{cases} \tag{13}$$

where $\tilde{S} = (l_i, m_i, u_i)$ and $\tilde{S}_j = (l_j, m_j, u_j)$.

Step 3. Calculate the weighted matrix. Assume that $W = (w_1, w_2, ..., w_n)$ is the weight of criteria. The weighted matrix is calculated by the following equation:

$$V = \left[\nu_{ij}\right]_{m\times n};\, \nu_{ij} = w_j\,.\,r_{ij};\ \ \forall i, j\,. \tag{14}$$

Step 4. Determine the positive ideal solution (A^+) and negative ideal solution (A^-) by the following equation:

$$\begin{aligned} A^+ &= (\nu_1^+, \nu_2^+, \ldots, \nu_n^+) \\ A^- &= (\nu_1^-, \nu_2^-, \ldots, \nu_n^-) \end{aligned} \tag{15}$$

where

$$\begin{aligned} \nu_j^+ &= \left(Max\,\nu_{ij}, j \in J_1\ or\ Min\,\nu_{ij}, j \in J_2\right);\ \ \forall j = 1,2, \ldots, n, \\ \nu_j^- &= \left(Min\,\nu_{ij}, j \in J_1\ or\ Max\,\nu_{ij}, j \in J_2\right);\ \ \forall j = 1,2, \ldots, n, \end{aligned} \tag{16}$$

in which J_1 represents the set of positive criteria and J_2 denotes the set of negative criteria. J_1 and J_2 correspond to the benefit and cost criteria, respectively.

Step 5. Compute the distance between the alternative and positive ideal solution and the distance between the alternative and negative ideal solution for all alternatives such that:

$$\begin{aligned} d_i^+ &= \sqrt{\sum_{j=1}^{n} (d_{ij}^+)^2} \\ d_i^- &= \sqrt{\sum_{j=1}^{n} (d_{ij}^-)^2} \end{aligned} \tag{17}$$

where

$$d_{ij}^{+} = \nu_j^{+} - \nu_{ij}; \quad \forall i = 1,\ldots,m,\ j = 1,2,\ldots,n,$$
$$d_{ij}^{-} = \nu_j^{-} - \nu_{ij}; \quad \forall i = 1,\ldots,m,\ j = 1,2,\ldots,n, \tag{18}$$

in which *i* and *j* are the number of alternatives and criteria, respectively.

Step 6. Calculate the relative closeness index of the alternatives by the following equation:

$$rc_i = \frac{d_i^{-}}{d_i^{+} + d_i^{-}}; \ \forall i = 1, 2,\ldots, m \,. \tag{19}$$

The relative closeness index ranges between 0 and 1. The bigger the relative closeness index is the better the alternative will be.

6. Applications

6.1. Calculation of criteria weights

In order to calculate the weight of each criterion, the importance degree of the seven criteria was defined by means of the linguistic variables, which were chosen based on the comments of the geostatistics specialist (Table 1). Once the relative importance matrix was determined by using the importance degrees (Table 1), the same stages described earlier in the section on Fuzzy-TOPSIS method (Fig. 3) were followed in order to achieve the fuzzy numbers. Finally, the weight of each criterion was obtained by means of the defined stages (Table 2).

Table 1 - The importance degree of the seven criteria.

	SOR	Real Corr.	leave-1-out cross validation	WM (SK)	NW	KV	KE
SOR	Eq.	Sl.	Sl.	St.	St.	VS.	VS.
Real Corr.		Eq.	Sl.	Sl.	Sl.	St.	St.
leave-1-out cross validation			Eq.	Sl.	Sl.	St.	St.
WM (SK)				Eq.	St.	Sl.	Sl.
NW					Eq.	Im.	Im.
KV						Eq.	Im.
KE							Eq.
Equally important (Eq.); Slightly important (Sl.); Important (Im.); Strongly important (St.); Very Strongly important (VS.); Infinitely important (In.).							

Table 2 - The weight of each criterion calculated based on the triangular fuzzy number.

criterion	SOR	Real Corr.	leave-1-out cross validation	WM (SK)	NW	KV	KE
Weight	0.224	0.170	0.175	0.146	0.134	0.088	0.063

6.2. Synthetic samples (2D Kriging)

In this section, the application of QKNA to a grid with low and high nugget effect is discussed. Synthetic data were generated by the Lower-Upper (LU) simulation program (Alabert, 1987) on a regular grid of 160×160×1 sample in x, y, and z dimensions [with 5 m spacing, which makes an area of 800×800 m^2 (Fig. 4)] and also using the following anisotropy model:

$\gamma = 0.1{*}Nug(0) + 0.9{*}Sph(50, 30)$, Az = 0 Dip = 0
$\gamma = 0.4{*}Nug(0) + 0.6{*}Sph(50, 35)$, Az = 0 Dip = 0

Then, a data set of 970 samples was chosen randomly as hard data (Fig. 4). This data set was later used to calculate the anisotropy model of the grid:

$\gamma = 0.1{*}Nug(0) + 0.9{*}Sph(55, 47, 0)$, Az = 0 Dip = 0
$\gamma = 0.4{*}Nug(0) + 0.6{*}Sph(58, 36, 0)$, Az = 0 Dip = 0

where γ represents the variogram value of the studied variable, *Nug* denotes the nugget effect. The *Sph* shows that the fitted model is spherical.

