An algorithm for quantifying regionalized ore grades

by B. Tutmez*, A.E. Tercan†, and U. Kaymak‡

Synopsis

We present a novel hybrid algorithm for quantifying the ore grade variability that has central importance in ore reserve estimation. The proposed algorithm has three stages: (1) fuzzy clustering, (2) similarity measure, and (3) grade estimation. The method first considers data clustering, and then uses the clustering information for quantifying the ore grades by means of a cumulative point semimadogram function. The method provides a measure of similarity and gives an indication of the regional heterogeneity. In addition, grade estimations can be obtained at different levels of similarity using a weighting function, which is the standard regional dependence function (SRDF).

Keywords: Grade, fuzzy clustering, similarity measure, point madogram, weighting function.

Introduction

Measured grades rely on the relative positions of measurement locations within the ore site. These measurements at a set of locations give some insight into regional variability. This variability determines the regional behaviour as well as the predictability of the grade. The larger the variability, the more heterogeneous is the geological environment[1]. One of the tools used to measure regional variability is the semivariogram (variogram), which provides a measure of spatial dependence among a multitude of locations as an alternative to the auto-covariance of a time series[2].

The classical variogram, although suitable for irregularly spaced data, has practical difficulties. One of the main drawbacks is that it is insufficient to analyse the regional heterogeneous behaviour of the grade. In general, ore deposits have heterogeneous properties rather than homogeneous structures. Heterogeneity means that the properties (grades) observed at different locations do not have the same value, and that different zones are observed in the ore site. In order to quantify the regional behaviours, a cumulative semivariogram (CSV) concept has been proposed by Şen[1] as an extension of the classical semivariogram. Şen[1] used the CSV in stochastic processes for analysing the regional correlation and concluded that CSV is a better tool than the classical variogram in identifying spatial dependence. Alternatively, a point cumulative semivariogram (PCSV) measure is proposed by Şen[3] in identifying the spatial behaviour of a regional variable around a location concerned. The basic principle of the technique is to compute experimental PCSVs for each data location, which leads to the estimation of the radius of influence around each location[4]. In some recent works[5,6], point cumulative semimadogram (PCSM) measure has been proposed instead of PCSV for modelling the regional spatial dependence due to the advantages of absolute difference[7].

This paper presents a hybrid methodology, which uses the fuzzy clustering based PCSM for identifying the regional dependence. The method proposed in the study uses both soft and probabilistic tools. The algorithm first considers the fuzzy information and then describes the regional variability based on the mean absolute difference measure. In addition, the algorithm allows the regional heterogeneity of the grade to be evaluated at fixed similarity levels. Finally, grade estimations are carried out at different levels using standard regional dependence function (SRDF).

* *Inonu University, Department of Mining Engineering, Malatya, Turkey.*

† *Hacettepe University, Department of Mining Engineering, Ankara, Turkey.*

‡ *Erasmus University Rotterdam, Econometric Institute, Rotterdam, The Netherlands.*

 SA ISSN 0038–223X/3.00 + 0.00. Paper received Dec. 2006; revised paper received Jan. 2008.

Algorithmic procedure

Determination of the regions by fuzzy clustering

In the first stage, the grade values are clustered into a number of groups using fuzzy clustering. For this purpose, the fuzzy c-means (FCM) clustering[8] is used. By this application, sample points within the subgroups are characterized by cluster centres. Each data point belongs to a cluster centre with a degree that is determined by the membership grade. The FCM algorithm is given in Appendix A[9].

The number of clusters is an important parameter that influences the accuracy and transparency of the models[10]. It has been considered extensively in the literature[11,12,13]. However, these methods are not suitable for describing spatial variability. In the present study, a novel cluster validity approach proposed in Tutmez *et al.*[6] specially for spatial estimations, is used. The approach is based on reproducing the variability of the sample data in the variability of the cluster centres, while using the minimum number of clusters possible:

$$\text{Min. } n_c \text{ under } \quad Std[g(x)\approx Std[g(c)] \qquad [1]$$

where n_c is the optimal number of clusters, $g(x)$ is the grade values in data set, $g(c)$ is the grade values in clusters, and *Std* is the standard deviation of the property values. In this technique, the number of clusters is plotted against the corresponding standard deviations of the cluster centre values and then the number of clusters satisfying Constraint [3] is retained as optimum.

