Discriminating copper geochemical anomalies and assessment of their reliability using a combination of sequential Gaussian simulation and gap statistic methods in Hararan area, Kerman, Iran

S. Abbaszadeh¹, F. Mohammad Torab¹*, A. Alikhani² and H. Molayemat¹

¹. Department of Mining and Metallurgical Engineering, Yazd University, Yazd, Iran
². Mines and Metals Development Investment Company, Iran

Received 3 May 2018; received in revised form 17 August 2018; accepted 25 August 2018

Keywords

Abstract

Sequential Gaussian Simulation
Threshold
Gap Statistics
Reliability
Hararan District

In geochemical exploration, there are various techniques such as univariate and multivariate statistical methods available for recognition of anomalous areas. Univariate techniques are usually utilized to estimate the threshold value, which is the smallest quantity among the values representing the anomalous areas. In this work, a combination of the Sequential Gaussian Simulation (SGS) and Gap Statistics (GS) methods was utilized as a new technique to estimate the threshold and to visualize the anomalous regions in the Hararan area, which is located in SE Iran, and consists of copper mineralization that seems to be connected to a porphyry Cu-Mo system. Furthermore, the most important advantage of this method is the reliable assessment of the anomalous areas. In other words, the anomalous areas were discriminated in terms of their probability values. The regions with high probability values were reliable and appropriate to locate the drilling points for a detailed exploration. It not only decreases the risk, cost, and time of exploration but also increases the drilling point reliability and precision of reserve estimation after drilling. In this research work, the results of analysis of 607 lithogeochemical samples for the element Cu were used. The SGS method was performed on the transformed data and 50 realizations were obtained. In the next step, the back-transformed realizations were utilized to obtain an E-type map, which was the average of 50 realizations. Moreover, the results of the GS method showed that the Cu threshold value was 228 ppm in the area. Therefore, using the E-type map, areas with values greater than 228 ppm were introduced as the anomalous areas. Finally, the probability map of the exceeding threshold values was acquired, and the anomalous districts located in the southern part of the studied area were considered as more reliable regions for future detailed exploration and drilling.

1. Introduction

The purpose of geochemical exploration is to discover and evaluate the anomalous populations of ore elements [1]. In mineral resource categorization, mine planning, and mineral exploration, identification of a geochemical background and its distinction from the geochemical anomalies are very important to recognize, delineate, and model the mineral zones [2]. Separating a geochemical background from the anomalous areas and then identifying the mineralized areas is the key to geochemical data processing and to recognize the threshold [3, 4]. There are various techniques available such as mean+ 2SDEV to estimate the threshold, which are still applied approximately 50 years after its beginning [5], and fence, median+2MAD, Gap Statistics (GS), and other spatial methods including the fractal and multi-fractal analysis techniques. Many authors have used the mentioned techniques in their research works.

* Corresponding author: fmtorab@yazd.ac.ir (F. Mohammad Torab)
In addition, there are some multivariate analysis methods involving factor analysis, cluster analysis, principal component analysis, and so on, which are used to identify the anomalous areas that have been applied by a lot of researchers [21-30]. In order to forecast the spatial attributes and model the uncertainty of predictions in locations that are un-sampled, the geostatistical techniques have been increasingly applied as powerful tools [2]. These methods have recently been applied in geochemical studies by many researchers [31-35]. An important geostatistical interpolation technique is called kriging, which is a robust estimator and visualizer, but its smoothing effect, particularly for the data that is skewed, is its major disadvantage [2]. In order to overcome the smoothing effect of the kriging estimator, the conditional stochastic simulation has been designed [36]. On the other hand, if the distribution of dataset is non-Gaussian and kriging is used, spatial heterogeneity that is the attribute of a lot of such datasets is not capable of being reproduced. To the contrary, there is Gaussian simulation as an alternative method, which provides more accurate results [2]. When the distributions of continuous variables transform to Gaussian or multi-Gaussian, they will be simulated by the Gaussian simulation methods [2, 37]. In mining, Gaussian simulations are most general. Although there are various Gaussian simulations that are used as well as others, Sequential Gaussian Simulation (SGS) is the most commonly used method [38]. SGS was first commenced by Isaaks (1990), and was based upon the multi-Gaussian RF model assumptions [38]. This method has been applied by many researchers in mining industries [2, 36, 39]. In this work, a combination of the GS and SGS methods was used to separate and delineate the anomalies from the background districts. Moreover, this method was utilized to assess the reliability of the anomalous regions, which is the most important ability of this technique. In other words, the probability values, which are the base of the discriminating regions, were calculated in the anomalous areas. In a detailed exploration, regions with high probability values are reliable and applicable to locate the drilling points. This makes the precision of reserve estimation to increase and the cost and time to decrease in a drilling project.