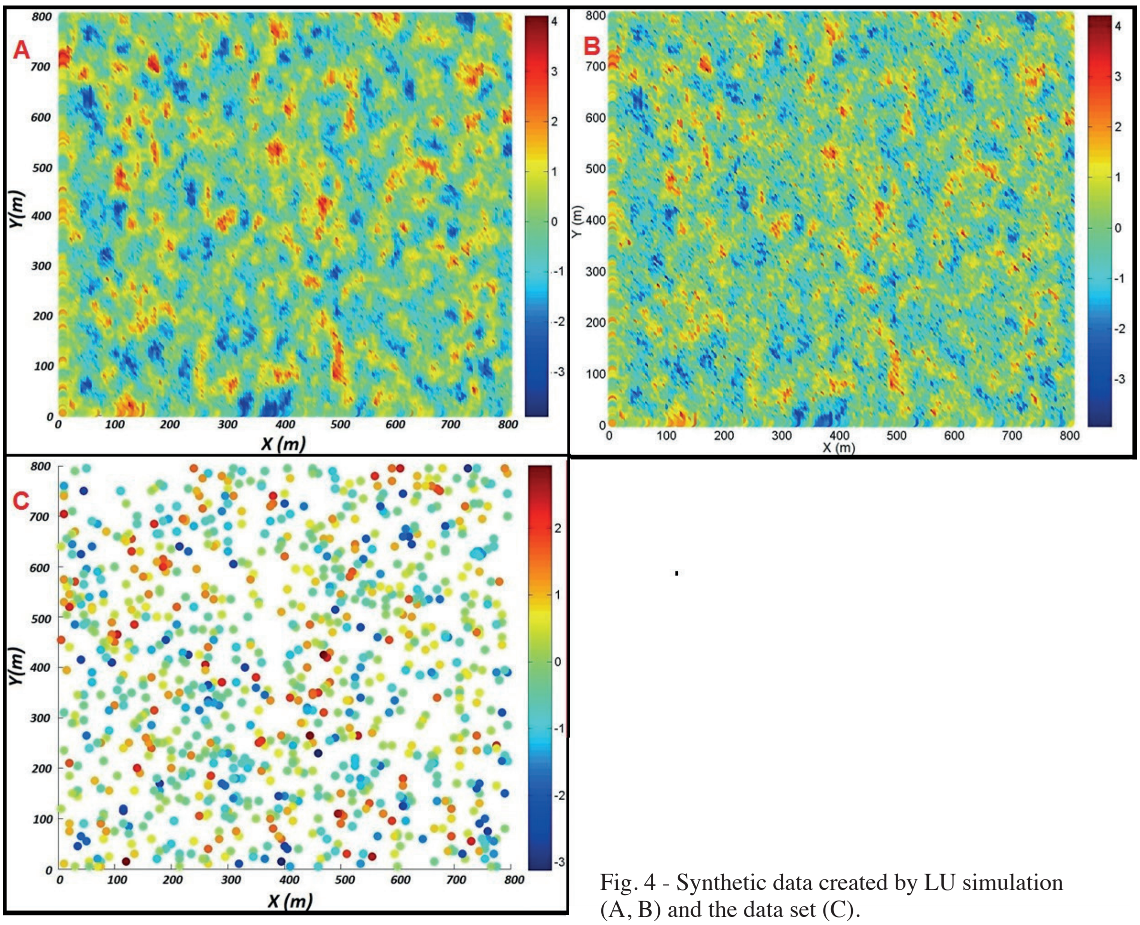

Fig. 4 - Synthetic data created by LU simulation (A, B) and the data set (C).

To analyse the effect of the search range on QKNA criteria, we calculated these criteria and applied them to 15 different search ranges (Table 3). In these two cases, the focus was on the search range and other parameters like min/max number of data or octant search were not examined. Cross validation and estimation were performed via both ordinary and SK for all search ranges. To perform the estimations and calculate the required parameters, a Matlab toolbox, namely mGstat, was used. Some parts of this code were changed in order to make it work properly.

