IJMGE
Int. J. Min. & Geo-Eng.
Vol.50, No.1, June 2016, pp.49-60.

Accuracy evaluation of different statistical and geostatistical censored data imputation approaches (Case study: Sari Gunay gold deposit)

Babak Ghane and Omid Asghari*

Simulation and Data Processing Laboratory, University College of Engineering, School of Mining Engineering, University of Tehran, Tehran, Iran

Received 26 Nov. 2015; Received in revised form 2 Jan. 2016; Accepted 5 Jan. 2016
**Corresponding author Email: o.asghari@ut.ac.ir*

Abstract

Most of the geochemical datasets include missing data with different portions and this may cause a significant problem in geostatistical modeling or multivariate analysis of the data. Therefore, it is common to impute the missing data in most of geochemical studies. In this study, three approaches called half detection (HD), multiple imputation (MI), and the cosimulation based on Markov model 2 (MM2) are used to impute the censored data. According to the fact that the new datasets have to satisfy the original data underlying structure, the Multidimensional Scaling (MDS) approach has been used to explore the validity of different imputation methods. Log-ratio transformation (alr transformation) was performed to open the closed compositional data prior to applying the MDS method. Experiments showed that, based on the MDS approach, the MI and the MM2 could not satisfy the original underlying structure of the dataset as well as the HD approach. This is because these two mentioned approaches have produced values higher than the detection limit of the variables.

Keywords: *censored data, collocated cosimulation Markov model 2, half detection, imputation, multidimensional scaling, multiple imputation.*

1. Introduction

A common problem in analyzing the mining datasets is the observations with values below the detection limit of the instruments (censored data). When the values approach zero and the precision of the laboratory instrument is not sufficient to detect the right values, then, the dataset will contain the missing values which are below the detection limit [1]. The level at which a measurement has a 95% probability of being different than zero is defined as the limit of detection [2]. Some researchers may believe that the censored data are unimportant since their values are extraordinary small. However, these data may influence the parameters of the distribution of the whole samples. Also, incorrect treatment of the censored data may produce biased mean and variance of the distribution [3]. Thus, the researchers have to

find an appropriate imputation approach for dealing with these data.

One of the most common and easiest approaches is called simple replacement, in which the censored values are replaced with a single value equal to zero, proportion of the detection limit (usually 1/2 or $1/\sqrt{2}$ times the detection limit), or the exact value of the detection limit [1] .In cases where missing data are less than 1 percent of the whole population, simple replacement by a single value of 0.33 to 0.5 times the detection limit is appropriate [4]. For instance, Carranza [5] replaced Arsenic missing data which contained 30% of the whole samples by a value equal to the half detection limit for As prior to mapping the geochemical anomalies.

Besides, there are other types of the imputation approaches which are based on statistical algorithms. Among them, multiple imputation (MI) and maximum likelihood estimation (MLE) approaches are preferred more by methodologists. MI [6] is a statistical technique for analyzing datasets containing missing data. The purpose of this approach is to replace each missing value by several plausible values and produce *n* complete dataset. Thus, there is uncertainty about the right value to impute [7, 8]. As a result, the n complete datasets are integrated to produce a final dataset. MLE [9], as another popular approach, maximizes the log-likelihood of each observation by estimating the unknown population parameters through iterative optimization [10].

In addition to simple and statistical based imputation approaches, geostatistical approaches can be used to impute the missing data. These approaches are applicable when analyzing the regional variables like geochemical ones. Geostatistical approaches estimate unknown data in an unbiased way and minimize the estimation variance [11]. As multiple imputation may not be suitable for geological data, geostatistical approaches can be merged with the MI, producing the parametric and non-parametric methods [10]. In another attempt to fill in missing data, Munoz et al [12] applied multiple imputation by means of geostatistical models to environmental data. They claimed that combining the multiple imputation method with the geostatistical models has the advantage of making it possible to impute missing data for both continuous and discrete environmental variables.

For another instance, Zhang et al. [13] applied two geostatistical methods named ordinary kriging (OK) and ordinary cokriging (OCK) to fill in missing data of remotely sensed atmospheric methane. They found that the two interpolation methods presented similar spatial patterns and provided acceptable results, while OCK method yielded better results than OK. However, these approaches cannot be used for imputing the censored data because they may estimate or simulate values larger than the limit of detection. To explore the ability of the geostatistical approaches for imputing the censored data, Collocated Cosimulation Markov model 2 is applied in this study.

According to the fact that the new datasets have to satisfy the original underlying structure of the data, multidimensional scaling (MDS) approach is used to explore this issue. MDS [14] is a multivariate technique that reduces the dimensions of multivariate datasets and also aims to reveal the underlying structure of the data. This method provides a geometrical configuration of the relations between the variables. In this configuration, variables which are more similar to each other have fewer distances [15]. MDS has been used in a small number of mining or geological studies during the past years. Deutsch and Deutsch [16] used MDS to plot the rock transition probability matrix. In another study, Boisvert and Deutsch [17] applied MDS to make sure of the positive definiteness of the resulting kriging system of equations for an attempt to incorporating locally varying anisotropy in kriging or sequential Gaussian simulation.

