Ore Geology Reviews 150 (2022) 105132

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeorev

Geostatistical modeling of heterogeneous geo-clusters in a copper deposit integrated with multinomial logistic regression: An exercise on resource estimation

Nasser Madani [a,*], Mohammad Maleki [b], Saeed Soltani-Mohammadi [c]

[a] School of Mining and Geosciences, Nazarbayev University, Astana, Kazakhstan
[b] Department of Metallurgical and Mining Engineering, Universidad Católica del Norte, Antofagasta, Chile
[c] Department of Mining Engineering, University of Kashan, Kashan, Iran

ARTICLE INFO

Keywords:
Multinomial logistic regression
Non-stationary
Heterogeneity
Sequential indicator simulation
Porphyry copper deposit

ABSTRACT

Resource estimation is the main and primary step in the development of a mining project. Principally, it is necessary to first identify the geological domains through boreholes, model them at unsampled locations, and then evaluate the grade(s) of interest inside each built domain. The traditional determination of these categorical domains over the sampling points is suboptimal as it considers mostly-one or two variables from core logging. This leads to the neglect of the influence of other significant variables. To circumvent the problem of estimation domain identification, spatially dependent clustering machine learning algorithms can be of great help in detecting such domains. However, one problem that may appear when using these techniques is that the resulting geo-domains (geo-clusters) obtained by the clustering technique might be heterogeneous and show a non-stationary property. The reason is that the aim of these spatially dependent techniques is to produce compact and spatially contiguous clusters, which are well suited to establishing non-stationary geo-domains. This makes the procedure of modelling challenging as it necessitates the use of advanced geostatistical techniques to propagate the heterogeneous geo-clusters at unsampled locations.

An algorithm is presented in this study that employs a non-stationary sequential indicator simulation paradigm to model such complex variability of heterogeneous geo-clusters. Since the spatial trends of underlying geo-clusters are required in this simulation method, in this study, we propose the use of multinomial logistic regression to infer these trends. The algorithm was tested using an actual case study from a porphyry copper deposit in Iran, where Cu, Mo, Au, Rock Quality Designation (RQD), mineralization zones, alteration types, and rock types were employed to identify and spatially model the heterogeneous geo-domains in the entire deposit. The results were compared with a conventional sequential indicator simulation where no trend was used. An examination of the resulting maps using several evaluation criteria including visual inspection of the realizations, probability maps, reproduction of proportion of each geo-cluster, connectivity measures, and trend analysis, showed that the findings of the proposed algorithm were superior in modelling heterogeneous geo-domains.

1. Introduction

Mineral resource modeling is an important task providing a basement in the value chain of mine development (Sinclair and Blackwell, 2006; Rossi and Deutsch, 2014; Abzalov, 2016). Proper modelling of ore grades in a deposit significantly impacts the long-term planning of a mining operation (Maleki et al., 2020; Maleki et al., 2021). In a geological borehole database, one usually deals with two types of variables: continuous (e.g., ore grades and mineral grades) and categorical (e.g., alteration, mineralization zones, rock types, lithology, weathering, and other geological characteristics) variables. A typical act in the mining industry is to first split the deposit into categorical sub-domains (estimation domains) and then, using geostatistical/interpolation techniques, to model the corresponding continuous variable(s) inside each domain separately (Rossi and Deutsch, 2014; Emery and Séguret, 2020). The benefit of this technique is that the continuous variable(s) inside the domains are deemed homogenous and stationary, which facilitates their prediction (Sinclair and Blackwell, 2006; Moon et al., 2005; Yunsel and

* Corresponding author.
E-mail address: nasser.madani@nu.edu.kz (N. Madani).

https://doi.org/10.1016/j.oregeorev.2022.105132
Received 13 June 2022; Received in revised form 25 September 2022; Accepted 27 September 2022
Available online 28 September 2022

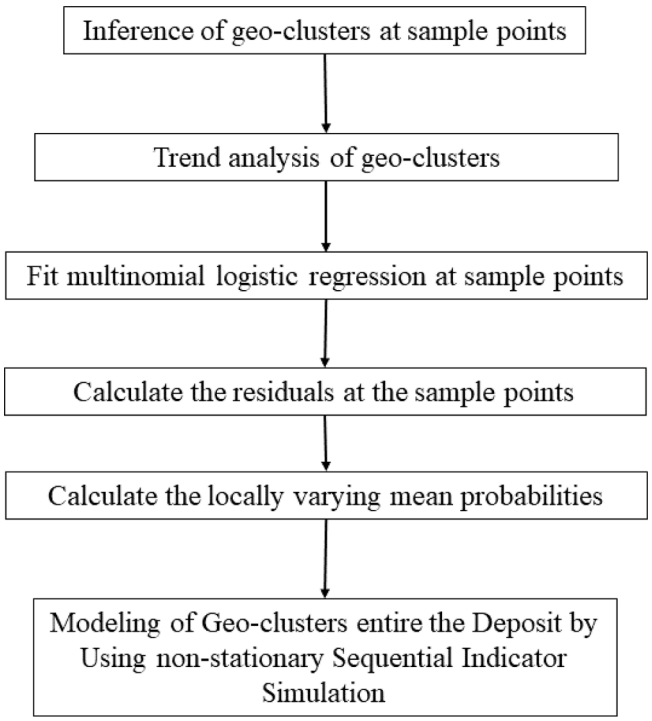

Fig. 1. Flowchart stating the main steps of the proposed approach in this study.

