

Using Sequential Gaussian Simulation to Assess Geochemical Anomaly Areas of Lead Element

Fengrui Chen¹, Shiqiang Chen² and Guangxiong Peng^{3,4,*}

¹ Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, Ministry of Education, Henan University, Kaifeng, China*

²The Center for Yellow River Civilization and Sustainable Development of Henan University, Kaifeng, China.

³School of Geosciences and Info-Physics, Key Laboratory of Metallogenic Prediction of Nonferrous Metals, Ministry of Education, Central South University, Chansha, China.

⁴Qinghai Institute of Geological Survey, Xining, China;

frulich@gmail.com pgxcusu@csu.edu.cn

Abstract. Mapping geochemical anomaly is essential for prospecting. Kriging is often used to characterize the spatial variability of geochemistry. However, due to its smooth effects, it is incapable of detecting multi-location uncertainty of geochemistry estimate. Sequential gaussian simulation (SGS) can model the spatial uncertainty through generation of several equally probable stochastic realizations. In this study, SGS is used to map Pb spatial distribution and to provide a quantitative measure of its spatial uncertainty in particular. The results show that the spatial pattern of Pb is well recognized using the SGS. The different among realizations is used to provide a quantitative measure of spatial uncertainty of Pb estimate. Knowledge of spatial uncertainty is helpful to evaluate abnormal areas of Pb element.

Keywords: geochemical anomaly, geostatistics, sequential Gaussian simulation

1. Introduction

With the development of exploration geochemistry, the evaluation of regional geochemical anomaly such as regional stream and secondary halo attract more and

*Corresponding author

more attention[1]. How to evaluate effectively the geochemical anomaly has been the focus for many exploration geochemists[2]. For geochemical elements carry some mineralization information. The fine spatial analysis of elemental geochemistry can reveals the tracing of mineralization directly or indirectly. The study of geochemical element distribution is one of the important ways to reveal the spatial variation, mineralization and enrichment of element[3, 4]. Geochemical anomaly analysis is the important work for prospecting, whose core content is to find, filtrate and evaluate anomaly.

For better understanding of geochemical anomaly, it is very important to predict the spatial distribution of geochemistry over the study area. Geostatistics has been used for spatial variability characterization and prediction of geochemistry. The most well-known geostatistics estimation technique is ordinary kriging (OK) called also the “best linear unbiased estimator” [5, 6]. However, the smoothing effect of OK results in less variation in the estimated values than in the observed values [7-9]. The estimation at each unsampled location includes an estimation error or uncertainty. This uncertainty should be quantified and accounted for in subsequent decision-making. For geochemical anomaly analysis, assessing the reliability of delineating an area as “normal” or “anomaly” requires the knowledge of estimation uncertainty at several locations simultaneously. Geostatistics simulation has the ability to not only predict the spatial distribution of the property at hand but also to assess both local and spatial uncertainty. The main objective of this study is to assess the uncertainty of delineating geochemical anomaly. For this the feasibility of using sequential Gaussian simulation (SGS) [10-12]is examined.

2. Materials and method

2.1 Study area and data source

The study area is located in the Mohailaheng Pb-Zn deposit of Zaduo County, Qinghai Province, China. It is located in $95^{\circ} 33' - 95^{\circ} 59' E$ and $32^{\circ} 39' - 32^{\circ} 58' N$, near the North of Sanjiang metallogenic belt (Fig. 1). Numerous of Pb-Zn-Ag-Cu ore deposits with significant economic values occur in Lanping basin and