Regional heterogeneity measure

Two grade values $g(x)$ and $g(x+h)$ at two points x and $x+h$ separated by the vector h are spatially correlated. As the distance between these grades increases, one would expect that the spatial correlation decreases and vice versa. The classical variogram is not convenient for describing the local variability. Therefore, the point semivariogram (PSV) function was proposed[1]. PSV can be used in determining the spatial behaviour of any variable around a particular data location. Its mathematical expression is given as follows,

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (g_m - g_{(m+h_i)})^2 \qquad [2]$$

where N is the number of data values, $N(h)$ is the number of pairs, g_m is the reference grade value, and $g_{(m+h)}$ is the grade value at a distance h.

Madograms are particularly useful for establishing the range parameter[14]. Therefore point semimadogram (PSM) was suggested[6] as an alternative measure for evaluating the local spatial behaviour of data and it was used by Tutmez and Hatipoglu[5]. By using this measure, the zone of influence around each point can be determined. The point madogram function is preferable to the point variogram. This function uses the absolute difference instead of squaring the difference between g_m and g_{m+h}. If the data-set includes the outlier values and the number of data is limited, the PSM is more convenient than the PSV due to the advantages of the absolute difference measure[6],

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left| g_m - g_{(m+h_i)} \right|. \qquad [3]$$

If the experimental variogram includes the outlier values and the number of data is limited, the PSM is more outlier resistant than the PSV[6]. In order to quantify the degree of regional variability for around each measurement location, the point cumulative semimadogram (PCSM) function is considered. This measure gives the regional effect of all the other data locations within the study area on the location concerned. The number of PCSMs is equal to the number of data locations.

In the present methodology, variogram modelling is carried out using the membership values, which are obtained from the fuzzy clustering algorithm and Cartesian coordinates. The methodology uses the assigned fuzzy sets of each location instead of its grade value. Traditional variogram function uses spatial coordinates and data values (grades) for analysis. However, our approach considers the membership values (μ_A) instead of the grade values. The main objective of this procedure is to evaluate the variability using fuzzy tools. The mathematical expression is given as,

$$2\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} \left[g(\mu_{Ai}) - g(\mu_{Ai} + h) \right]^2 \qquad [4]$$

where, $g(\mu_{Ai})$ is the membership value of the fixed (pivot) location considered.

Estimation by standard weighting

A sample PCSM leads to a non-decreasing function with distance. In this section, the standard regional dependence function (SRDF)[15,5], is applied. The SRDF provides weights for different regional locations depending on the distance from the pivot location. This weighting function value is calculated using the following steps:

- ➤ Find the maximum PCSM value, (γ_m), which is taken at the greatest distance, (d_m)
- ➤ Divide all the PCSM values by (γ_m). The result appears as a scaled form of the sample PCSM values within limits of zero and one
- ➤ Subtract the dimensionless PCSM values from one at each distance. The resulting non-decreasing function is named the standard regional dependence function (SRDF).

Before the interpolation, determination of the regional locations employed in estimation process is critical. For this purpose, a search domain is constructed. According to [5], a location x is defined to belong to domain Ω if the Euclidean distance between pivot location p and x_i is not greater than the range a of the location considered.

$$x_i \in \Omega \qquad \text{if } d(p, x_i) \le a \quad i = 1,2,\ldots,N \qquad [5]$$

where N is the number of data.

Finally, each grade value is multiplied by the corresponding standard weight and contributions for each location are calculated. For a pivot location, grade is estimated by dividing the total contributions by the total standard weights.

The steps of the proposed algorithm can be summarized as follows:

- *Step 1*—Collect ore grade values at different locations in the field
- *Step 2*—Normalize the data for the Cartesian (x,y) product space and the grades by using linear transformation
- *Step 3*—Cluster the data into c clusters
- *Step 4*—Determine the optimal number of clusters using the clustering validity index outlined earlier
- *Step 5*—Obtain the partition matrix, whose jkth element $\mu_{jk} \in [0,1]$ is the membership degree of data object x_k in cluster j
- *Step 6*—Obtain the one-dimensional fuzzy sets μ_{Aji} by projection onto the space of the input variables x_i, where the jth row of U contains a pointwise definition[16] of a multidimensional fuzzy set. For this procedure, the expresion $\mu_{Aj,i}(x_{jk}) = \mathrm{proj}_i(\mu_{jk})$ is used
- *Step 7*—Compute the PCSMs for each location and plot the PCSMs against the corresponding distances. For this, use the Cartesian coordinates and memberships only derived in Step 6
- *Step 8*—Demonstrate the grade variability by similarity maps for different fixed levels of PCSMs
- *Step 9*—Calculate the standard weightings for each location using distances between the pivot location and its neighbour locations
- *Step 10*—Evaluate the grades at different levels of similarity using similarity maps. Spatial variability of the grades is assessed using similarity measures at fixed levels of PCSM. The similarity measure indicates the presence of the heterogeneity between two locations[3]
- *Step 11*—Estimate the grades at different levels of similarity. Numerical value of the radius of influence at any location recognizes the neighbouring locations that should be taken into consideration in quantifying the grade variability at this location. The weights for different regional locations depending on the distance from the pivot location are calculated by the SRDF.