Table 3 shows the result of QKNA when the search range expands from the 1st ellipsoid SOR and reaches its proper values in the 7th ellipsoid. As given in the table, other criteria like KV, NW, WM, KE, and leave-one-out correlation are also acceptable. Hence, it can be said that it is the best search range for this case. It should be considered that in reality it is not possible to calculate the real correlation which is shown in this table. Thus, choosing the best search range through these criteria is not practical. As explained before, the best search range should be determined via other criteria.

In this table, the real correlation is calculated according to the estimation result and actual ones (obtained through LU simulation) and the leave-one-out correlation is gauged based on this algorithm. There exists an optimal range for some other parameters such as WM. This search range is expected to be at most 20% proportional to WM in SK. NWs are assigned, because of the screen effect, to marginal data which have no significant correlation with the estimation location but are involved in the estimation process. NWs are not generally problematic if they represent a small proportion (up to 5%) of the total weight; even so, it is better to choose NW proportions less than 5% for special cases like high-grade zones of gold and copper deposit.

Table 3 shows the decision making matrix of QKNA parameter for all search ranges. According to the second step of the algorithm proposed for selecting the optimal search range, the normalised decision matrix was obtained by means of normalised Euclidean distance. Finally, the relative closeness of each search range was determined via the normalised decision matrix and weight of each criterion. Thus, the Fuzzy-TOPSIS algorithm was used for discovering the optimal search range. Table 4 presents the result of Fuzzy-TOPSIS method for the synthetic samples.

Table 4 and Fig. 5 show that the optimal search range introduced by QKNA is a search range between ellipsoid 6 and 8 in major axes. To find the optimal search range, then, the algorithm reruns between these two ranges and the best one is selected (see Table 5).

Boyle (2010) reported that neighbourhood analysis adds value in the presence of high nugget effect. For this case, the range between ellipsoids 8 and 9 could be chosen as the optimal search range (neighbourhood with 81 m length in the major axis). It can be seen in Fig. 5 and Table 5 that the best estimation quality occurs in this neighbourhood.

Regarding KV, it is obvious that it should not be very high compared to the total variance of the data set. Besides, in the stability of other parameters (including variance of the estimated vector), the less the KV is, the better the estimation will be. High KV is due to the number of data in the search ellipsoid. Table 3 indicates that in ellipsoids 7 and 8, fewer than 9 data exist, which explains the reason for the high KV. KE depends on nugget effect and KV, with the latter ignoring local grade variability. In a kriging program, besides other parameters, KE should be in the appropriate range. In Table 5, to evaluate KE, the linear minimum variance estimation must be calculated. To this end, a SK estimation was performed for each estimation using all available data, and the minimum variance was found to be about 0.20 and 0.45 for low and high nugget effects. This value was, then, used to calculate KE as $\sigma^2_{GSK}(u)$ (Eq. 8).

Moreover, when the number of data in the neighbourhood reaches 30, it is possible to achieve smoother results. If the choice between more accurate and smoother results and less accurate and less smooth results is desirable, the authors suggest adopting the strategy proposed by Deutsch *et al.* (2014), who demanded that the type of estimation also be considered while defining search ranges.

In fact, in any estimation program, search range could be investigated by means of the QKNA parameter so as to insure the quality of the results.

Table 3 - Search ranges used in the estimations and QKNA parameter calculated for each result.

Ellipsoid No.	major search range	Mean data in neighbor	Real Corr	SOR	KV	KE	NW	WM (SK)	Leave 1 out Corr
				(proportion)					
	Low nugget effect								
1	9.16	0.1	0.215	0.000	1.00	0.20	0	0.998	0.06
2	18.32	0.4	0.254	0.074	0.99	0.20	0.000	0.967	0.08
3	27.48	1	0.417	0.317	0.94	0.21	0.000	0.845	0.15
4	36.64	2	0.497	0.533	0.87	0.23	0.002	0.667	0.36
5	45.8	3	0.521	0.798	0.76	0.26	0.007	0.473	0.69
6	54.96	4	0.511	0.944	0.66	0.30	0.019	0.311	0.810
7	64.12	5	0.492	0.9867	0.56	0.35	0.036	0.197	0.812
8	73.28	6	0.479	0.996	0.49	0.40	0.064	0.127	0.813
9	82.44	7	0.478	0.998	0.44	0.45	0.108	0.085	0.811
10	91.6	9	0.478	0.999	0.39	0.49	0.167	0.059	0.809
11	100.76	11	0.482	0.999	0.37	0.53	0.233	0.044	0.810
12	109.92	12	0.485	0.999	0.34	0.57	0.306	0.033	0.809
13	119.08	14	0.482	0.999	0.33	0.61	0.378	0.026	0.808
14	128.24	16	0.483	0.999	0.31	0.64	0.430	0.020	0.808
15	137.4	18	0.476	0.999	0.29	0.68	0.467	0.017	0.808
	High nugget effect								
1	9.66	0.2	0.138	0	0.95	0.47	0	0.999	0.06
2	19.32	0.5	0.155	0.074	0.94	0.48	0.000	0.979	0.08
3	28.98	1	0.274	0.204	0.92	0.49	0.000	0.895	0.13
4	38.64	2	0.336	0.410	0.88	0.51	0.001	0.760	0.33
5	48.3	4	0.358	0.620	0.83	0.54	0.004	0.601	0.59
6	57.96	4	0.353	0.849	0.77	0.58	0.009	0.450	0.710
7	67.62	5	0.339	0.955	0.71	0.63	0.019	0.327	0.726
8	77.28	5	0.327	0.988	0.66	0.68	0.033	0.233	0.727
9	86.94	8	0.319	0.997	0.61	0.74	0.053	0.167	0.721
10	96.6	10	0.316	0.998	0.58	0.78	0.083	0.123	0.722
11	106.26	12	0.312	0.998	0.55	0.82	0.123	0.093	0.722
12	115.92	15	0.307	0.999	0.53	0.85	0.169	0.072	0.722
13	125.58	18	0.306	0.999	0.52	0.86	0.226	0.057	0.722
14	135.24	21	0.303	0.999	0.51	0.88	0.285	0.046	0.722
15	144.9	23	0.294	0.999	0.50	0.90	0.342	0.038	0.723

