Using multivariate statistical analyzes in geochemical data: Example of lead-zinc deposit

Ozge Ozer Atakoglu [*1] *and Mustafa Gurhan Yalcin* [2]

Geostatistical methods are frequently used in the geochemical evaluation of mineral deposits. Geostatistical methods are preferred by researchers for reasons such as less error rates than traditional methods and being based on precise results. The most commonly used geostatistical methods are as follows: means, measures of variability, classification, regression and correlation analysis, analysis of variance, correlation function (variogram), estimation error variance, kriging and simulation. The aim of this study is to emphasize the importance of geostatistical methods in the evaluation of geochemical analyzes during the exploration and operation of mineral deposits. It is within the scope of the study to highlight the details of the multivariate data analysis, which is mostly preferred for geochemical analyzes containing large numbers of data, with examples in the literature. The error rates of traditional geometric methods such as polygon, triangle and section used in the calculation of reserve and grade of mineral deposits are higher than the error rates of the values calculated using geostatistical methods. In this respect, it is suggested as a result of the study that researchers who will work on the evaluation of mineral deposits should first turn to geostatistical methods.

2020 MSC: 62H10, 62H05

KEYWORDS: Geostatistical analysis, Geochemistry, Elemental distribution, Prediction, Pb, Zn

Introduction

Geostatistical methods are a type of applied statistics used for the interpretation of geochemical analyzes and estimation of problems [17]. The foundations of the theory were laid by the French mining engineer G. Matheron [17]. Geostatistical interpretation methods are being applied by researchers by expanding them. When geostatistical methods are used for the evaluation of mineral deposits, it is an important method used in many aspects such as determining the extent of mineralization, monitoring and quality extraction. It is possible to gather the main geostatistics under two main roofs: Random variables and data analysis. Data analysis from a geoscience perspective is the most common geostatistical method for the evaluation of mineral deposits. Multivariate data analysis constitutes the first step of the project by providing accurate summary and interpretation of geochemical data [17]. The purpose of factor analysis, which is one of the multivariate statistical analyzes, is to reduce the high number of variables to a small number and to make the data easy to understand [1], [7], [18]. As a result of factor analysis, a common variance value is learned [1], [14]. In this study, the benefits of data analysis, which is one of the geostatistical methods, in the evaluation of mineral deposits are emphasized. Data analysis

covers the determination of the basic statistical properties of the results of geochemical analysis containing a large number of data and presenting them in a summary form. Histograms, one of the graphical statistical methods, are the sample counterpart of the density function of the random variable. However, while they allow graphical representations, they do not display descriptive statistical data numerically. For this purpose, descriptive statistical analyzes are also performed. Descriptive statistics are divided into three groups: Measures of central tendency (Arithmetic mean, Geometric mean, Harmonic mean, Mode, Median); Measures of variability (range of variation, mean absolute deviation, variance, standard deviation, coefficient of variation); Distribution shape measures (skewness measures, Pearson asymmetry measures, Bowley asymmetry measures, Kurtosis and Skewness measures) [6]. In other words, descriptive statistics are divided into three as location, spread and shape statistics. Location statistics describe the center of the distribution. Diffusion statistics express the degree of variability of the data. Shape-measuring statistics measure the tail length of distribution types. The common feature of all descriptive statistics is that they simply present the data numerically. Histogram plots and relational lineage plots divide the data into specific class ranges, and it is possible to subtract the number of data groups in each class range and obtain the frequency value [10], [11], [12], [13]. Geometric traditional methods such as polygon, triangle and section are used in the calculation of mineral deposits reserve. The basis of these methods is based on geostatistics. Considering a mineral deposit, the difference between the grade values will be f(x)=z(a)-z(a+x) when the grade value of the ore at point a is indicated by the grade value z(a+x) at a distance x from here. Calculation and modeling of the function of distance X will be important in reserve calculation methods. However, this relationship was later destroyed by the theory of regional variables [8], [5]. In geostatistics, the grade changes of regional variables with distance are modeled as stationary random functions. A few of the studies dealing with geostatistical methods in the evaluation of mineral deposits are summarized as follows; Geological modeling of the polymetallic mineral deposit containing Pb, Zn and Ag in Balikesir-Balya Hastanesitepe locality was made by simulation and Krigging methods. Simulation method was used for the distribution of host rock and ore at certain elevations of the underground. However, they thought that the simulation was weak in showing the subterranean melting spaces (lens, karstic void, alteration zone, etc.) [3]. The data obtained as a result of the geochemical analyzes on the urban soils of the Brazilian Parana region were evaluated geostatistically. The Matheron Semivariogram method was used on the normally distributed data. The semivariogram was made according to 3 theoretical models: spherical, exponential and Gaussian. The study showed that geostatistical analysis can make good predictions for trace metals [2]. The physical and chemical properties of the soils of the Bindiba mining area in eastern Cameroon were analyzed. Forecast maps at unsampled locations in the region were extracted by variographic analysis and normal Krigging (OK) geostatistics methods. The spatial variability of the metal concentrations of the region was learned with the maps drawn from the geostatistical models [9]. Triangular prism, trapezoid, polygon, isopaque, etc. for reserve calculations of ore deposits planned to be operated and produced. Traditional methods are often preferred. However, it has been observed that the error rates of the predictions made by traditional methods are at high levels. Contrary to these methods, the reasons why geostatistical modeling techniques are more reliable than traditional methods can be summarized as follows: it provides estimation and simulation in unsampled areas, the continuity, continuity and anisotropy of the ore bed are measured in the light of the estimations, suitable location estimates are made for sampling, the grade distribution of the ore in the deposit is made, Error rates of predictions and simulations are calculated, and their reliability and validity