But, applying the MDS is problematic due to the compositional nature of the geochemical data. The restriction of constant sum may yield spurious correlations between the variables because, by increasing the portion of one variable, other variables should be decreasing. Therefore, correlations are not free to vary from -1 to +1 [18]. So, those multivariate statistical methods that are based on correlation coefficients between the variables are not appropriate for untransformed compositional data. As a result, closed number system compositional data should be open prior to

further analysis [19- 24]. To open the compositional data, three types of log-ratio transformations have been introduced: additive log-ratio (**alr**) transformation [19], centered log-ratio (**clr**) transformation [19], and isometric log-ratio (**ilr**) transformation [25], respectively. Log-ratio transformation methods have been widely used during the past years in geochemical studies. For example, Carrenza [5] applied three kinds of log-ratio transformations to the stream sediment data for mapping and analyzing the geochemical anomalies. Also, Caritat and Grunsky [26] performed both alr and clr transformations to soil geochemical data prior to applying multivariate analysis to identify the pathfinder elements associated with mineral deposits.

The methodology used in this study are presented in section 2. The study area and geochemical datasets are discussed briefly in section 3 where also practical examples are given individually for all imputation approaches. Finally, Sections 4 and 5 present the discussion and conclusion, respectively.

2. Methods

In this paper, three different imputation approaches are applied: 1. simple imputation, 2. multiple imputation approach using Markov Chain Monte Carlo (MCMC) method, and 3. Collocated Cosimulation Markov model 2.

2.1. Simple Imputation

In simple imputation approaches, missing data are imputed by a single value. There are several single imputation approaches, such as Hot Deck imputation, Mean imputation, and Naïve-Bayes imputation. However, in this study, half detection approach (HD) is applied, which imputes for the whole missing data a single value equal to half detection of the elements.

2.2. Multiple Imputation (MI)

Multiple imputation replaces each missing datum with more than two possible values, and, as a result, creates n complete datasets. It also provides the uncertainty of the imputed value. In addition, each complete dataset is analyzed by using standard procedures and finally the analyses are combined in order to get the final result [7].

There are a few methods available in the MI approach, the choice of which depends on the type of the missing data pattern [28]. These are:

1. Monotone missing data pattern, for which both parametric regression method assuming multivariate normality and nonparametric method using propensity score is applicable.
2. Arbitrary missing data pattern, in which case Markov Chain Monte Carlo (MCMC) method, assuming multivariate normality, is applicable.

The monotone missing pattern exists when a variable x_i and all subsequent variables x_k are missing at location j. Otherwise, the missing data have an arbitrary pattern in which case MCMC is proposed to apply [28].

MCMC is a set of methods which simulates direct draws from the distribution in question. In this method, each value is used to randomly generate the next value so that the Markov chain will be created. In this case, the distribution of each sample relies on the value of the previous one [29]. The goal in using MCMC is to achieve a stationary common distribution. It is done by performing the Markov chain long enough to the distribution of samples. MCMC simulates draws from the stationary distribution by repeatedly simulating steps of the chain [30]. MCMC is applied to identify posterior distributions which are the information about the unknown parameters in Bayesian inference. In MCMC method, joint posterior distribution of the unknown parameters can be simulated to achieve the estimated posterior parameters gained by simulation.

2.3. Collocated Cosimulation Markov model 2

Cokriging, as a method to estimate the primary variable using a secondary variable which is more dispersed in the area, is widely applied in geostatistical modeling. Implementation of the cokriging requires calculating a Linear Model of Coregionalization (LMC) variograms which can be tedious and take a long time. Almeida [31] and Almeida and Journal [32] introduced the Markov model in which the estimations are performed without modeling LMC. This method requires only modeling the univariate variogram and calculating the co-located correlation

coefficient between the primary and the secondary variables [33].

In Markov model, only the collocated hard data have influence on estimating the unknown points. In the primary Markov model (MM1), the secondary variable correlogram is not required; however, MM1 requires the sample cross-correlogram $\hat{\rho}_{12}(h)$ shape and continuity which conform to sample primary correlogram $\hat{\rho}_1(h)$. However, the sample cross correlogram tends to obey the secondary correlogram $\hat{\rho}_2(h)$ in practice [34]. Thus, an alternative Markov model was proposed which was named Markov model 2 (MM2), and which requires a model for secondary correlogram $\rho_2(h)$ and a cross-correlogram $\rho_{12}(h)$. A cross-correlogram is made proportional to $\rho_2(h)$ and under the MM2 the following can be written [34]:

$$\rho_{12}(h) = \rho_{12}(0)\,\rho_2(h) \tag{1}$$

where $\rho_{12}(0)$ is the cross-correlation between the primary and the secondary variables. Therefore, in order to estimate a primary variable in each node of the more abundant secondary variable, collocated models could be implemented. As a result, Sequential Gaussian Cosimulatian is applied in this study using Markov model 2 and it simulates the missing data of the concerned variables by secondary variables which are more abundant and have acceptable correlations with the primary variables.

2.4. Log-ratio transformation

Compositional data (e.g., raw geochemical data) are the ones in which the elements are non-negative, and their sum is constant. Therefore, the variables in such data are not independent (compositional data only carry the relative rather than the exact information) [35]. These data are defined as:

$$S^D = \{(x_1, x_2, \ldots, x_D) : x_1 > 0,\ x_2 > 0, \ldots, x_D > 0;\ x_1 + x_2 + \ldots + x_D = k\} \tag{2}$$

where S^D is the D dimensional compositional data complex. The constant sum constraint laying in the compositional data leads to the limitation of geo-information, which means if an element in such data rises up in value, the other elements should decrease due to the constant sum constraint. Therefore, this results in the fact that correlation coefficients between the elements are not real and are negatively biased [36, 37].