The utilized data was obtained from the lithogeochemical samples in the Hararan area, which appeared to possess the potential for Cu-Mo porphyry mineralization. The deposit was located in the northern latitudes 56°, 40′, 49″ to 56°, 42′, 40″ and the eastern longitudes 29°, 27′, 20″ to 29°, 29′, 46″ in the Baft geological sheet (1:100,000 series) in SE Iran [40]. The samples were systematically collected according to a regular grid pattern. The sampling density was 48.56 samples/Km² that corresponded to 1:5000 lithogeochemical surveys. 607 rock samples were analyzed by Amdel laboratory for 45 elements. However, only the results for the element Cu were used in this work. The map of the studied area and the location of the sampling points are shown in Figure 1(a). In this research work, the geostatistical studies were applied using the SGeMs, WinGslib, Golden Surfer, and GSFT softwares.

2. Geological setting
The most ancient rock units cropped out extensively in most parts of the studied area belong to the Eocene period. These units including the andesite and andesite porphyry and andesite volcanic breccias diorite to granodiorite rocks placed in the north, SE, and south parts of the studied area, respectively. Moreover, the tonalite and granodiorite rocks belong to the Eocene period placed in the central, west, and NW parts of the area. As well, dacite dykes with NE to SW trend are dispersed in most parts of the area. In addition to the felsic to intermediate rocks mentioned earlier, the other outcrops with Quaternary period mainly consist of colluviums with andesite volcanic, colluvium with agglomerate component, and quartz stockworks [41] (Figure 1(b)).

3. Methodology
3.1. Sequential gaussian simulation
Sequential simulation is a stochastic modeling algorithm obtaining multiple realizations based on the same input data. This data could be either categorical or continuous [42, 43]. Regarding the kind of data, Sequential Gaussian Simulation (SGS), Direct Sequential Simulation (DSS) or Sequential Indicator Simulation (SIS) are utilized [44]. The mainly straightforward algorithm to produce realizations of a multivariate Gaussian field is given by the sequential theory [45, 46]. SGS requires standard Gaussian data with unit variance and zero mean, so for SGS, the data is transformed to Gaussian through a quantile transformation [47]. Each variable is simulated sequentially in accordance with its normal conditional cumulative distribution function.
(CCDF) via a simple kriging estimation method. The conditioning data comprising all the formerly simulated values and all the original data exists within a neighborhood of the position being simulated [45, 46, 48]. Performance of the SGS method consists of the following steps [38].

![Diagram](image)