Application to grade estimation

Statistical data properties

The Sivas-Kalburcayiri lignite field in Turkey is considered in this case study. The locations of the 42 records for the upper sector of the field[6] were randomly selected (Figure1). The data were scaled by using a linear transformation between 1034 and 1735. Figure 2 gives the basic descriptive statistics and the distribution. As stressed by Wellmer[17], if the distribution has a single peak and is approximately symmetrical, then the assumption of normal distribution generally leads to acceptable results for geological and geochemical problems. Moreover, it is possible to treat the methodology presented earlier for lognormal distributed data. In this case, the grade values that describe a lognormal distribution are transformed logarithmically to a normal distribution.

For FCM clustering, the fuzziness parameter was selected arbitrarily as m=1.4 and three-dimensional clustering spaces were comprised of the Cartesian coordinates and grade

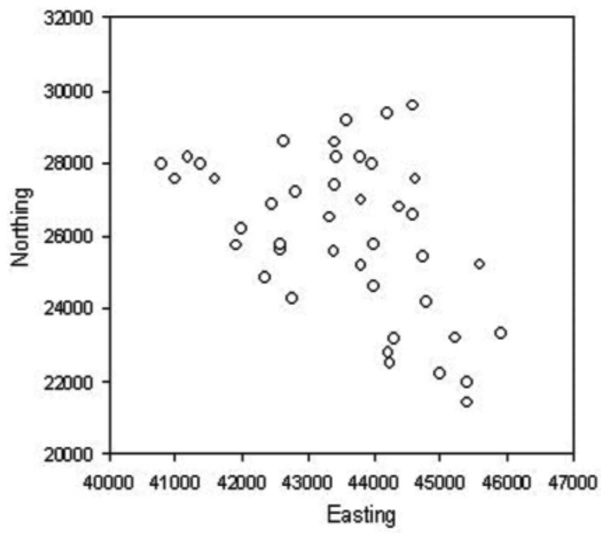
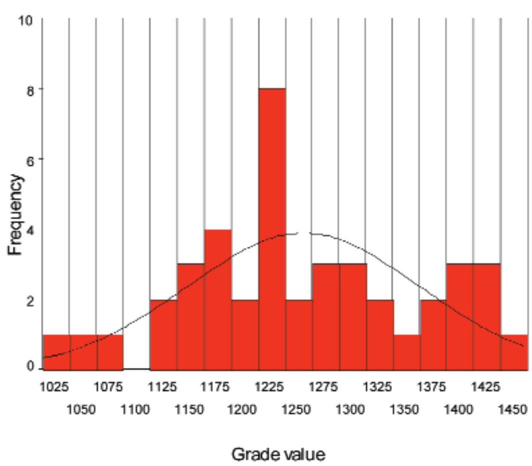


Figure 1—Sample locations of lignite data



Statistics	
Mean	1254.9
Std deviation	123.2
Skewness	0.06
Kurtosis	-0.58

Figure 2—Descriptive statistics and distribution

values. The optimal number of clusters was determined using the validity method presented earlier. In Figure 3, the number of clusters has been plotted against the corresponding standard deviations of the cluster centre grades. As can be seen from Figure 3, the appropriate number of clusters is six.

Using the FCM algorithm, a preliminary determination of the positions of different regions can be carried out. For representing the grade of each location, the membership to the closest prototype (max. membership) is calculated by one dimensional pointwise operator[16], which has been stated in the sixth step of the algorithm.

Analysis of regional variability

Since the grade distribution of the ore site is heterogeneous (see Figures 4 and 5), individual PCSMs for the grades show different behaviour. Calculated PCSMs are plotted on the vertical axis versus the corresponding distances on the horizontal axis and representative experimental PCSM diagrams are obtained, as shown in Figure 6.

Experimental madograms show that some of the PCSMs have straight lines on the horizontal distance axis. Most of the straight lines show the heterogeneity involved around the the location concerned at different distances. In addition, all the PCSMs pass through the origin. It means that there are no nugget effects (residual influence) within the regional grade variability ($C_0=0$).

Similarity evaluation

As a result of the PCSM analyses, the interpolated distances at these levels are obtained. Table I shows the distance values for two different levels. Fuzzy clustering based PCSM is an indicator of cumulative similarity of the variability for a location. Similarity contour maps at 0.2 and 0.4 levels of PCSM are shown in Figure 7 and Figure 8, respectively.