Yushu and Tuotuohe areas on the eastern and northern margins of Tibetan orogenic belt. Being hosted by sediments and having no obvious affinity to igneous activities, these deposits occur in continent-continent (Indo-Asian) collisional orogen and are controlled by thrust-nappe structures, thus obviously different from the well known sediments- hosted base metal deposits in the world[13]. They were formed in Tertiary foreland basins developed from Paleo-Tethyan rifting basins and Mesozoic depression basins. During mineralization, thrust-nappe structures juxtaposed Mesozoic strata over Tertiary strata through a series of parallel reverse/ thrust faults in the basins. The deposits are strictly controlled by thrust-nappe structures. The rock types of magmatic are complete with frequent magmatic activity, while the exposed area is small with scattered distribution. Sampling points get from geochemical data of stream sedimentation in Yushu region of South Qinghai in 2009. This data set includes 864 Sampling points with 13 elements of Ag, As, Au, Ba, Bi, Cd, Cu, Hg, Mn, Mo, Pb, Sb, Sn, W and Zn.

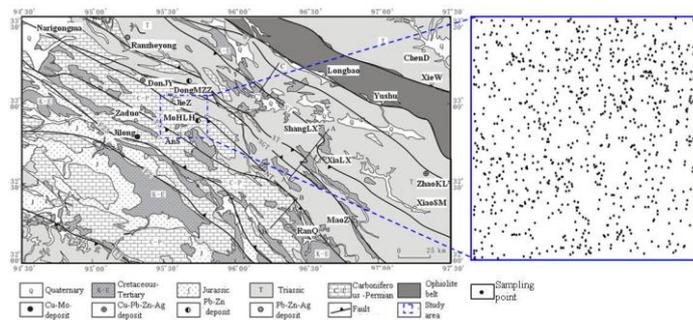


Fig. 1. Geological map of south Qinghai and the spatial distribution of sampling points.

2.2 Kriging

Kriging is a linear interpolation procedure that provides a best linear unbiased estimation for quantities, which varies in space, Kriging estimates are calculated as weighted sums of the adjacent sampled concentrations[14, 15]. That is, if data appear to be highly continuous in space, the points closer to those estimated receive higher weights than those farther away.

In geostatistical analysis, the semivariogram is used to quantify the spatial dependence of the Pb concentration:

$$r(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (1)$$

where h is the distance between both locations, x_i and x_i+h , and $N(h)$ is the number of pairs for x_i and x_i+h . The experimental semivariogram was then fitted with a suitable theoretical model: spherical, exponential or gaussian. The models provide information about the spatial structure as well as the input parameters for kriging interpolation. Kriging is considered as the optimal spatial interpolation for making best linear unbiased estimation of regionalized variables at unknown locations. The spatial prediction of the value of Pb at an unknown point is calculated as a weighted average:

$$z^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (2)$$

where $z^*(x_0)$ is the value to be estimated at the location x_0 ; $z(x_i)$ is the known value at the sampling site x_i and λ_i is the weight. There are n sites within the search neighborhood around x_0 used for estimation, and the magnitude of n will depend on the size of the moving search window and on user's definition. Kriging differs from other methods, such as inverse distance weighted, in that the weight function λ_i is no longer arbitrary, and is calculated from the parameters of the fitted variogram model with unbiasedness and minimized estimation variance for interpolation.

2.3 Sequential Gaussian simulation

The basic idea of sequential simulation is that the conditional distribution of the observed variable can be used for the simulation of subsequent grid points. The SGS is based on multi-Gaussian assumption of a random function model[16]. The data requires then a prior normal score transformation to ensure the normality of at least the univariate distribution of data. The normal score transformation[5] for a continuous variable z at location $x_a, a=1, \dots, n$, is given as:

$$y(x_a) = G^{-1}[F^*(z(x_a))] \quad a = 1, \dots, n \quad (3)$$

where $G^{-1}(\cdot)$ is the inverse Gaussian cumulative distribution function (CDF) of the random function $Y(x)$, and F^* is the sample CDF of z . A back transformation of the normal scores to the original space is achieved by applying the inverse of the normal score transform introduced in equation (3).