As a result, these spurious correlations yield unreliable results when standard statistical methods, such as principal component analysis and multidimensional scaling, which are based on data correlation coefficient matrix, are performed [19]. Thus, to deal with the compositions problems, log-ratio transformations are applied to open the closed system data and convert them to Euclidean space [38, 5]. Three types of log-ratio transformations, as mentioned before, can be applied to compositional data. Among them, Additive log-ratio transformation (alr) is used in this study to open the data prior to applying MDS approach. The alr transformation can be shown as the following [39]:

$$i = 1, \ldots, D-1, y_i = \log\left(\frac{x_i}{x_D}\right) \tag{3}$$

where x_D is one of the components which must be the same for the whole data, strictly positive for all components [39] and also of no importance in the study. By applying alr transformation, the component which is the denominator is eliminated, and subsequently D-1 transformed component will be produced [39]. Also, since only some of the components are of interest in some studies, the concept of subcomposition needs to be introduced. A subcompositin is part of a full composition which is normalized. In addition, its covariance relationship between the components in subcomposition is different from the one in full composition [20]. The subcomposition can be indicated as the following [24]:

$$x = C(x) = \left[\frac{k.x_1}{\sum_{i=1}^{D} x_i}, \frac{k.x_2}{\sum_{i=1}^{D} x_i}, \ldots, \frac{k.x_D}{\sum_{i=1}^{D} x_i}\right] \tag{4}$$

where, k is the constant sum. Thus, the first step to apply log-ratio transformation is to construct the subcomposition through equation (4).

2.5. Multidimensional scaling

Multidimensional scaling (MDS) is a multivariate technique which reduces the dimensionality of the dataset and aims to

disclose the relationships existing between the variables in the complex. The relationships between the variables are shown in one dimensional or multidimensional configuration in which each variable is represented by a point. The input data used for MDS can represent either similarities or dissimilarities. The distance between the points in the configuration reveals the relationships between the variables. If the data indicate similarities, a shorter distance represents more similarity and a longer distance represents less similarity. This is the other way around if the input data represent dissimilarities [15].

MDS includes two main methods named metric MDS and non-metric MDS. Metric MDS assumes that the distances (similarity or dissimilarity) between the variables are Euclidean distances. It aims to achieve distances in the configuration which are similar in value to the observed ones. In contrast, absolute values are not meaningful in Non-metric MDS. What is important is the distance between a special pair in relation to distances between other pairs of variables. In this case, the distances between variables are ranked with the largest distance ranking first, and the smallest distance ranking last. Then, monotonic transformation of the similarities is calculated to obtain scaled similarities (f(p)).

In Non-metric MDS, the aim is to achieve a configuration in which the distances between each pair have the same rank order compared to the observed pairs' ranks [40]. Non-metric MDS is applied in this study to find the real underlying structure of the original data and to compare it with the structures gained from imputed data by the already mentioned approaches. There are some ways to examine the validation of the MDS results. Stress function is one way. The smaller the stress value becomes, the more reliable the results will be. Stress function is as follows [40]:

$$STRESS = \sqrt{\frac{\sum\left(f\ (p)-d\right)^2}{\sum d^2}} \qquad (5)$$

The Stress function is the squared differences between the scaled similarities (f(p)) and distances between the points in the MDS configuration (d).

3. Case study

3.1 Case study area and geochemical data

3.1.1. Location and brief information of the gold deposit

The Sari Gunay Au/Sb prospect is located in the SE corner of Kordestan Province in NW Iran, approximately 60 km NW of Hamadan (Fig. 1). Primary investigation was done by the Rio Tinto Mining and Exploration Limited in May and August 1999. This investigation resulted in the identification of a 16 km^2 hydrothermal alteration system associated with the mineralization of gold, antimony, and arsenic, which occurred in Oligo-Miocene dacitic to andesitic porphyries and vent tuffs. Then, Rio Tinto/CESCO Joint Venture started to perform the surface exploration in the first half of 2000 which successfully led to identifying a 1300 m x 400 m well-defined and zoned gold soil anomaly [27].

3.1.2. Channel sampling geochemical data

In the study area, 1064 lithogeochemical samples were collected along six trenches drilled with NW/SE strike. ICP-OES Multi-element analysis was performed for all 1064 samples. Among the datasets, there are several elements that have various percentages of the missing data. The existence of the missing data was due to the lower amount of the analysis instrument than the detection limit. Among the elements which had missing data, four with different percentages were selected:

1. Co contains 52 missing data equal to 4.9 percent of the whole data
2. Ti contains 172 missing data equal to 16.2 percent of the whole data
3. Hg contains 376 missing data equal to 35.3 percent of the whole data
4. Li contains 662 missing data equal to 62.2 percent of the whole data.