**Figure 1.** (a) Location of the Hararan area and its sampling point: left picture shows the geographical position of important copper porphyry deposits and Cu porphyry mineralization of Hararan in the Urumieh-Dokhtar magmatic arc. Location of the studied area is defined by a triangle shape. Right picture depicts the lithogeochemical sampling points in the Hararan area. (b) Geological map of the Hararan area (1:5000): most rock units cropped out extensively in most parts of the area belong to the Eocene age.
1. Complete a whole exploratory data analysis of the original data comprising the domain definition and variography.
2. After identifying the domains, examine whether the data is required to be de-trended, i.e. whether the simulation should be applied to the residuals.
3. Obtaining the matching Gaussian distribution, carry out the normal score transformation to the original data.
4. Acquire the variogram models with a Gaussian distribution for the variable that is transformed.
5. Draw a chance path through each domain to be simulated. To avoid artifacts, the route for the simulation is randomly delineated.
6. For each node to be simulated, estimate the conditional distribution via simple kriging in the Gaussian space. The variance of conditional distribution is called the simple kriging variance (σsk (u)), and its mean is named the estimated simple kriging value (Y*(u)). The Gaussian mean of distribution is zero while simulation is applied to residuals later than de-trending.
7. Obtaining a simulated value for each node, Ys (u), draw from the conditional distribution at random.
8. Add the simulated value (Ys (u)) as the conditioning data for the nodes that are simulated later. This is essential to guarantee the variogram reproduction.
9. Loop the process until all domains and all nodes have been simulated.
10. Ending the simulation, check that histogram (univariate distribution) of values that is simulated is Gaussian; also verify the variogram model of simulation the same as the original model variogram.
11. Back-transform the normalized simulated values into the original variable space.
12. Add back the trend if the simulation was performed on residuals.
13. Check that the original distribution and the variogram of the original values are the same as the histogram for the back-transformed data and the variogram that is obtained from the simulated values, respectively.

It should be regarded although there are statistical fluctuations in simulation while data distribution is transformed to Gaussian and vice versa; they should be unbiased and logical in the variance and mean [49].

The following controls should be carried out after having all nodes simulated. Reproduction of [49]:
1) The actual summary statistics;
2) The data values at data positions;
3) The enter covariance model;
4) The actual histogram.

3.2. Gap statistics
The gap statistics (GS) method is applied while there is a more delicate gap in the data. In this method, the data distribution should be Gaussian; thus in the first step, the data must be converted so that they can conform the normal form as well as possible. The second step is the standardization of data so that they have a mean with zero value and a variance equal to one; the goal of standardization is to remove the effects of scales. The standardized data is called (Z). The third step is to sort the transformed data in an ascending or descending form; the next steps include the following [50, 51]:

The mean of two standardized values that were placed sequentially was obtained. These values were called (mi).

Obtain the absolute distinction between the resulting succeeding values in the ordered array, which may be named the standardized gaps. The greatest standardization gaps will tend to take place next to tails of distribution, and approximately never take place next to the mean. Nevertheless, the geochemical threshold values can take place anywhere through the variety of values; this trend should be removed. This is applied by multiplying the gap standardized by the supposed frequency at the center of gap as specified from an appropriate normal curve. Therefore, the calculated values are called the adjusted gaps (Gi).

The greatest value is selected among the calculated Gi. Then mi corresponding to Gi is introduced as the gap statistics value. Finally, the threshold is acquired using Equation 4, in which the calculation stages are shown by the following equations:

\[ G_i = F(m) \left[ Z_{j+1} - Z_j \right] \quad i = 1, 2, 3 \ldots \]  \hspace{1cm} (1)

\[ F(m) = 0.3989e^{-\frac{1}{2m^2}} \]  \hspace{1cm} (2)

\[ m_i = \frac{Z_{j+1} + Z_j}{2} \]  \hspace{1cm} (3)

**Threshold = (gap statistics × standard deviation) + mean**  \hspace{1cm} (4)

4. Results and discussion
4.1. Descriptive statistics
Before calculating the primary statistical parameters, preprocessing of the geochemical data including replacement of the censored and outlier
data was performed. The histogram and descriptive statistics of raw copper concentrations from 607 samples are presented in Figure 2 and Table 1, respectively. The values representing the mean, standard deviation, and skewness were equal to 118.9 ppm, 243.35 ppm, and 5.82 ppm, respectively. As illustrated in the histogram and skewness coefficient, the copper distribution is highly far from normal and positively skewed. In order to perform the SGS and GS methods, the data should be transformed by a suitable transformation function to follow a Gaussian distribution. Note that due to the influence of skewness, a method like ordinary kriging can produce poor results. Since the input data for the SGS and GS methods have to be a standard Gaussian distribution, first, the copper data was transformed by utilizing a log transformation function. Then the logarithmic data was standardized using the Z score method; the mean distribution was subtracted from each observation and divided by the standard deviation of the distribution. The statistical parameters of the transformed data (i.e. the variance value close to 1 and mean of about 0) control the accuracy of the transformation. The histogram of the new variable with Gaussian distribution and its statistical parameters are shown in Figure 3 and Table 1, respectively. Regarding the statistical values such as variance (equal to 1), mean (equal to 0), and skewness (equal to 0), the correctness of transformation was confirmed.