The SGS consists in defining a regularly spaced grid, covering the region of interest and establishing a random path through all grid nodes, such that each nodes is visited only once in each sequence. For every grid node the normal score conditional cumulative distribution function (CCDF) is estimated according to a simple kriging system conditioned to all Pb known data and previously simulated grid nodes within a neighborhood of the location to be simulated. In this stage the semivariogram model of Pb normal scores is used. Afterwards a simulated value is drawn from the estimated CCDF and then added to the conditioning data set to be used for simulating other grid nodes. This procedure will be continued until all the grid nodes are simulated. The simulated normal score CCDFs then need to be back transformed to obtain the CCDFs of Pb in original space. These sequential steps build up only the first realization, which is only one of equally probable models of Pb spatial distribution. To generate multiple realizations, this sequence should be repeated with different random paths passing through all nodes for each realization. Indeed the only difference among the several individual simulations is the random-number seed, which initiates the simulation process and affects the random drawing of a value from the CCDF used to generate each simulated value. Therefore, in different realizations different values will be assigned to the same location. Pixel-by-pixel averaging of generated maps of Pb spatial distribution provides the so-called E-type[17] estimate at each grid node. The corresponding conditional variance, which represents the spread of CCDF and the probability of exceeding or not exceeding a critical threshold of Pb at each grid node also can be produced through post processing.

3. Results and discussions

3.1 descriptive statistics and spatial autocorrelation analysis

Descriptive statistics and the histogram of Pb observations are presented in Table 1 and Fig. 2, respectively. The study area has a large range of Pb element, and the maximum and minimum Pb elements are 6.2 and 944.2 g/t. The skewness of Pb observations is 14.15, which indicates that the data is anormal distribution. Therefore, before the implementation of kriging, the Pb data should be transformed. The skewness of $\ln(\text{Pb})$ is 0.8, and the $\ln(\text{Pb})$ data are almost normal distribution. In Fig. 2(b), the Pb element data have been transformed by a normal score transformation.

The statistics (i.e. the mean value of 0 and the variance of 1) of transformed data check the correctness of the transformation.

Table 1. Summary statistics of Pb data and its transformation

	<i>mean</i>	<i>STD</i>	<i>variance</i>	<i>maximum</i>	<i>minimum</i>	skewness
Pb	31.9	46.3	2144.4	6.2	944.2	14.15
Normal score of Pb	0	1	1	3.24	-3.24	0
ln(Pb)	1.4	0.224	0.05	0.79	2.98	0.68

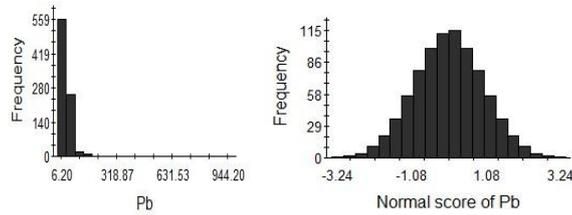


Fig. 2. Histogram of the raw and normal score transformed data for Pb

The experimental semivariogram of ln() and normal transformed data are presented in Table 2 and Fig. 3. Also plotted on each semivariogram there is the best-fitted model to the experimental points. As it is shown in Fig.3, the isotropic semivariogram of normal score of Pb follows an exponential structure with a nugget effect of 0.29, and it reaches to a sill of 0.70 at a range of 4125 m. Compared with the semivariogram model of normal score of Pb, the semivariogram of ln(Pb) follows again an exponential structure, and have a more or less continuity at space (3084). However, its partial sill/sill (0.74) is larger than that of former (0.63), which indicates that ln(Pb) has more spatial correlation.