3.2. Results

The main purpose of this study is to perform three different approaches of Half Detection Imputation, Multiple Imputation, and Collocated Cosimulation Markov model 2 to impute the censored data and finally define the best approach. The comparison between the approaches is done by MDS by calculating:

1. the distances between the variables in the configurations which are gained from the imputation approaches

2. the distances between the variables in the configuration which is gained from the original data.

A subcomposition of a geochemical dataset is selected to impute. The dataset includes 1064 samples containing Co, Hg, Li, and Ti with different percentages of the missing data (Table 1). The samples were collected from six trenches. First, the relationships between the variables which were gained by applying the MDS to the original data and imputed data will be demonstrated (Fig. 2) and, then, the results and the detailed discussion for each approach will be taken into account.

Table 1. Mean and standard deviation with detection limits and missing percentage of the elements.

Elements	Mean (ppm)	Stdev (ppm)	detection limit (ppm)	Missing (%)
CO	8.25	6.46	1	4.9
Hg	15.24	33.204	1	35.3
Li	4.18	2.77	2	16.2
Ti	33.32	47.13	10	63.2

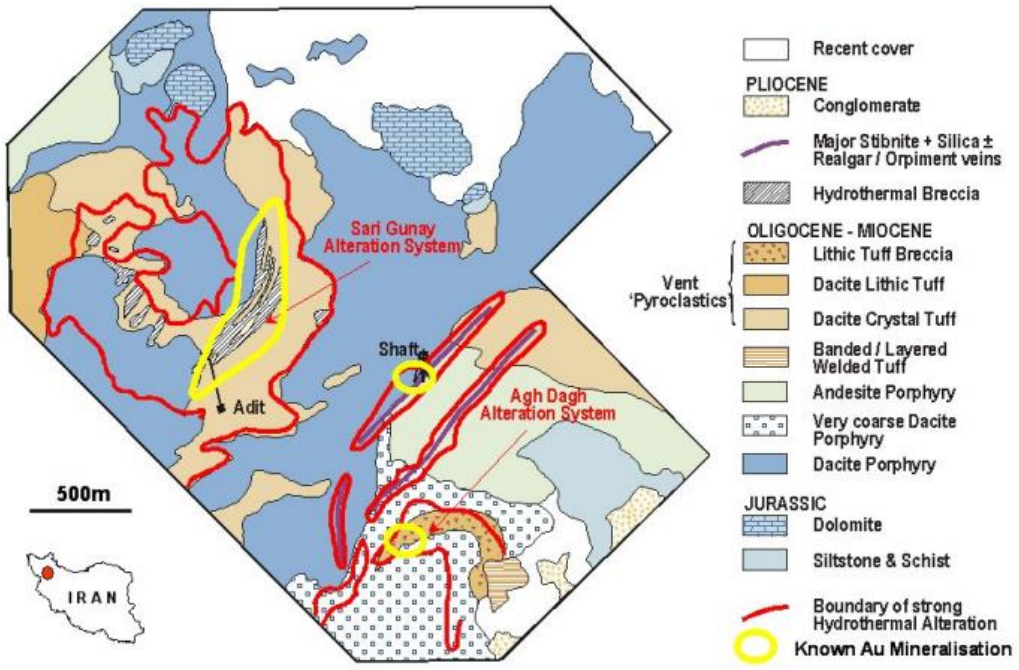

Fig. 1. Sari Gunay area geology and alteration. The map shows the presence of the eroded dacitic to andesitic dome complex related to the large hydrothermal alteration system. This complex includes a large eroded strato-volcano or two smaller volcanoes (Sari Gunay and the Agh Dagh).

3.2.1. Original data

First, in order to discover the multivariate structure of the original data by performing the MDS, log-ratio transformation using alr transformation is applied to the subcomposition to open the closed system. This is done using Ce as the denominator. Then, correlation coefficients (Table 2) of the transformed data are calculated and used as the input data for performing the MDS. Non-metric MDS is used in this study for reducing the dimensions of the composition to both 2D and 3D configurations. Due to the difficulty of displaying the 3D configuration of the data, 2D configuration of the original data and imputed data are shown in this study to demonstrate the relationships between the already mentioned elements (Fig. 2). However, to compare the results produced by the imputation approaches, distances obtained from 3D configuration are used due to being more reliable.

Table 2. Correlation coefficient matrix of the original data.

Elements	CO	Hg	Li	Ti
CO	1	-0.094	0.487	0.447
Hg	-0.094	1	-0.093	0.032
Li	0.487	-0.093	1	0.671
Ti	0.447	0.032	0.671	1