are given with precise results. In the literature, there are studies on the interpretation of geochemical analyzes made during the exploration and operation of mineral deposits using different geostatistical methods. However, no research has been found that emphasizes the importance of geostatistical methods in the evaluation of mineral deposits and addresses all methods. In this context, the aim of this research is to touch on all geostatistical methods that can be used in the evaluation of mineral deposits and to emphasize their importance.

Main results

The results of the geochemical analysis of the lead-zinc deposit in the Silesian-Cracow region in Southern Poland were interpreted by statistical analysis. Some descriptive statistical data (minimum, maximum, standard deviation, variance, coefficient of variation) of ore samples and wall rock (dolomite) samples collected from the study area were compared with the anomaly values known in the literature such as earth crust, continental crust, carbonate rocks. The cumulative frequency distribution curve was drawn on the logarithmic probability paper of the zinc concentrations of the samples collected from different locations. Here, it was thought that the zinc was close to the lognormal. Cumulative frequency distribution curves were also drawn for zinc and lead. It has been discussed whether the samples collected from different locations of the region are within the operable limits. The histogram of zinc in the content of the dolomite samples, which is the country rock of the region and found with the ore, was drawn. In the study, statistical analyzes of lead-zinc content in dolomite made it easier to distinguish the threshold values of ore bodies [15]. The geochemical analysis data of the Mehdiabad zinc-lead deposit in Iran has been interpreted geostatistically. Correlation and regression analysis were applied to the data that were adjusted to normal distribution. Elements were categorized by principal component matrix (PCA) analysis method. The distribution of the selected elements is visualized by the 3D modeling method. In this study, Pearson correlation analysis and eigenvalues/eigenvectors covariance analysis were applied. As a result of the study, the geostatistics results of the samples collected from different locations divided the elements into three main categories. 1. For the study, Si, Al, Na, Ni, P, Ti, Ce, Cr, La, Li, V and Zr elements are abundantly found in the deep parts of the ore deposit in the structure of quartz, ilmenite, zircon, monazite and clay group minerals. 2. Elements forming the side rock of the ore, Ca and Mg, were chemically deposited in the limestone and parent rocks. 3. They are hydrothermal fluids that bring the ore-forming elements Pb, S, Zn, Ag and Cd. With the results of the factor analysis, it was seen that Fe, Ba and Mn had a different origin. It is thought to have a different origin from the hydrothermal fluids that make up the ore [1].

The geochemical data of lead-zinc in the metallogenic belt in Kerman region of Iran are statistically explained. Analysis data has been normalized. Variograms of Pb and Zn concentrations were extracted for each exploration well. In addition, the behavior of Pb and Zn was tried to be measured precisely with the K-means clustering method. As a result of the K-means clustering analysis, it was determined that Pb and Zn exhibited geochemically similar linear behavior [4]. The Zn geochemical analysis values of 50 ore samples collected from the lead-zinc deposit of the Bolkardagi (Ulukisla-Nigde) region in Turkiye are given in Table 1 [16]. The descriptive statistical values of the samples are given in Table 2. The histogram graph of the Zn distribution is given in Figure 1-a and the box-plot graph is given in Figure 1-b.