| Table 1. Descriptive statistical parameters of element Cu in the Hararan area. |
|------------------|------------------|------------------|------------------|------------------|------------------|
|                  | Mean (ppm)       | Median (ppm)     | Std. Deviation (ppm) | Variance (ppm)^2 | Skewness         |
| Cu               | 118.9            | 62               | 243.35            | 59220.68         | 5.82             |
| Transformed data | 0                | 0                | 1                 | 1                | 0                |

4.2. GS on simulated data
In order to compute the threshold of the Cu data in the Hararan area, the GS method was used. Based on the principles of GS mentioned in Section 3.2, a programming code was used. Consequently, the threshold value of the Cu data was obtained to be 228 ppm in the studied area.

4.3. Simulation of copper concentration based on SGS
4.3.1. Spatial analysis
In order to explain the spatial structure and control the anisotropy of the transformed data, the omni-directional and several directional semi-variograms (for directions N-S, N45°E, E-W, and N45°W) were computed with 22.5 degree tolerance, and were modeled. Then the theoretical models of spatial variability were fitted to the experimental semi-variograms. The spatial model of variability for the transformed Cu data depicted two components of continuity: a spherical structure with geometry anisotropy and a nugget effect. Interestingly, the region had a maximum range (600 m) in the azimuth 90° that could verify the main trend of Hararan mineralization as being in the W-E orientation with an anisotropy ratio of 1.5.

Figure 4 displays the semi-variograms in two directions on the basis of the transformed data with the spherical model that is fitted to the experimental variogram, while Table 2 depicts the parameters of the theoretical model fitted to the
semi-variograms. As it can be seen, a high content of the nugget effect demonstrates the high variability of the regional variable (Cu) even at short ranges, and the variograms show periodic variations (hole effect), especially in the N-S direction (AZ = 0), representing periodic grade fluctuation or vein type mineralization.

4.3.2. Geostatistical simulation

SGS was implemented by the SGSSIM algorithm in the SGeMS software, and a regular 2-D grid with the 64 × 62 m blocks were produced within the estimation space. For each block, fifty conditional simulations were produced by utilizing the semi-variogram model parameters of the transformed copper data and the ordinary kriging estimator. The representation of two realizations, which were selected randomly, and an E-type map, which was obtained from averaging 50 realizations, are displayed in Figures 5(a-c). As mentioned, the final results are presented in an E-type map. Therefore, in order to compare the SGS results with the kriging method, the ordinary kriging map of data was obtained and displayed in Figure 5(d). Unlike the kriging technique, the SGS technique can simulate the data for each block in whatever number the user needs. This property is the main benefit of the SGS technique in comparison with the kriging interpolation technique because it provides the user with the capability of having wider possibility models of related distribution in an ore body. Therefore, as it can be seen in Figures 5(c) and 5(d), there is a much lower smoothing effect in the spatial distribution of the E-type map in comparison with the kriging map.

Table 2. Variogram model parameters of transformed Cu data in the Hararan area.

<table>
<thead>
<tr>
<th>Variable</th>
<th>C₀</th>
<th>Sill</th>
<th>Range (m)</th>
<th>Maximum Continuity Direction</th>
<th>Anisotropy Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformed data</td>
<td>0.6</td>
<td>1</td>
<td>600</td>
<td>W-E</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Anisotropy Ratio: major axis/medium axis; Direction: major axis orientation for the ellipse of the spherical structure; Range: spherical structure major range (m); Sill: (C+C₀) for spherical structure; C₀: nugget effect.