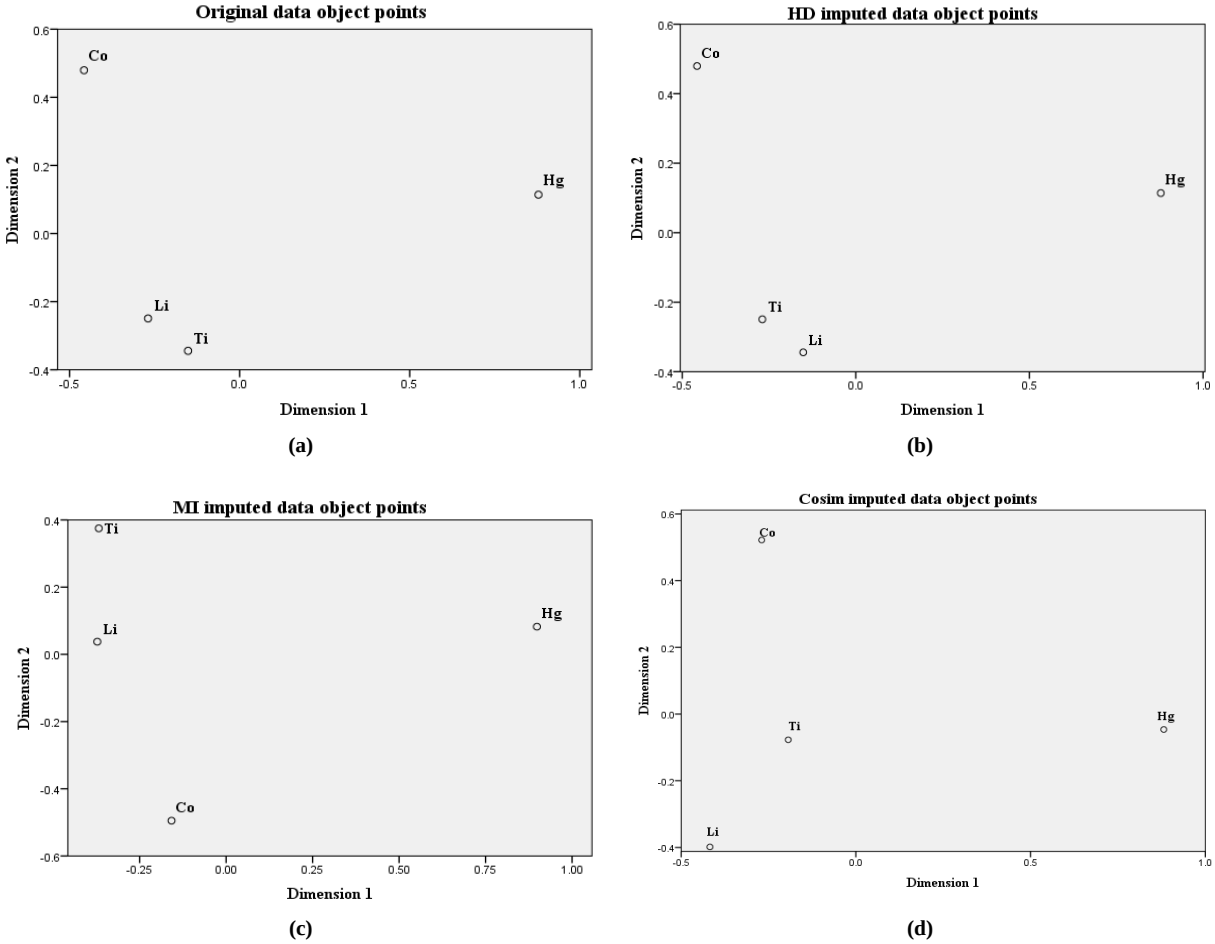

Fig. 2. 2D configuration gained by applying the MDS. (a) configuration of the original data, (b) configuration of the Half detection imputed data, (c) configuration of the MI imputed data, (d) configuration of the Cosimulation imputed data.

As shown in Figure 2a, elements which have the higher correlation coefficients are closer to each other in the configuration and those which have weaker correlations are far from each other. It means the distance between Li and Ti which have a correlation coefficient equal to 0.671 has to be shorter than the distance between the other pairs of elements. Consequently, the distance between Co and Hg which have the strongest negative correlation coefficient has to be more than the distance between the other pairs. As mentioned before, these relations should be maintained after imputing the missing data.

3.2.2. Half detection imputation

Half detection imputation is applied to the data as the first approach by replacing the missing data with a value equal to the half detection limit of the elements. After replacement, alr transformation is performed on the complete dataset, using Ce as the denominator to open the subcomposition. Then, the correlation coefficients of the transformed data are computed, and used as the input data to apply Non-metric MDS. Moreover, the distances between the elements are calculated in 3D configuration. To demonstrate the changes having occurred in the relationships between

the elements, 2D configuration is used (Fig. 2b). As shown in Fig. 2b, two of the elements, Li and Ti, were altered and took each other's positions in the configuration. However, they seemed to hold their distance. The alteration in the positions was due to the changes that occurred in their correlation against the correlation between Hg and Co. The distances between the elements calculated in the 3D dimension are presented in Table 3.

3.2.3. Multiple imputation

Multiple imputation is applied in this study for imputation using Fully Conditional Specification (FCS). In addition, the Predictive Mean Matching (PMM) model is used for scale variables. FCS is an iterative Markov Chain Monte Carlo (MCMC) method which can be used for the arbitrary missing data patterns, and is applicable in this study. In this method, a univariate model is fit for each iteration and variable, using all other variables in the model as the predictors. Then, the missing data for the variable being fit are imputed, and the iteration will continue up to the maximum number of the iterations (which is ten in this study). Finally, the imputed values obtained from the maximum iteration are saved to the impute dataset (SPSS 17.0 tutorial). PMM is one of the several kinds of linear regressions, which results in a good compatibility between the imputed values and the closer observed data (SPSS 17.0 tutorial).