	%Zn		%Zn		%Zn		%Zn		%Zn
1	4.87	**11**	3.25	**21**	5.23	**31**	13.04	**41**	2.87
2	5	**12**	5.74	**22**	5.18	**32**	12.03	**42**	2.99
3	4.87	**13**	5.27	**23**	5.22	**33**	12.24	**43**	2.96
4	5.10	**14**	5.65	**24**	5.00	**34**	11.82	**44**	2.94
5	5.93	**15**	5.61	**25**	5.23	**35**	11.61	**45**	2.82
6	5.43	**16**	9.02	**26**	5.85	**36**	16.06	**46**	2.79
7	6.21	**17**	9.09	**27**	6.12	**37**	16.73	**47**	2.54
8	4.97	**18**	9.13	**28**	6.28	**38**	16.49	**48**	2.07
9	5.18	**19**	8.64	**29**	5.79	**39**	15.69	**49**	2.68
10	4.96	**20**	8.03	**30**	6.14	**40**	16.03	**50**	2.79

Table 1: Zn concentrations of Turkiye Bolkardagi Zn-Pb deposit [16]

Statistics	**Results**
Mean	6.94
Median	5.52
Mode	2.79[a]
Std. Deviation	4.12
Variance	16.98
Range	14.66
Skewness	1.16
Kurtosis	0.36

a. Multiple modes exist. The smallest value is shown.

Table 2: Descriptive statistical values of Zn concentrations of Turkiye Bolkardagi Zn-Pb deposit

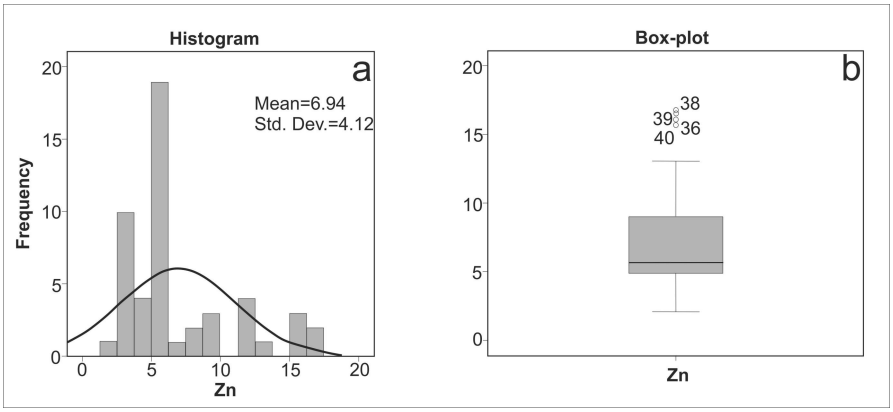


Figure 1: a) histogram plot and b) box-plot plot of Zn distribution

Conclusion

Summary values of 50 Zn samples collected from Bolkardagi zinc lead deposit were presented. The central tendency measures of the descriptive statistics were found as follows, respectively: Arithmetic mean: 6.94; Median (median data when series is ordered from largest to smallest): 5.52; mod (most observed value from series): 2.79.

The results of the scatter shape measurements were found as follows, respectively: Standard deviation (the square root of the mean squared difference of the series from the arithmetic mean): 4.12; variance (square of standard deviation): 4.12; Range (The difference of the maximum value in the series from the minimum value): 14.66. The results of the distribution shape measurements are as follows, respectively: Skewness (can take values in the range of $-\infty$ $+\infty$, if it takes values in the range of ± 3, it is assumed that the distribution is normal): 1.16. Kurtosis (related to the steepness or kurtosis of the scatter plot, value $(+)$ means the curve is steep, $(-)$ means the curve is flat): 0.36. Since the skewness value fell within the range of ± 3, the distribution was considered to be normal. However, the plotted histogram graph showed that the distribution was not normally distributed. According to the histogram graph, the distribution is right skewed and tends to low values. In histogram plots tending to low values, it is thought that there is a relationship between measures of central tendency of the data as follows: Arithmetic mean $>$ Median $>$ Mode. The situation is the same for Bolkardagi Zn samples. Arithmetic mean (6.94) $>$ Median (5.52) $>$ Mode (2.79). The box-plot plot enabled the detection of outliers/outliers in the series. It also made it easier to understand in which range the data were mostly classified. Within the scope of the study, a few simple geostatistical analyzes were tried to be exemplified with the results of the analysis compiled from a previous study from the literature. With geostatistical analysis, it is aimed to summarize the data and make it easier to understand. The importance of geostatistical analysis in the evaluation of mineral deposits has been revealed.