Table 4 - The relative closeness of each alternative in the synthetic samples.

Alternative	Low nugget effect			High nugget effect		
1	0.324	0.157	0.326	0.323	0.147	0.313
2	0.308	0.158	0.340	0.307	0.148	0.325
3	0.254	0.184	0.420	0.268	0.167	0.384
4	0.149	0.261	0.637	0.164	0.245	0.599
5	0.065	0.317	0.830	0.094	0.290	0.756
6	0.040	0.344	0.900	0.042	0.327	0.885
7	0.039	0.346	0.899	0.044	0.336	0.844
8	0.043	0.343	0.890	0.030	0.339	0.918
9	0.050	0.339	0.872	0.035	0.339	0.910
10	0.060	0.335	0.848	0.054	0.332	0.860
11	0.076	0.330	0.812	0.068	0.327	0.828
12	0.096	0.326	0.773	0.084	0.323	0.793
13	0.133	0.320	0.706	0.103	0.319	0.756
14	0.084	0.330	0.798	0.125	0.316	0.716
15	0.158	0.319	0.669	0.150	0.314	0.676

Table 5 - Estimation parameter calculated for the optimal search range. Low (1st row) and high (2nd row) nugget effect cases.

Search range (m)	Mean data in neighbor	Real Corr	Min est var	SOR	KV	KE	NW	WM (SK)	Leave 1 out Corr
				(proportion)					
68	5	0.49	0.1972	0.995	0.5	0.05	0.4	0.2	0.812
81	6	0.32	0.4497	0.9976	0.63	0.048	0.713	0.196	0.727

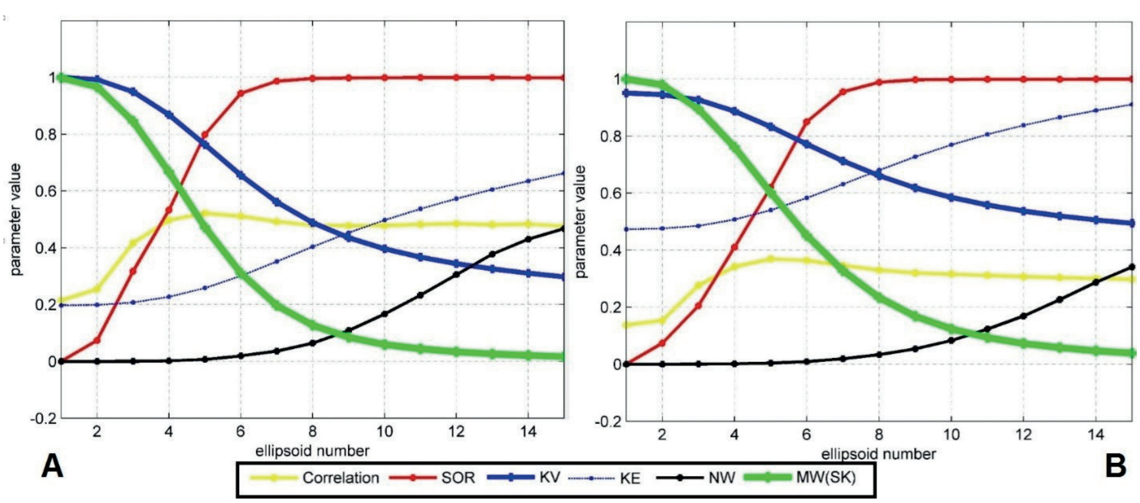

Fig. 5 - Sensitivity of QKNA parameter to search range. Low (A) and high (B) nugget effect cases.