In this application, the radius of influence is determined to vary between 0.2 (60–250 m) and 0.4 (100–450 m) levels, respectively. The smaller distance indicates the more grade variability and the smaller regional dependence. For example, in Figure 7, the map shows three regions, which have intense grade variability with different similarity contours. On the other hand, Figure 8 presents the lower level of similarity. By means of these maps, the differentiated variability zones (heterogeneity) can be determined easily.

Weightings and estimations

In order to obtain the estimated values, search domains have been constructed. Figure 9 shows the sample search domain determination for location no. 6 at 0.2 level of PCSM. The search radius of this domain, which is a fixed distance at 0.2 level of PCSM, has been defined using Figure 6 (PCSM no. 6).

By using the search domain, scaled distances and distance ratios have been calculated at the different levels of similarity. In addition, SRDF weightings have been obtained by the difference between the locations in the domain and the pivot location (no. 6) easily. The sixth column in Tables II and III includes these SRDF weightings for location no. 6, which can also be taken from the graph in Figure 10.

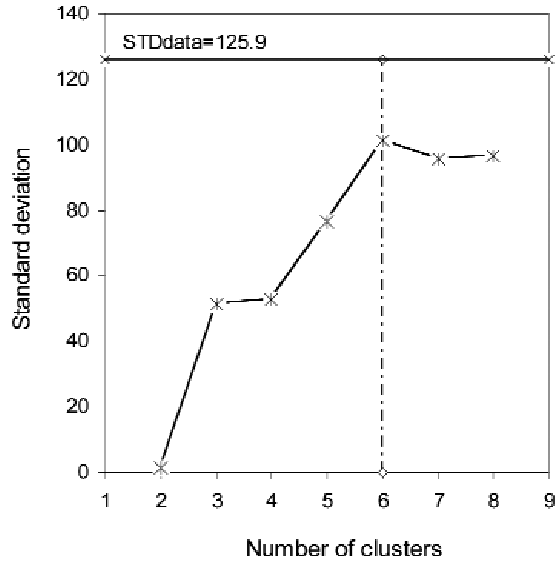


Figure 3—Cluster validity study

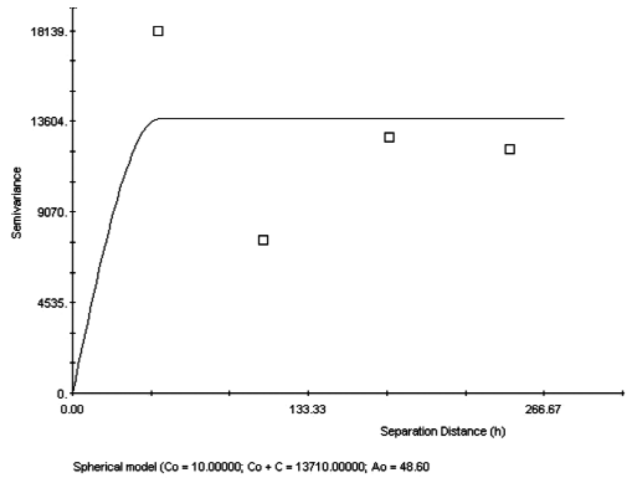


Figure 4—Variogram model of traditional method

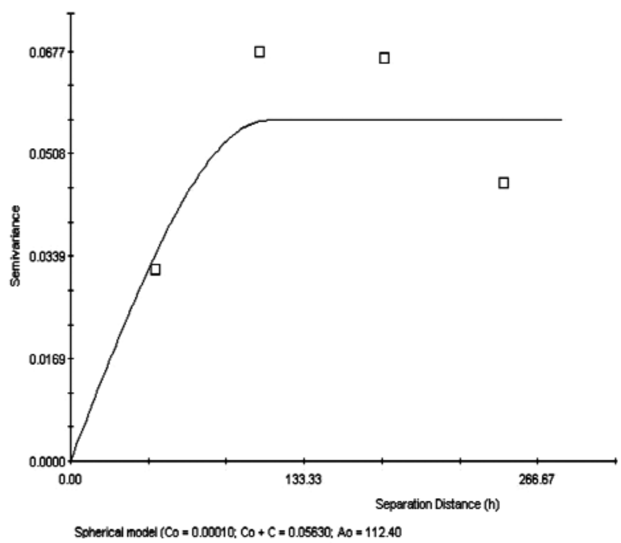


Figure 5—Variogram model of the proposed method

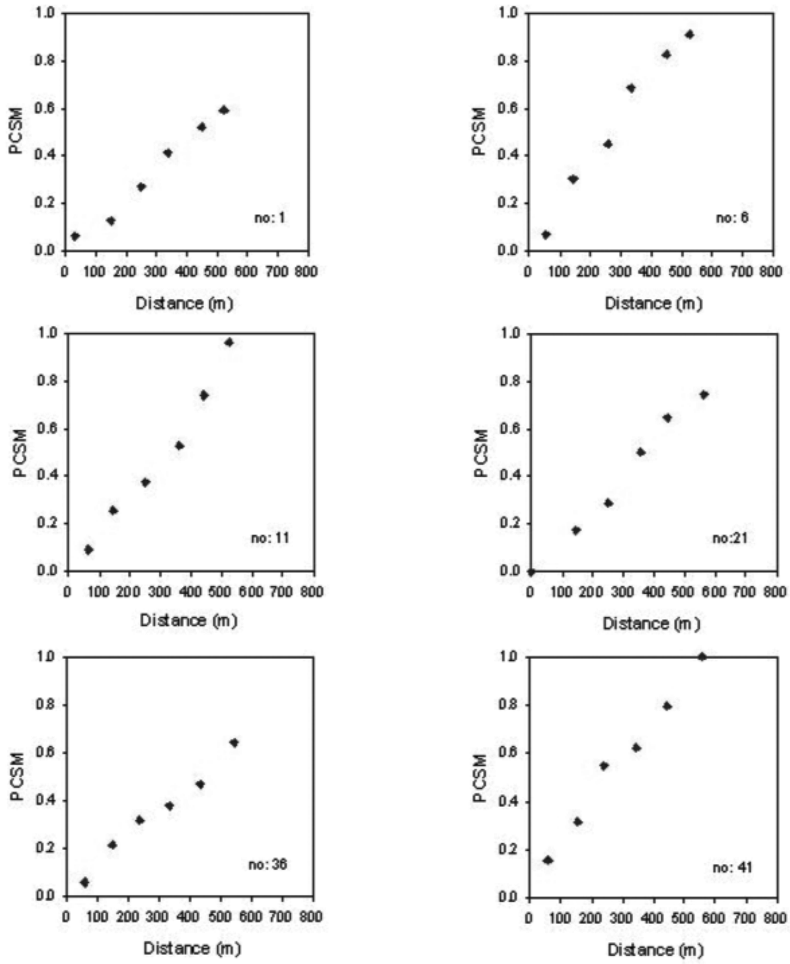


Figure 6—Sample experimental PCSMs at different measurement locations

Table I

Regional variations at fixed levels of PCSM

Location number	0.2	0.4	Location number	0.2	0.4
1	185	330	22	206	457
2	168	358	23	84	294
3	173	302	24	118	169
4	132	297	25	137	324
5	118	212	26	98	229
6	96	231	27	67	101
7	218	382	28	97	249
8	156	370	29	184	391
9	225	382	30	79	259
10	120	332	31	145	335
11	111	268	32	217	266
12	228	337	33	139	270
13	98	169	34	243	427
14	204	387	35	181	389
15	142	359	36	134	351
16	181	324	37	116	237
17	59	223	38	124	284
18	158	328	39	157	307
19	183	368	40	157	336
20	154	256	41	74	192
21	167	285	42	99	238