Thus, after applying the multiple imputation using the already mentioned methods, alr transformation is applied to the imputed dataset, using Ce as the denominator like the half detection. Then, correlation coefficients between the transformed data are calculated, and used as input data for MDS. The 2D configuration of the imputed dataset is shown in Figure 2c. The general relationships between the elements remain similar to the original ones in the configuration; however, as can be seen in Figure 2c, there are more changes in the distances between the pairs (Table 3) from the original ones compared to the distances gained from the half detection. This means that half detection could better reproduce the original relationships between the elements. It may be due to the origin of the missing data in this study. The missing data exist owing to being less in value than the detection limit of the analysis instrument. Therefore, imputing some values which are more than the detection limit of the elements may cause these discrepancies for the imputed data.

3.2.4. Collocated cosimulation Markov model 2

Sequential Gaussian cosimulation is applied in this study, using Markov model 2 as the collocated Cokriging method. Secondary variables are chosen from among the geochemical dataset variables for each of the elements. These secondary variables have to satisfy two conditions:

1. the secondary variables should be more abundant than the primary variables and sampled for each node of the collocated grid
2. They should have high correlations with the primary variables.

To meet the two mentioned conditions, three different elements are chosen as the secondary variables. Au is considered the secondary variable for simulating the Hg and it has a correlation coefficient equal to 0.617 with Hg. Ni is regarded as the secondary variable for simulating the Co and it has a correlation coefficient equal to 0.852 with Co. Also, Mg is considered the secondary variable for simulating both Li and Ti and it has a correlation coefficient equal to 0.716 with Li and 0.680 with Ti.

Then, variograms of the secondary variables are computed (Fig. 3). After performing cosimulation for each of the elements, the imputed dataset is transformed by alr transformation, using Ce as the denominator and, finally, all steps are taken as previously mentioned. The MDS configuration of the imputed dataset obtained from cosimulation is shown in Figure 2d, and the distances between each of the elements are calculated (Table 3). The change in distances between the elements in the configuration is shown in Figure 2d. This difference is more easily noticed in position change of Ti and Li which are reported to stay farther from each other than their positions in the original data.

Results reveal that the half detection approach produces better results compared to the two other approaches. Also, compared to MI, it is cosimulation approach that yields better result according to the imputed data structures obtained.