6.3. Real examples (block kriging)

A topic that has not been tackled so far concerns validating kriging reproduction of global and local means for estimated domains by using different search ranges. The economic evaluation of a mining project is highly dependent on estimation results, and using biased and inaccurate results for this purpose could lead to drastic errors. In the remainder of this study, some more examples are presented to highlight how search range can influence the reproduction of samples average. For this, three different data sets with different distributions are explored. These data sets represent real case studies. It should be noted that statistical analysis and un-clustering were performed and then the data sets were run.

6.3.1. Statistical analysis of data sets

Sungun copper deposit is one of the biggest copper deposits in north-western Iran, containing about 817 m ore and a mean grade of 0.62% (Pars Olang Engineering Consultant Company, 2006). In this study, some of the information related to this deposit was used to clarify the effect of data distribution on QKNA results. The data set comprised 396 and 801 borehole samples for the first and second estimation phases, respectively, with lognormal distribution and positive kurtosis in two case studies (Table 6, Figs. 6 and 7). To analyse the estimation type and understand the results, in addition to the two cases of this example, the program was run for these two data sets and the results are presented in Table 7. The borehole distance is 90 and 65 for the first (Fig. 6) and second (Fig. 7) phases, respectively; therefore, these two cases can be considered distinct estimation stages.

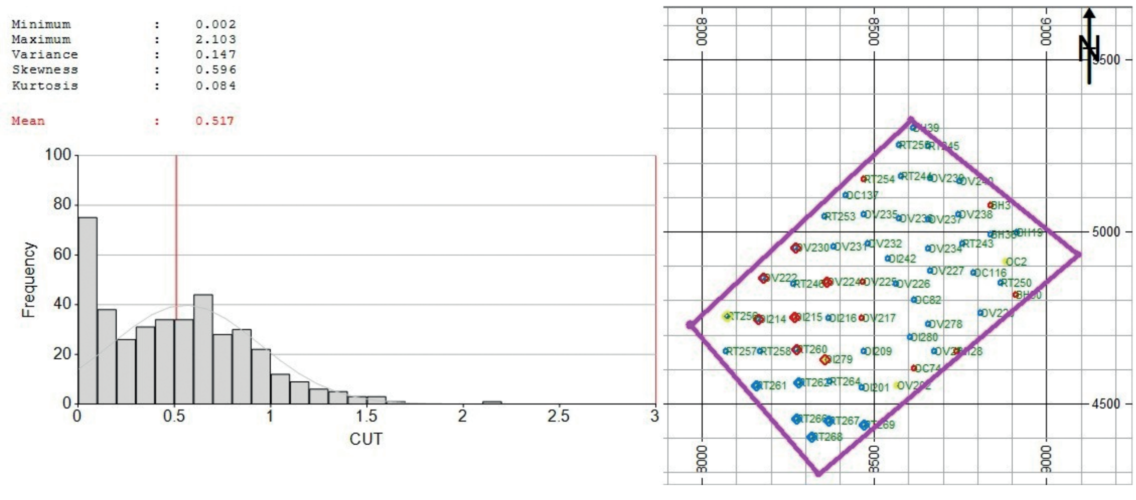

Fig. 6 - The position of drill holes (right) and histogram of Fe (left) in Sungun data set phase 1.

Although this deposit is part of a big iron ore deposit which has an excellent data set, the exploration program is no more running in this part of the deposit and there is no certainty regarding the mean of the data. This problem could have both positive and negative effects on the results. The negative point is that an extra uncertainty is added to the calculations. Consequently, in this regard, the authors decided to show how the results would be in the absence of certainty.