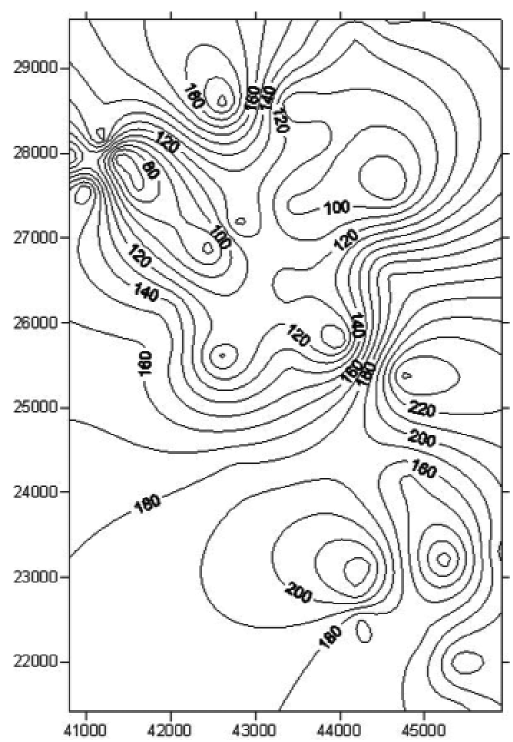


Figure 7—Similarity map at 0.2 PCSM value

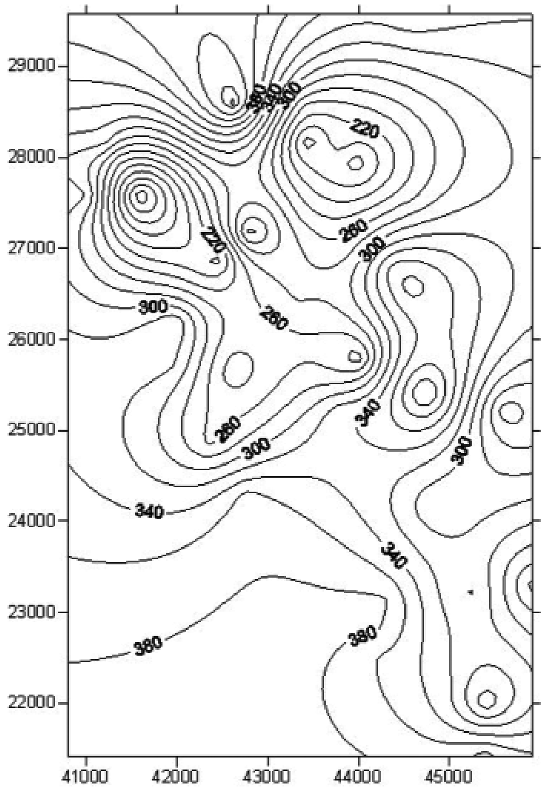


Figure 8—Similarity map at 0.4 PCSM value

In Figure 10, the closest location to the pivot contributes the highest weight, and the furthest ones relatively contribute the least weights. In order to obtain the estimated grades, spatial interpolations have been carried out for each location. The seventh column in Table II is the grade contribution, which was calculated by multiplying the grade by the SRDF values as weights. For location no. 6, substitution of the values leads to the estimation of grade at level 0.2 level as 3706.3/3.3 = 1133.8 (see Table II).

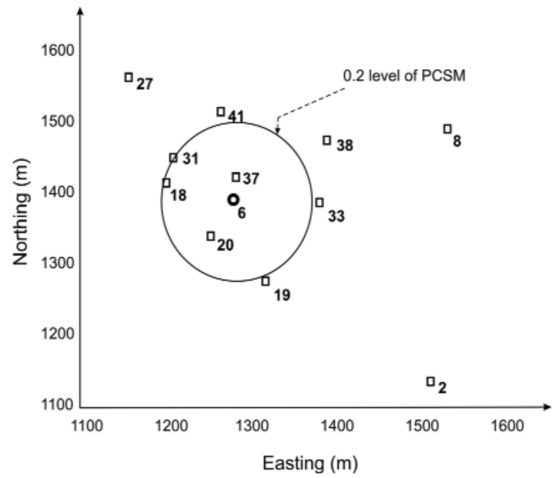


Figure 9—Determination of search domain for location no 6

Table II

Grade estimation for location no. 6 at 0.2 similarity level

Location distance	Grade ratio	PCSM weighting	Scaled	Distance	SRDF	Contribution	Estimation
6	1136	0.0	0.0	0.00	1.00	1136.0	1136
37	1155	9.5	17.0	0.18	0.90	1037.0	
20	1145	14.0	71.1	0.74	0.85	972.6	
18	1075	44.5	93.9	0.98	0.52	560.6	
31	1039	93.0	96.0	1.00	0.00	0.0	
				∑=3.27	∑=3706.3	3706.3/3.27=1133	

Table III

Grade estimation for location no. 6 at 0.4 similarity level

Location	Grade	PCSM	Scaled distance	Distance ratio	SRDF weighting	Contribution	Estimation
6	1136	0.0	0.0	0.00	1.00	1136.0	1136
37	1155	9.5	17.0	0.08	0.98	1132.3	
20	1145	14.0	71.1	0.33	0.97	1111.9	
18	1075	44.5	93.9	0.43	0.91	976.2	
31	1039	93.0	96.0	0.44	0.81	839.4	
33	1168	109.0	109.0	0.50	0.77	905.0	
41	1273	177.5	111.3	0.51	0.63	806.1	
19	1158	188.5	116.5	0.54	0.61	707.0	
38	1180	210.5	127.2	0.58	0.57	666.8	
36	1224	254.5	138.1	0.63	0.47	580.4	
40	1225	299.0	169.9	0.78	0.38	468.2	
28	1308	385.0	189.0	0.87	0.20	267.5	
26	1178	406.0	193.7	0.89	0.16	189.8	
11	1247	461.5	203.9	0.94	0.05	58.0	
1	1116	471.5	210.4	0.97	0.03	28.8	
27	1161	484.0	217.6	1.00	0.00	0.0	
	∑=8.55	∑=9873.4	9873.4/8.55=1155				