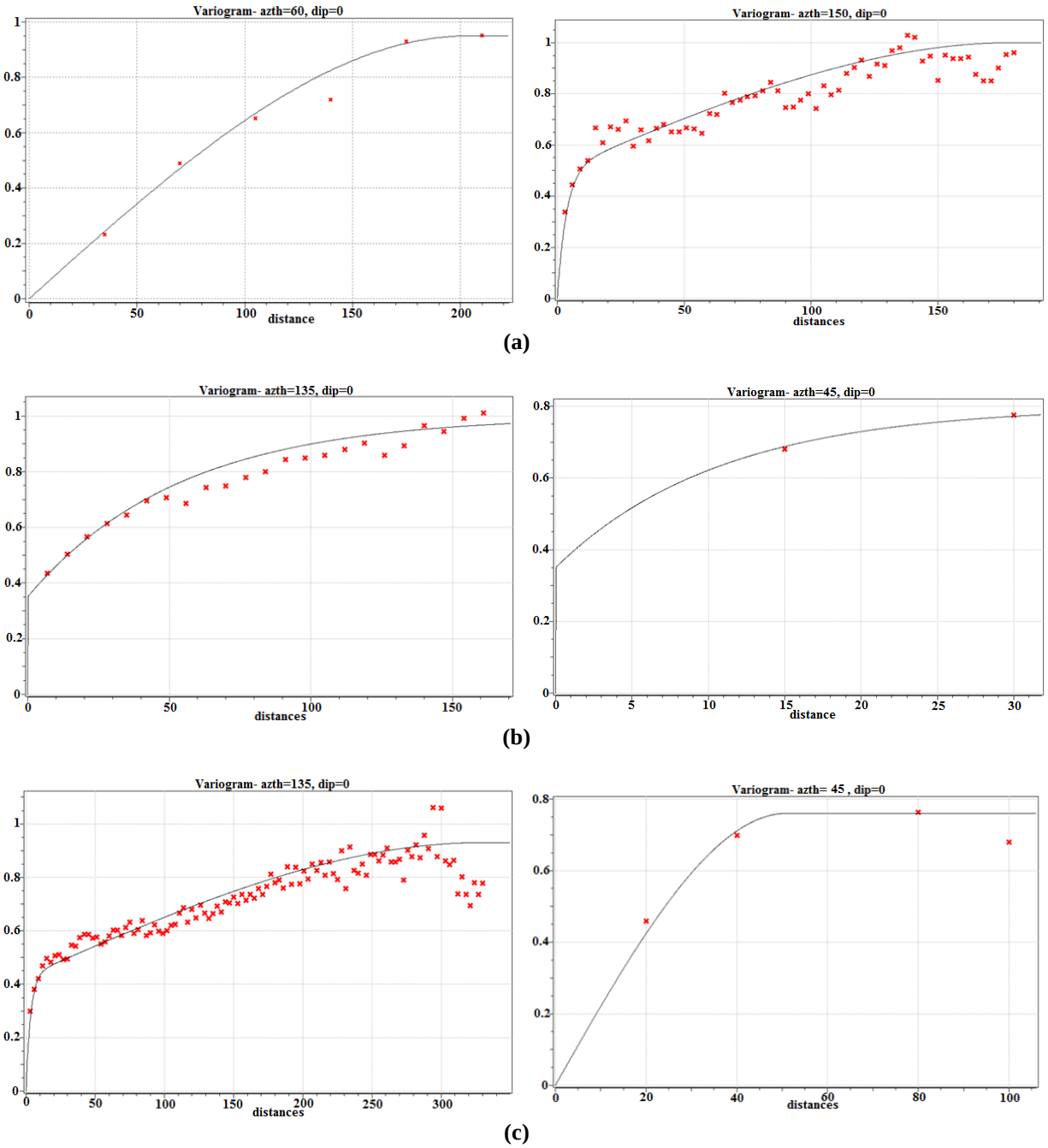

Fig. 3. Variograms of secondary variables. (a) Variograms of Au in major (left) and minor (right) direction, (b) Variograms of Mg in major (left) and minor (right) direction, (c) Variograms of Ni in major (left) and minor (right) direction

4. Discussion

For comparison between the approaches, results obtained from applying the MDS method to the original data and to the outcomes gained from the three imputation approaches including the element distances, the stress value, and the SSE value (which is computed by considering the elements distances in the original configuration and the distances in the imputed data configuration) of each of the approaches are computed (Table 3).

Table. 3. MDS results containing the 3D elements distances, Stress value and SSE value for each of the approaches

Data		Distances				Stress (3D)	SSE (3D)
		Co	Hg	Li	Ti		
Original	Co	0				0.0	-
	Hg	1.301	0				
	Li	0.792	1.165	0			
	Ti	0.858	1.101	0.612	0		
Half Detection	Co	0				0.0	0.016
	Hg	1.301	0				
	Li	0.858	1.101	0			
	Ti	0.792	1.165	0.612	0		
Multiple Imputation	Co	0				0.0	0.06
	Hg	1.126	0				
	Li	0.742	1.22	0			
	Ti	0.836	1.237	0.682	0		
Cosimulation	Co	0				0.0	0.048
	Hg	1.208	0				
	Li	0.882	1.274	0			
	Ti	0.725	1.086	0.659	0		

The zero value for the Stress function shown in the Table 3 represents the reliability of the outcomes gained from applying the MDS method. As can be seen in Table 3, distances between the Co-Hg and Li-Ti remain constant after imputing the missing data using half detection compared to the distances between the elements in each pair in the original data. However, the distances between the elements in other pairs reveal the substitution in the positions of Li and Ti according to the substitution of the distances between Li-Co, Ti-Co, Li-Hg, and Ti-Hg, which is obvious from the configurations illustrated in the previous section. On the other hand, relocations between the elements in the configuration obtained from the multiple imputation caused more differences between the distances between elements compared to the original ones. As a result, the SSE value for this approach increases.

However, the lower SSE value for the cosimulation approach uncovered the fact that there is less relocation in the configuration gained from cosimulation against the multiple imputation approach. Therefore, the cosimulation approach could better reproduce the variable structures in comparison to the multiple imputation approach. Finally, considering the SSE values for all of the approaches, MDS determines the approaches which produced better results. These approaches are sequentially the half detection approaches, the Collocated Cosimulation Markov model 2, and the multiple imputation approach.

5. Conclusion

In this study, three different approaches called Half Detection Imputation, Multiple Imputation, and Collocated Cosimulation Markov Model 2 are used to impute the censored data. The results obtained from these approaches are compared to each other. The results of this study are as follows:

a) Multiple imputation approach utilized in this study using MCMC method to impute missing data leads to greater differences in the configuration obtained from this approach compared to the original configuration of the data. Also, the SSE value computed for this approach was higher than those of the other approaches.

b) Collocated Cosimulation using Markov model 2 as estimation method was another approach used in this study for imputing the missing data. It produced better results compared to multiple imputation approach because its SSE computed value was lower than the value computed in the multiple imputation method.

c) Half detection approach was also applied in this study. This approach demonstrated the better result in MDS configuration due to having the least SSE value and reproduced the original configuration better than the other two approaches.

Acknowledgment

The authors would like to thank CESCO mining company for providing the geochemical data.

References

[1]. Croghan, C., & Egeghy, P. P. (2003). Methods of dealing with values below the limit of detection using SAS. Southern SAS User Group, St. Petersburg, FL, 22-24.

[2]. Taylor, J. K. (1987). Quality assurance of chemical measurements. CRC Press.

[3]. Lyles, R. H., Fan, D., & Chuachoowong, R. (2001). Correlation coefficient estimation involving a left censored laboratory assay variable. *Statistics in Medicine*, *20*(19), 2921-2933.

[4]. Grunsky, E. C., & Smee, B. W. (1999). The differentiation of soil types and mineralization from multi-element geochemistry using multivariate methods and digital topography. Journal of Geochemical Exploration, 67(1), 287-299.

[5]. Carranza, E. J. M. (2011). Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. Journal of Geochemical Exploration, 110(2), 167-185.

[6]. Rubin, D. B. (1978). Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In Proceedings of the survey research methods section of the American statistical association., American Statistical Association, Vol. 1, pp. 20-34.

[7]. Rubin, D. B. (1988). An overview of multiple imputation. In Proceedings of the survey research methods section of the American statistical association, pp. 79-84.