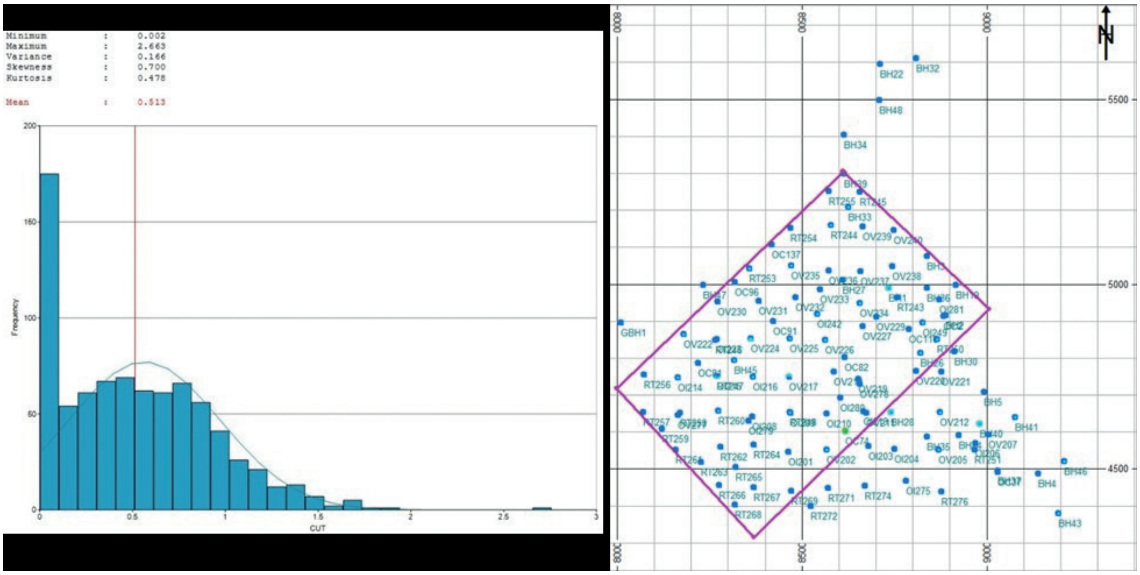

Fig. 7 - The position of drill holes (right) and histogram of Fe (left) in Sungun data set phase 2.

For these cases, the average of the estimated grid was calculated and found to be a further criterion among others. The related changes in different search ranges are also plotted in the graphs. Then, it is possible to investigate the reproduction of the real mean in the estimation result. To observe the reproduction of the local mean, the mean square error of the mean value of the model and composites was also assessed by swat plots.

Table 6 - Statistics of real case data sets.

Data set	Sungun phase 1	Sungun phase 2
Mean	0.5171	0.5132
Standard Deviation	0.383	0.4078
Sample Variance	0.147	0.1663
Kurtosis	0.084	0.488
Skewness	0.596	0.701
Distribution type	Lognormal (positive Skewness)	Lognormal (positive Skewness)
	0.0579	0.0608

6.3.2. Sungun results

The variogram model of Sungun 1 and 2 data sets is shown in the following equations (the variogram is calculated for de-clustered data; hence, it is different from primary data):

$$\gamma = 0.06 + 0.14*Sph(320, 161, 100), \mathrm{Az} = 135$$
$$\gamma = 0.06 + 0.13*Sph(311, 150, 90), \mathrm{Az} = 135.$$

The results of QKNA analysis are illustrated in Table 7, Figs. 8 and 9. Different steps of the Fuzzy-TOPSIS algorithm was applied to the decision making matrix (Table 7) in order to determine the optimal search range (Table 8). The results of the relative closeness analysis demonstrated that the optimal search ranges determined through QKNA criteria are ellipsoids 13 and 7 for phase 1 and 2, respectively (Table 8). As explained before, for some cases there is not a single option, and the selection of optimal search range would be an option. In such circumstances, one option is selected according to the estimation type and by considering the test parameter. For example, in the Sungun 1, which is designed to be a primary estimation for ellipsoid 13, the mean reproduction is excellent and after this range no significant improvement occurs in the test parameter; therefore, this search range could be a good choice. On the other hand, for Sungun 2, which is a resource estimation project, the minimum search range is the best option; hence, ellipsoid 7 should be chosen.