Table 2. The semivariogram analysis of ln() and normal score transformed data for Pb

	<i>model</i>	<i>nugget</i>	<i>range</i>	<i>sill</i>	partial sill/sill
Normal score of Pb	Exponential	0.29	412	0.7	0.63
	1		5	9	
ln(Pb)	Exponential	0.01	308	0.039	0.74

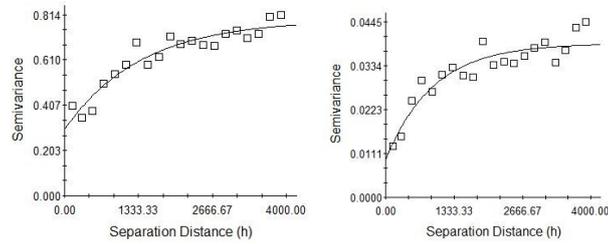


Fig. 3. Experimental semivariogram and best fitted model of the $\ln()$ and normal score transformed data for Pb.

3.2 Simulation result

Three hundred realizations with equal probability of Pb spatial distribution are generated using SGS on a 30-m grid over the study area. Simulation is performed using the simple kriging estimator, and the semivariogram model of Pb normal scores. The four randomly select realizations are display in Fig.4 (a)-(d). Each realization represents a realistic spatial distribution of Pb without a smoothing effect. Fig. 4(e) and (f) is the E-type of simulation and OK. They all indicated the spatial distribution of Pb element; however, there are a lot differences among them. It is evident that simulation realizations present more detail information than OK map. The maps of E-type and OK are smoother.

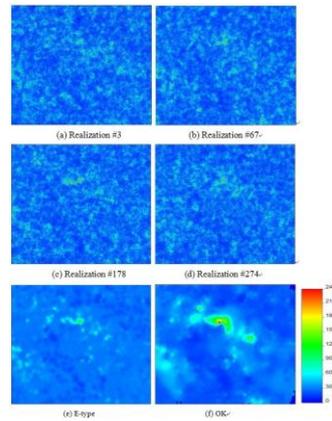


Fig. 4. The spatial distribution of Pb by sequential Gaussian simulation and OK over the study area

Table 3 represents summary statistics on Pb output result by both SGS and OK. SGS individual realizations well preserve the range of variation of the measured Pb

data compared to OK map where the minimum value is overestimated while the maximum value is underestimated. This reveals the smoothing effect, a typical property of kriging. This is also evident from the kriging variance, which is much smaller than the actual variance. OK dose not reflect the true variability, and hence is not appropriate for the underlying outlook to this study. Summary statistics of E-type are similar to those for the kriging map. This is an expected result in the sense that it is well-know that when the number of realization is large enough, the mean of the realizations converges to the estimation result.

Table 3. Basic statistics of Pb output results

	<i>mean</i>	<i>minimum</i>	<i>maximum</i>	variance
Actual	31.9	6.2	944.2	2144.4
Realization#3	32.1	6.2	944.2	2205.1
Realization#67	31.6	6.2	944.2	2123.7
Realization#178	31.5	6.2	944.2	2189.4
Realization#274	32.3	6.2	944.2	2167.6
E-type	31.5	9.1	683.7	1653.4
OK	32.2	9.1	723.4	1878.3

As mentioned before, the critical value is 124.5 g/t, which is determined by mean plus two times STD for Pb data. The OK estimation and Probability map of exceeding critical value is in Fig. 5. The OK map is the unique value, therefore, it critical area for OK map is certainty. However, there must be some degree of uncertainty for estimation. For example, an estimated value in OK map is larger than critical value; if it has a smeller probability of occur (0.001), then this value can be assigned to critical area. SGS can solve this problem, it provide the uncertainty of estimated value by all simulation realizations. From Fig. 5(b), the critical area is not certainty, and it is the represent of probability. The larger the value, the higher it belong to critical area. The critical area can be determined by setting a certain threshold which can take into account the situation of economic, social and management.