[8]. Van Buuren, Stef, and Karin Oudshoorn. (1999). "Flexible multivariate imputation by MICE." Leiden, The Netherlands: TNO Prevention Center, Netherlands.

[9]. Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the royal statistical society. Series B (methodological), 1-38.

[10]. Barnett, R. M., & Deutsch, C. V. (2015). Multivariate Imputation of Unequally Sampled Geological Variables. Mathematical Geosciences, 1-27.

[11]. Goovaerts, P., Geostatistics for natural resources evaluation. (1997). Oxford University Press, New York, 483 p.

[12]. Munoz, B., Lesser, V. M., & Smith, R. A. (2010). Applying Multiple imputation with Geostatistical Models to Account for Item Nonresponse in Environmental Data. Journal of Modern Applied Statistical Methods, 9(1), 27.

[13]. Zhang, X., Jiang, H., Zhou, G., Xiao, Z., & Zhang, Z. (2012). Geostatistical interpolation of missing data and downscaling of spatial resolution for remotely sensed atmospheric methane column concentrations. International journal of remote sensing, 33(1), 120-134.

[14]. Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. Psychometrika, 17(4), 401-419..

[16]. Deutsch, J. L., & Deutsch, C. V. (2014). A multidimensional scaling approach to enforce reproduction of transition probabilities in truncated plurigaussian simulation. Stochastic Environmental Research and Risk Assessment, 28(3), 707-716.

[17]. Boisvert, J. B., & Deutsch, C. V. (2011). Programs for kriging and sequential Gaussian simulation with locally varying anisotropy using non-Euclidean distances. Computers & Geosciences, 37(4), 495-510.

[18]. Pawlowsky-Glahn, V., & Egozcue, J. J. (2006). Compositional data and their analysis: an introduction. Geological Society, London, Special Publications, *264*(1), 1-10.

[19]. Aitchison, J. (1983). Principal component analysis of compositional data. Biometrika, 70(1), 57-65.

[20]. Aitchison, J. (1986). The Statistical Analysis of Compositional Data, first ed. Chapman and Hall, London, UK, 416 pp.

[21]. Aitchison, J. (1999). Logratios and natural laws in compositional data analysis. Mathematical Geology, 31(5), 563-580.

[22]. Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A., & Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. Mathematical Geology, 32(3), 271-275.

[23]. Buccianti, A., & Pawlowsky-Glahn, V. (2005). New perspectives on water chemistry and compositional data analysis. Mathematical Geology, 37(7), 703-727.

[24]. Buccianti, A., & Grunsky, E. (2014). Compositional data analysis in geochemistry: Are we sure to see what really occurs during natural processes?. Journal of Geochemical Exploration, 141, 1-5.

[25]. Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. Mathematical Geology, 35(3), 279-300.

[26]. de Caritat, P., & Grunsky, E. C. (2013). Defining element associations and inferring geological processes from total element concentrations in Australian catchment outlet sediments: multivariate analysis of continental-scale geochemical data. Applied Geochemistry, 33, 104-126.

[27]. Wilkinson, L. D. (2005). Geology and mineralization of the Sari Gunay gold deposits, Kurdistan province, Iran. Open-File ReportRio-Tinto Mining and Exploration Ltd.

[28]. Yuan, Y. C. (2010). Multiple imputation for missing data: Concepts and new development (Version 9.0). SAS Institute Inc, Rockville, MD.

[29]. Ni, D., & Leonard, J. D. (2005). Markov Chain Monte Carlo Multiple imputation for Incomplete ITS Data Using Bayesian Networks.

[30]. Schafer, J. L. (1997). Imputation of missing covariates under a multivariate linear mixed model. Unpublished technical report.

[31]. Almeida, A. S. (1993). Joint simulation of multiple variables with a Markov-type coregionalization model. Unpublished doctoral dissertation, Stanford University, Stanford, 199 p.

[32]. Almeida, A. S., and Journel, A. G. (1996). Joint simulation of multiple variables with a Markov-type coregionalization model. Math. Geology, v. 26, no. 5, p. 565–588.

[33]. Journel, A. G. (1999). Markov models for cross-covariances. Mathematical Geology, 31(8), 955-964.

[34]. Shmaryan, L. E., & Journel, A. G. (1999). Two Markov models and their application. Mathematical geology, 31(8), 965-988.

[35]. Egozcue, J. J., & Pawlowsky-Glahn, V. (2005). Groups of parts and their balances in compositional data analysis. Mathematical Geology, 37(7), 795-828.

[36]. Thomas, C. W., & Aitchison, J. (2006). Log-ratios and geochemical discrimination of Scottish Dalradian limestones: a case study. Geological Society, London, Special Publications, 264(1), 25-41.

[37]. Wang, W., Zhao, J., & Cheng, Q. (2014). Mapping of Fe mineralization-associated geochemical signatures using logratio transformed stream sediment geochemical data in eastern Tianshan, China. Journal of Geochemical Exploration, 141, 6-14.

[38]. Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society. Series B (Methodological), 139-177.

[39]. Job, M. R. (2012). Application of Logratios for Geostatistical Modelling of Compositional Data (Doctoral dissertation, University of Alberta).

[40]. Wickelmaier, F. (2003). An introduction to MDS. Sound Quality Research Unit, Aalborg University, Denmark.