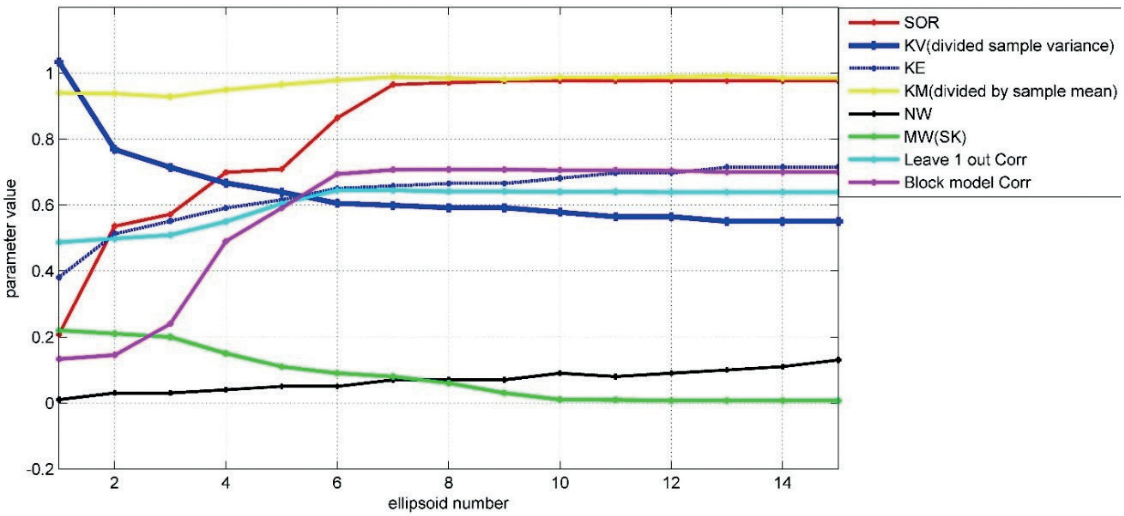

Fig. 8 - Alteration of QKNA parameter in Sungun 1 versus search range expansion.

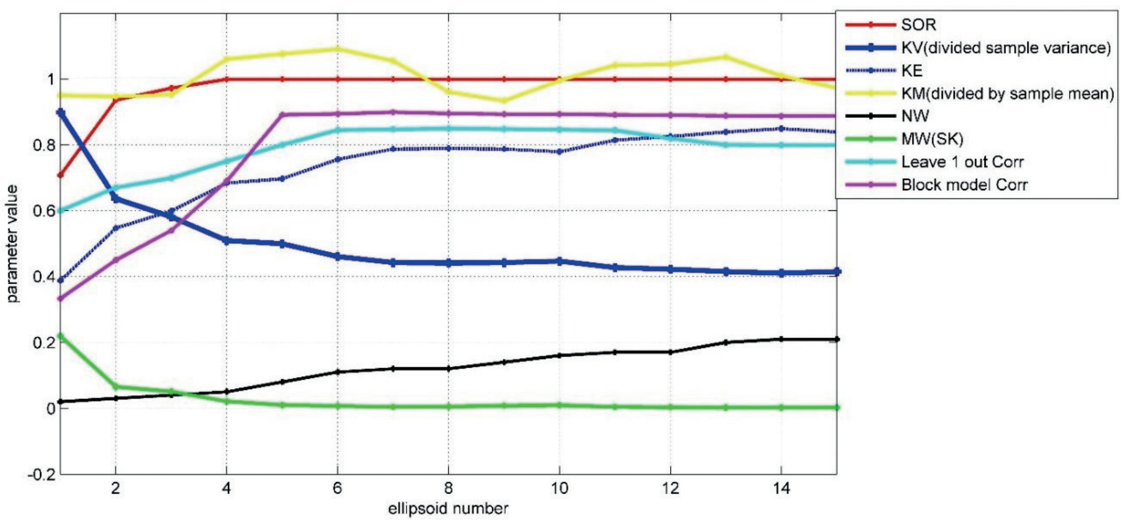

Fig. 9 - Alteration of QKNA parameter in Sungun 2 versus search range expansion.

Table 7 - Search ranges and QKNA parameter used in estimation of Sungun phase 1 and 2.

Ell. No.	Ave. data	major range	Estimation mean	SOR	KV	KE	NW	WM (SK)	leave-1-out cross validation	Block model Corr	MSE
Sungun phase 1											
1	2	39	0.486	0.208	0.152	0.38	0.01	0.22	0.487	0.133	0.100
2	5	78	0.485	0.536	0.113	0.51	0.03	0.21	0.498	0.145	0.140
3	8	117	0.480	0.572	0.105	0.55	0.03	0.20	0.509	0.240	0.020
4	11	156	0.491	0.699	0.098	0.59	0.04	0.15	0.550	0.490	0.170
5	16	195	0.499	0.709	0.094	0.62	0.05	0.11	0.604	0.591	0.084
6*	21	234	0.506	0.864	0.089	0.65	0.05	0.09	0.645	0.694	0.070
7	25	273	0.511	0.965	0.088	0.66	0.07	0.08	0.645	0.707	0.064
8	27	312	0.509	0.971	0.087	0.67	0.07	0.06	0.641	0.708	0.126
9	29	351	0.507	0.976	0.087	0.67	0.07	0.03	0.641	0.707	0.125
10	32	390	0.510	0.977	0.085	0.68	0.09	0.01	0.640	0.706	0.068
11	36	429	0.510	0.976	0.083	0.70	0.08	0.009	0.640	0.706	0.083
12	38	468	0.511	0.987	0.083	0.70	0.09	0.007	0.639	0.704	0.073
13	***40***	***507***	***0.516***	***0.997***	***0.081***	***0.71***	***0.1***	***0.007***	***0.639***	***0.700***	***0.060***
14	45	546	0.509	0.997	0.081	0.71	0.11	0.007	0.639	0.700	0.203
15	49	585	0.509	0.997	0.081	0.71	0.13	0.007	0.639	0.700	0.064
Sungun phase 2											
1	4	53.16	0.4876	0.708	0.1494	0.41	0.02	0.219	0.600	0.333	0.055
2	6	106.32	0.4856	0.936	0.1058	0.57	0.03	0.065	0.670	0.450	0.079
3	9	159.48	0.4889	0.972	0.0967	0.63	0.04	0.051	0.699	0.540	0.053
4	13	212.64	0.5440	0.999	0.0847	0.72	0.05	0.021	0.750	0.690	0.045
5	18	265.8	0.5521	0.999	0.0831	0.73	0.08	0.010	0.800	0.891	0.033
6*	25	318.96	0.5596	0.999	0.0766	0.79	0.11	0.007	0.845	0.894	0.026
7	***30***	***372.12***	***0.5214***	***0.999***	***0.0736***	***0.83***	***0.12***	***0.004***	***0.849***	***0.899***	***0.026***
8	34	425.28	0.4931	0.999	0.0734	0.83	0.12	0.005	0.849	0.895	0.035
9	36	478.44	0.4791	0.999	0.0736	0.83	0.14	0.008	0.848	0.893	0.036
10	39	531.6	0.5106	0.999	0.0743	0.82	0.16	0.009	0.846	0.893	0.037
11	42	584.76	0.5344	0.999	0.0711	0.86	0.17	0.004	0.844	0.891	0.037
12	44	637.92	0.5359	0.999	0.0702	0.87	0.17	0.003	0.819	0.890	0.027
13	50	691.08	0.5472	0.999	0.0690	0.88	0.20	0.002	0.800	0.888	0.034
14	56	744.24	0.5178	0.999	0.0682	0.89	0.21	0.002	0.799	0.887	0.066
15	59	797.4	0.4994	0.999	0.0690	0.88	0.21	0.002	0.799	0.887	0.066