Figure 4. (a) Sample semi-variogram of transformed Cu data for the directions N-S (a) and E-W (b). The black line corresponds to the theoretical model fitted to the experimental semi-variograms.
Figure 5. Representations of two randomly chosen realizations (a, b), E-type map (c), and kriging map (d) of spatial distribution of Cu element in the Hararan area through applying SGS (the results obtained are on the basis of the normal score transformed).
4.3.3. Validation of simulation results

The simulation results are considered to be acceptable while their validation was done. Validation of the output of a sequential geostatistical simulation (realizations) was carried out in two steps. The first step was applied by comparing the experimental semi-variogram model of the transformed Cu data to the semi-variograms of a set of realizations. However, some differences between the sample model and the variation realizations named fluctuations are reasonable, which may have special causes such as: (a) the parameters of the semi-variogram model, (b) the algorithm that is used for the simulation, and (c) the number of conditioning data to be used for the simulation. The results of this comparison were depicted in Figure 6. (The green line and black lines are representative of the semi-variogram model and semi-variogram realizations, respectively.)

The second step used to examine the validity of the simulation was comparison of the cumulative distribution frequencies (CDFs) of all realizations with the transformed Cu data (CDF). Figure 7 shows the CDFs of some realizations and raw data in various colors. As depicted in Figures 6 and 7, not only there is a suitable coincidence between the semi-variograms obtained from the simulation and the semi-variogram of the raw data but also there is an appropriate coincidence between their CDFs. Hence, in this work, the simulation results are acceptable and can be utilized for the next steps.

Figure 6. Comparison of experimental variogram model reproduction acquired by realizations (black lines) with experimental omni-directional variogram of original data (green line) in the Hararan area.

Figure 7. CDFs of some realizations and primary Cu data in the Hararan area.
4.3.4. Display of anomalous areas
Having produced and validated the ultimate simulated models (50 realizations), the E-type map, which is the average of realizations, was obtained, and selected to display the anomalous areas in the studied area. As mentioned earlier, the threshold value for the element Cu, acquired by the GS method, is equal to 228 ppm in the studied area. Therefore, Figure 8(a) displays the map of the anomalous areas, which exceeds the target value for the Cu element data produced by the E-type map in the Hararan area.

4.4. Determination of reliability of anomalous areas
As mentioned earlier, the SGS algorithm is able to create various realizations, i.e. each grid node in the studied area is estimated as the number of realizations. Therefore, this capability was utilized to define a criterion of reliability of anomalous areas of Cu mineralization. However, there are some degrees of uncertainty for estimation of anomalous regions. In order to evaluate the quality of estimations, a probability map of threshold exceeding was generated using the realizations. Consequently, at each grid node, a bulk of 50 realizations was transformed into a probability of exceeding the target value. Figure 8(b) displays the probability map of exceeding the target value for the Cu element data in the Hararan area. As it can be seen in the probability map, the anomalous areas were discriminated on the basis of various probabilities shown in different colors. The regions that represented high probability values (70-100%) are reliable, and can be used for a detailed exploration and drilling to estimate more accurate reserves. This affair will decrease the risk, time, and cost of a detailed exploration. In Figure 8(b), these districts are depicted in red color, and are located in the southern part of the studied area. Moreover, due to the lack of drilling in the studied area, the copper values obtained from the mineralized points in the outcrop zones were utilized for validation of high probability areas. These points are represented in the probability map (Figure 8b) as black and white asterisks. As it can be seen in this figure, the validation points with high outcrop copper values (white asterisks) are often consistent with high probability points, especially in the southern part. It emphasizes that the outcrops and anomalous areas in the southern part of the district have more importance and priority for a detailed exploration, and could be explored at deeper levels by drilling for a probable deep and high grade mineralization.