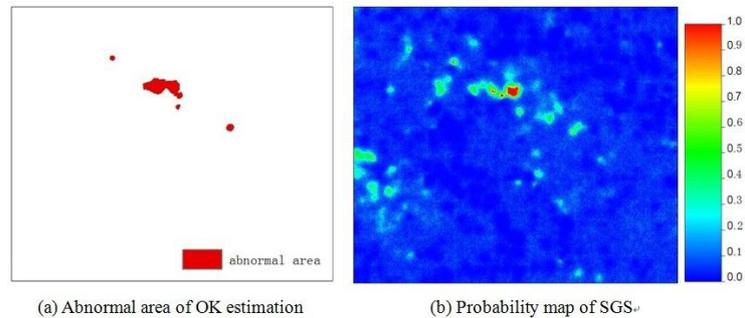


Fig. 5. OK estimation and probability map of exceeding 124.5 g/t for Pb.

4. Conclusion

Mapping geochemical anomaly is essential for prospecting. Traditional methods often make use of the interpolation results to attain anomalistic areas, without considering the uncertainty of estimate. In this study, SGS is used to evaluate the anomalistic areas of Pb element. The simulated maps are summarized to provide the probability map of Pb exceeding the critical value. The probability map is used to delineate the anomalistic areas. The difference among 300 realizations provides a quantitative measure of spatial uncertainty of Pb estimate. Eventually SGS is recommended as a powerful and convenient simulation-based algorithm for not only spatial prediction but also for evaluating both local and spatial uncertainty, which should be incorporated in many decision-making processes.

Reference:

1. W. Ruiting.: Facts and problems of the evaluation of regional geochemical anomalies. *Chinese Geology*.23, 168-175(2005).
2. W. Xueqiu.: Exploration geochemistry:past achievements and future challenges. *Earth Science Frontiers*. 10, 239-248(2003).
3. X. Shuyun, B. Zhengyu.: The method o f moments and its application to the study of mineralization in Shao guan district, north Guangdong , China. *Journal of Jilin University : Earth Science Edition*. 33, 443- 448(2003).
4. C. Qiuming.: Multifractal theory and geochemical element distribution pattern. *EarthScience. Journal of China University of Geo sciences*. 25, 311-318(2000).

5. P. Goovaerts.: *Geostatistics for natural resources evaluation*. Oxford University Press, New York (1997).
6. E. Isaaks, R. Srivastava.: *An introduction to applied geostatistics*. Oxford University Press, New York (1989).
7. E. Pardo-Iguzquiza, P. Atkinson.: Modelling the semivariograms and cross-semivariograms required in downscaling cokriging by numerical convolution and deconvolution. *Computers and Geosciences*. 33, 1273-1284(2007).
8. P. Kyriakidis.: A Geostatistical Framework for Area-to-Point Spatial Interpolation. *Geographical Analysis*. 36, 259-290(2004).
9. P. Goovaerts.: Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical geology*. 40, 101-128(2008).
- 10.A. Castrignan, G. Buttafuoco.: Geostatistical stochastic simulation of soil water content in a forested area of south Italy. *Biosystems engineering*. 87, 257-266 (2004).
- 11.R. Dimitrakopoulos, X. Luo.: Generalized sequential Gaussian simulation on group size v and screen-effect approximations for large field simulations. *Mathematical geology*. 36, 567-591(2004).
- 12.M. Delbari.: Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *Catena*. 79, 163-169(2009).
- 13.H. Zengqian.: Thrust-controlled, sediments-hosted Pb-Zn-Ag-Cu deposits in eastern and northern margins of Tibetan orogenic belt. Geological features and tectonic model. *Mineral Deposits*. 27, 123-145 (2008).
- 14.R. Webster, M. Oliver.: *Geostatistics for environmental scientists*. John Wiley & Sons Inc, New York(2007).
- 15.H. Wackernagel.: *Multivariate geostatistics: an introduction with applications*. Springer Verlag, Berlin(2003).
- 16.P. Goovaerts.: Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*. 89,1-45(1999).
- 17.C. Deutsch, A. Journel.: *GSLIB: Geostatistical software library and user's guide*. Oxford University Press, New York(1998).