*search ranges calculated through variogram. Best values are indicated in blue, non-acceptable values in red. Italics indicate optimum search range.

7. Conclusions

This paper investigated determination of kriging search range. It was discovered that selecting the search range through variogram range is not always the best choice, for it may cause

Table 8 - The relative closeness of each alternative in Sungun phase 1 and 2.

Alternative	Sungun phase 1			Sungun phase 2		
	d^+	d^-	rc	d^+	d^-	rc
1	0.644	0.323	0.334	0.194	0.153	0.441
2	0.304	0.420	0.580	0.167	0.157	0.485
3	0.249	0.521	0.676	0.157	0.162	0.508
4	0.308	0.821	0.727	0.057	0.215	0.789
5	0.401	0.958	0.705	0.083	0.207	0.713
6	0.940	0.718	0.433	0.056	0.219	0.796
7	0.919	0.892	0.492	0.045	0.283	0.863
8	0.888	0.227	0.204	0.071	0.212	0.749
9	0.772	0.728	0.485	0.055	0.222	0.802
10	0.307	0.744	0.708	0.052	0.224	0.812
11	0.225	0.849	0.790	0.085	0.206	0.709
12	0.431	0.926	0.682	0.107	0.197	0.649
13	0.002	0.963	0.998	0.152	0.191	0.557
14	0.245	0.612	0.714	0.153	0.190	0.554
15	0.487	0.910	0.651	0.154	0.192	0.555

conditional bias. Biasedness of results, especially in the final estimation, is not acceptable, and many strategies should be utilised during the estimation process to eliminate or at least reduce it. On the other hand, as an essential parameter for estimation quality, KE is directly related to search range and the number of data used. Using restricted neighbourhood could not satisfy conditional unbiasedness and afford high quality estimation. While large search ranges are not always the solution for low quality and biased result, establishing search ranges for different estimation applications is influenced by both QKNA criteria and KE.

Several case studies presented here displayed how estimation results are affected by neighbourhood parameter, nugget effect, data distribution, and data availability. The optimal search range was determined and estimation quality was explored by calculating QKNA. Choosing the best method with the high precision is consequential in selecting an optimal search range. In this study, the MCDM method was deployed to achieve this goal. In fact, after the QKNA criteria (as a decision making matrix) were computed for different search ranges based on the kriging method, the optimal search range was found using the MCDM method. Among various MCDM methods, Fuzzy-TOPSIS is a useful instrument for ranking alternatives. In this regard, the TOPSIS method was combined with the triangular fuzzy number to select the optimal alternative. A remarkable finding of this study is that the optimal search range is associated with many factors, which makes its discovery challenging. Some of the more salient factors in this context were discussed in this study.

As a recommendation for future work the authors suggest to use QKNA for kriging in a simulation program to see the effect of QKNA in simulation result.

Acknowledgements. The authors would like to acknowledge the financial support of University of Tehran for this research under grant number 27350/1/05.