Figure 8. (a) Anomalous areas (discriminated by gap statistic method) displayed in red color, produced by simulation in the Hararan area. (b) Probability map of exceeding values in the Hararan area.
5. Conclusions
In exploration geochemistry, recognition of the anomalous areas is necessary and important. In order to recognize them, there are various methods. In this work, a new technique, i.e. a combination of the SGS and GS methods was applied. Compared to various existing methods of identifying the anomalous areas, this method not only separates background from anomaly and represents the anomalous areas but also identifies the reliability of the exceeding critical values. The analysis of 607 lithogeochemical samples for the element Cu was utilized in this work. After transforming the data to a standard Gaussian distribution, 50 realizations were simulated by the SGS algorithm, and an E-type map that was the average of 50 realizations was obtained. As well, based upon the principles of the GS method, the threshold of the transformed data (228 ppm) was acquired. Therefore, regions with values higher than the threshold value were represented in an E-type map, and introduced as the anomalous districts in the studied area. Consequently, a probability map of the anomalous areas (with values greater than 228 ppm) was obtained from 50 realizations, and the anomalous districts that were located in the southern part of the studied area displayed a high probability and were identified as the reliable districts for a detailed exploration.

Eventually, a combination of the SGS and GS methods is suggested as a new powerful technique to identify the anomalous areas and the assessment of their reliability, which is the most important advantage of it. The determination of reliability is important and can be used in decision–making procedures such as drilling projects in a detailed exploration to decrease the risk, cost, and time of exploration.

References


چکیده:
در اکتشافات ژئوشیمیایی تکیه‌گاهی متنوع آماری به همراه روش‌های آماری تک و چند متغیری برای تشخیص مناطق آنومال در دسترس قرار دارد. تکنیک‌های تک متغیر معمولاً برای تخمین مقدار حد آستانه به کار برده می‌شود که به دلیل نتایج محدود حداکثر آنومال در منطقه است. در این پژوهش، تکنیک‌های روش‌های ژئوشیمیایی گویی که از روش‌های ژئوشیمیایی SGS (شکسته‌سر ساختاری) و آماده اشغال قابلیت اعتماد آستانه در منطقه حیران واقع در جنوب شرق ایران به کار گرفته شد که به نظر می‌رسد ناحیه کناری‌ای سازی سیستمی مربوط به یک سیستم سیستم‌گذاری شده. در این، مهم‌ترین مزیت این روش امکان ارزیابی قابلیت اعتماد مناطق آنومال است؛ به عبارت دیگر، مناطق آنومال برحسب مقدار احتمالی شان تفکیک می‌شود. به‌طور مشابه، گویی که مناطق با احتمال بالاتر برای تعیین موقعیت نقاط آنومال در اکتشافات نقصی با قابلیت اعتماد بالاتر به کار برده می‌شود، این روش به تنها در بلوک خطرپذیری، هزینه و زمان عمليات اکتشاف را کاهش می‌دهد. به‌طور دیگر، از ابزارهای ذکر شده به عنوان ابزار قابلیت اعتبار نقاط حفاری و دقت پیش‌بینی آن‌ها به عنوان ابزار قابلیت اعتبار نقاط حفاری در عملیات حفاری می‌شود. در این پژوهش تحقیق، تنها آنالیز ۶۰ نمونه متعلق به موقعیت نقاط حفاری از دستگاهی SGS بر روی داده‌های تبدیل یافته، اجرا و ۵۰ ثابتی مختلف به عنوان مدل به‌طور مثال از روش SGS شناسایی که حد آستانه می‌سازد. از آن‌ها، نتایج نشان داد که حد آستانه می‌سازد و با توجه به این نتایج، تشخیص منطقه متمایز از آستانه به دست آمده و منطقه آنالیسم که پیش‌بینی در بخش جنوبی منطقه مورد مطالعه متمرکز هستند، به عوامل منطقه با قابلیت اعتماد بالاتر در نظر گرفته شده و برای انجام اکتشافات نفیسی و خطرپذیر، پیشنهاد شده.

کلمات کلیدی: ژئوشیمیایی، گویی که، حد آستانه، امور انفعال، قابلیت اعتماد، منطقه حیران.