Acknowledgments

This paper is dedicated to Professor Yilmaz SIMSEK on the occasion of his 60th anniversary.

References

[1] Z. Bonyadi, *3D modeling and determination of factors responsible for zinc-lead mineralization in the mehdiabad deposit, central Iran, based on statistical analysis of geochemical data*, Acta Geologica Sinica-English Edition **96 (6)**, 2040–2055, 2022.

[2] W. H. Dib, J. D. F. G. M. Cicarello, L. Kummer and M. L. C. Del Monego, *Geostatistical assessment of trace metals in urban soils, Paraná, Brazil*, Revista Ibero-Americana de Ciências Ambientais **11 (6)**, 358–370, 2020.

[3] E. Ekici and A. Ersoy, *Jeoistatistik yontemlerle polimetalik maden yataklarinin jeolojik modellenmesi*, 2012.

[4] S. S. Ghannadpour and A. Hezarkhani, *Lead and zinc geochemical behavior based on geological characteristics in Parkam Porphyry Copper system, Kerman, Iran*, Journal of Central South University **22**, 4274–4290, 2015.

[5] A. G. Journel and J. C. Huijbregts, *Mining geostatistics*, Academic Press, New York, N.Y., 1978.

[6] S. Kalayci. *SPSS uygulamali cok degiskenli istatistik teknikleri*, Asil Yayin Dagitim, Ankara, Turkey, pp. 321–331, 2010.

[7] P. Kline, *An easy guide to factor analysis*, Routledge, 2014.

[8] G. Matheron, *Principles of geostatistics*, Economic geology **58 (8)**, 1246–1266, 1963.

[9] M. M. Njayou, M. Ngounouno Ayiwouo, L. L. Ngueyep Mambou and I. Ngounouno, *Using geostatistical modeling methods to assess concentration and spatial variability of trace metals in soils of the abandoned gold mining district of Bindiba (East Cameroon)*, Modeling Earth Systems and Environment **9 (1)**, 1401–1415, 2023.

[10] O. Ozer Atakoglu and M. G. Yalcin, *Geostatistical analysis and spatial distribution map of geochemical contents of the Susuzdag Formation limestones*, Pamukkale Universitesi Muhendislik Bilimleri Dergisi **28 (6)**, 851–862, 2022.

[11] O. Ozer Atakoglu and F. Yalcin, *Evaluation of the surface water and sediment quality in the Duger basin (Burdur, Turkey) using multivariate statistical analyses and identification of heavy metals*, Environmental Monitoring and Assessment **194 (7)**, 2022.

[12] O. Ozer Atakoglu and M. G. Yalcin, *Explanation of the Sutlegen bauxites to some REE contents by statistical approach and inequality expressions*, Gazi University Journal of Science Part A: Engineering and Innovation **8 (4)**, 391–401, 2021.

[13] O. Ozer Atakoglu, M. G. Yalcin, Y. Leventeli and B. T. San, *Geochemistry of red soils in the Kas district of Antalya (Turkiye) using multivariate statistical approaches and GIS*, Minerals **13 (6)**, 2023.

[14] K. Popov, *3D modelling of the geochemical associations in the Assarel porphyry copper deposit (Bulgaria)*, Comptes 2016.

[15] T. J. Smakowski, *Geological prognosis of potential zinc and lead resources for Triassic formations in the Silesian-Cracow region in southern poland*, Mathematical geology **24**, 693–703, 1992.

[16] S. Temur, *Bolkardagi yoresi Zn-Pb yataklarinin jeokimyasal incelemesi*, TJK Bulteni **35 (2)**, 101, 1992.

[17] A. E. Tercan and C. Sarac, *Maden yataklarinin degerlendirilmesinde jeoistatistiksel yontemler*, TMMOB Jeoloji Muhendisleri Odasi Yayinlari **48**, 1998.

[18] N. Zandy Ilghani, F. Ghadimi and M. Ghomi, *Application of alteration index and zoning for Pb-Zn exploration in Haft-Savaran area, Khomein, Iran*, Journal of Mining and Environment **9 (1)**, 229–242, 2018.

Department of Geological Engineering, Faculty of Engineering, Akdeniz University, Antalya, Turkiye [*1]

Department of Geological Engineering, Faculty of Engineering, Akdeniz University, Antalya, Turkiye [2]

E-mail(s): oozer@akdeniz.edu.tr [*1] (corresponding author), gurhanyalcin@akdeniz.edu.tr [2]