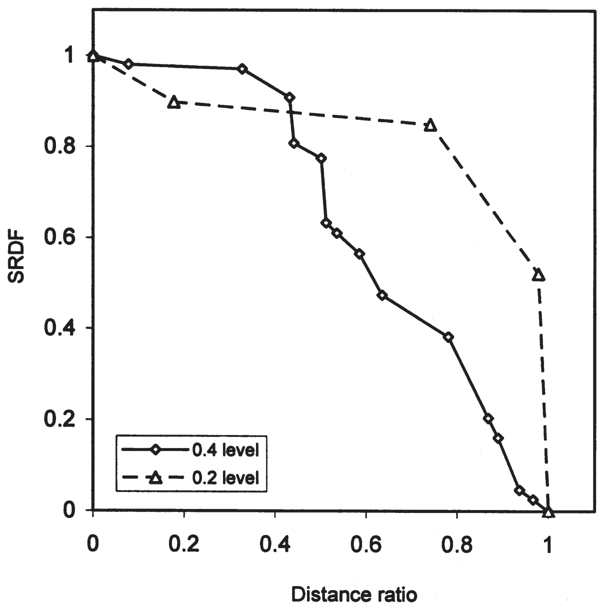


Figure 10—SRDF graphs at fixed levels for location no 6

Results and discussion

In the last step, it was tried to measure the strength of the linear relationship that exists between actual and estimated grades. In order to determine this relation, the coefficient of correlation, r, has been used. Figures 11–12 show these relationships for different fixed levels. The performance evaluation indicated better performance at 0.2 similarity level than at 0.4 similarity level.

As seen in Figures 11–12, for relatively small radius of influences (60–250 m) determined, more successful estimations have been obtained. On the other hand, with the increasing distances (100–450), less successful results have been taken. It might result from the manner in which the madogram function increases at small distances characterizes the degree of spatial continuity of the variable under study.

Conclusions

Quantification of grade variability has crucial importance in reserve estimation. Therefore different well-known methods are used based on pure probabilistic theory for this purpose. This study presented an integration of the probabilistic and fuzzy approaches. The proposed method uses membership grades and quantifies the variability using the point cumulative semimadogram function.

The approach can be used especially for determining the regional dependence and the radius of influence (using fixed PCSM levels). In addition, the method provides a general information to interpret spatial variability at each location rather than regionally. By this method, grade estimations also can be carried out at different PCSM levels by determining the search domains and standard regional weightings.

Appendix A

Fuzzy c-means (FCM) clustering

Given the data set **X**, choose the number of clusters $1 \prec c \prec N$, the fuzziness (weighting) exponent $m \succ 1$, the termination tolerance $\epsilon \succ 0$ and the norm-inducing matrix **A**. Initialize the partition matrix randomly, such that $U^{(0)}$.

Repeat for $l = 1, 2, \ldots$

Step 1: Compute the cluster prototypes (means):

$$\mathbf{v}_i^l = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m \mathbf{x}_k}{\sum_{k=1}^{N} \left(\mu_{ik}^{(l-1)}\right)^m}, \quad 1 \le i \le c.$$

Step 2: Compute the distance:

$$D_{ikA}^2 = \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right)^T \mathbf{A} \left(\mathbf{x}_k - \mathbf{v}_i^{(l)}\right), 1 \le i \le c, 1 \le k \le N.$$

Step 3: Update the partition matrix:
if $D_{ikA} \succ 0$ for $1 \le i \le c$, $1 \le k \le N$,

$$\mu_{ik}^{(l)} = \frac{1}{\sum_{j=1}^{c} \left(D_{ikA} / D_{jkA}\right)^{2/(m-1)}},$$

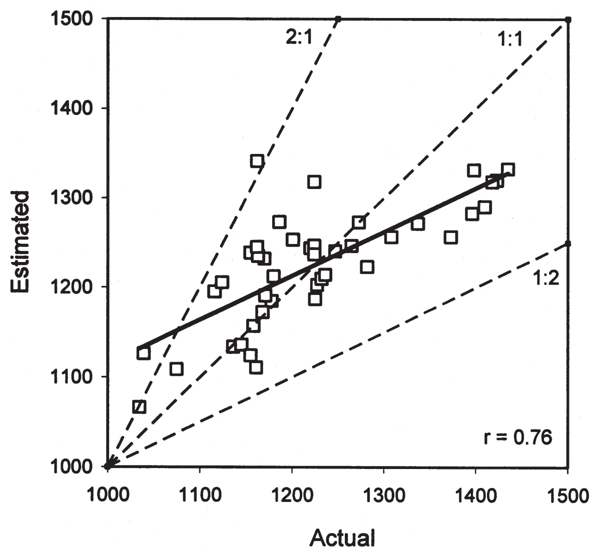


Figure 11—Performance evaluation at 0.2 level

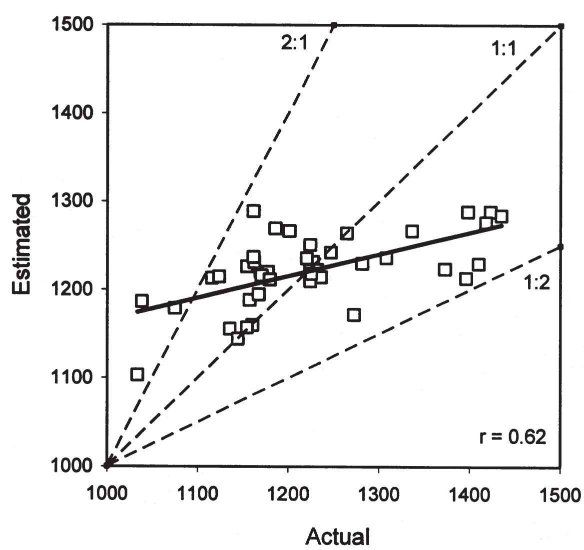


Figure 12—Performance evaluation at 0.4 level

otherwise

$$\mu_{ik}^{(l)} = 0 \text{ if } D_{ikA} \succ 0, \text{ and } \mu_{ik}^{(l)} \in [0,1] \text{ with } \sum_{i=1}^{c} \mu_{ik}^{(l)} = 1.$$

$$\textbf{until} \left\| U^{(l)} - U^{(l-1)} \right\| \prec \in .$$

References

1. Şen, Z. Cumulative semivariogram model of regionalized variables. *Mathematical Geology*, vol. 21, 1989. pp. 891–903.
2. Journel, A.G. and Huijbregts, Ch.J. *Mining Geostatistics*. Academic Press, 1981.
3. Şen, Z. Point cumulative semivariogram for identification of hetero-geneities in regional seismicity of Turkey. *Mathematical Geology*, vol. 30, no. 7, 1988. pp. 767–787.
4. Oztopal, A. Artificial neural network approach to spatial estimation of wind velocity data. *Energy Conversion and Management*, vol. 47, no. 4, 2006. pp. 395–406.
5. Tutmez, B. and Hatipoglu, Z. Spatial estimation model of porosity. *Computers & Geosciences*, vol. 33, 2007. pp. 465–475.
6. Tutmez, B., Tercan, A.E., and Kaymak, U. Fuzzy modeling for reserve estimation based on spatial variability. *Mathematical Geology*, vol. 39, no. 1, 2007. pp. 87–111.
7. Goovaerts, P. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 1997.
8. Bezdek, J.C., Ehrlich, R., and Full, W. FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences*, vol. 10, nos. 2-3, 1984. pp. 191–203.
9. Babuska, R. *Fuzzy Modelling for Control*. Kluwer Academic, 1998.
10. Sousa, J.M.C. and Kaymak, U. *Fuzzy Decision Making in Modelling and Control*, World Scientific, Singapore, 2002.
11. Xie, X.L. and Beni, G.A. A validity measure for fuzzy clustering. *IEEE Trans Pattern Anal Mach Intell*, vol. 13, no. 8, 1991. pp. 841–847.
12. Kaymak, U. and Babuska, R. Compatible cluster merging for fuzzy modelling. *Proc. FUZZ-IEEE/IFES '95*, Yokohama, 1995. pp. 897–904.
13. Kaymak, U. and Setnes, M. Extended fuzzy clustering algorithms. *ERIM report series Research in Management*, Erasmus University Rotterdam, ERS-2000-51-LIS. 2000.
14. Deutsch, C.V. and Journel, A.G. GSLIB: *Geostatistical Software Library and User's Guide*. 2nd edn. Oxford University Press, New York, 1998.
15. Tarawneh, A.D. and Şahin, A.D. Regional wind energy assessment technique with applications, *Energy Conversion & Management*, vol. 44, no. 9, 2003. pp. 1563–1574.
16. Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. *Fuzzy Cluster Analysis*, John Wiley & Sons, Chichester, 1999.
17. Wellmer, F.-W. *Statistical Evaluations in Exploration for Mineral Deposits*, Springer, Heidelberg, 1998. ◆