Ersoy, 2011; Haldar, 2018; Rossi and Deutsch, 2014). These categorical domains are the geological equivalent to the stationary distribution of mineralization. The latter step (modelling the continuous variable(s)) is straightforward, but the former step is challenging because there exist different complexities in identifying such "estimation domains". This challenge stems from first recognizing them over the sampling points (borehole logs) and, second, modeling them at target or unsampled points. For the former, there are different approaches that can be carried out based on the geological setting of a deposit, where the interpretation of categorical variables plays an important role. For this, geological core logging is the foundation for characterizing such geo-domains (Soltani and Hezarkhani, 2011; Adeli and Emery, 2017). For instance, in porphyry copper deposits, a mineralized zone (oxide/sulfide) or rock types can be considered the estimation domains (Madani et al., 2021a; Koike et al., 2022), or the use of lithology in iron deposits (Maleki et al., 2021; Hosseini et al. 2021).

Grade domaining is another alternative for defining these geo-domains. The grade shells obtained by this method are based on truncating the distribution of a continuous variable using the specified thresholds (Emery and Ortiz, 2005; Yunsel and Ersoy, 2011; Iliyas and Madani, 2021). To do so, the main element in the deposit is split into the categories where each belongs to an interval. The method is simple, but the obtained domains should be in agreement with the geological logging interpretation of the sample points.

However, these methods are labor intensive, time consuming, and subject to manual interpretation of the mineral deposits (Fouedjio et al., 2018). Thus, domaining is a clustering problem; another option is to use clustering machine learning algorithms to automatically and quickly

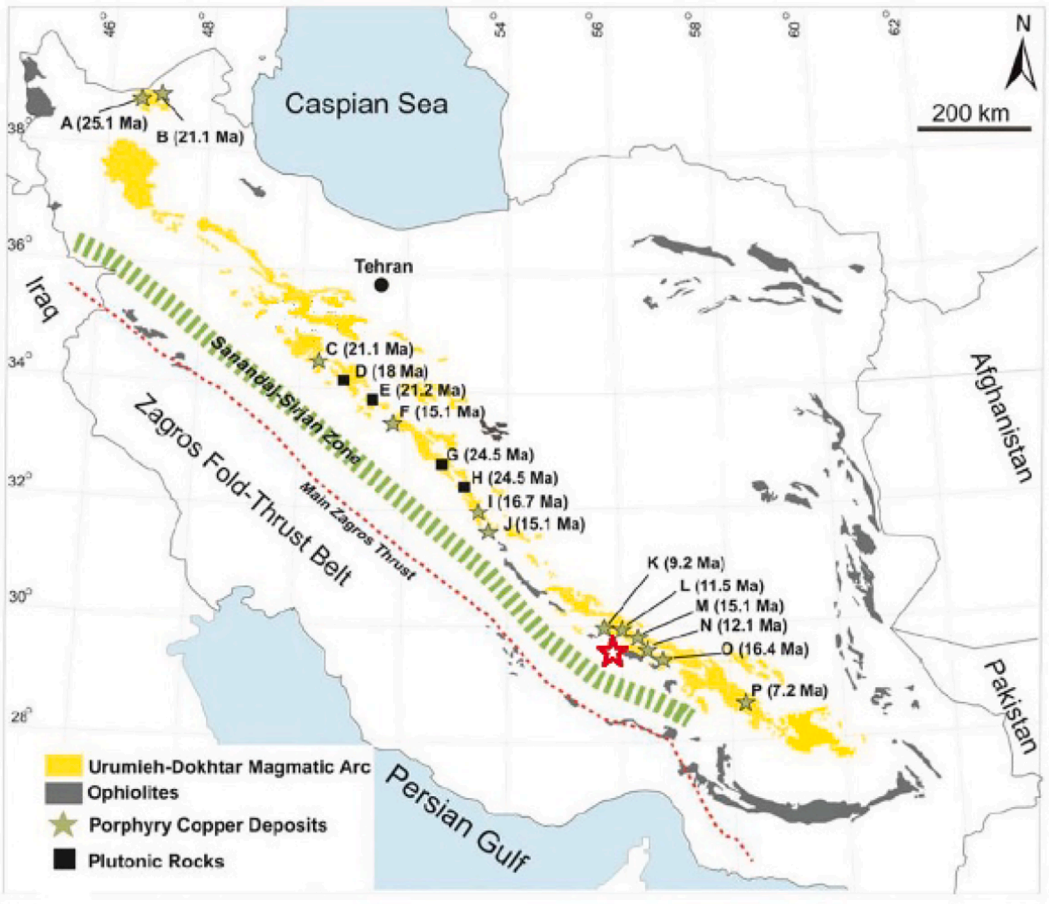

Fig. 2. Simplified geological map of Iran (after Mirnejad et al., 2019). The location of known copper deposits in the UDMA are marked. Sarkuh area is marked by red star. Abbreviations: A: Qarachilar; B: Sungun; C: Dali; D: Kashan; E: Natanz; F: Kahang; G: Ardestan; H: Nodoushan; I: Ali-Abad; J: Darreh-Zerreshk; K: Iju; L: Miduk; M: Sarkuh; N: Sar Cheshmeh; O: Darreh Zar; P: Kerver.

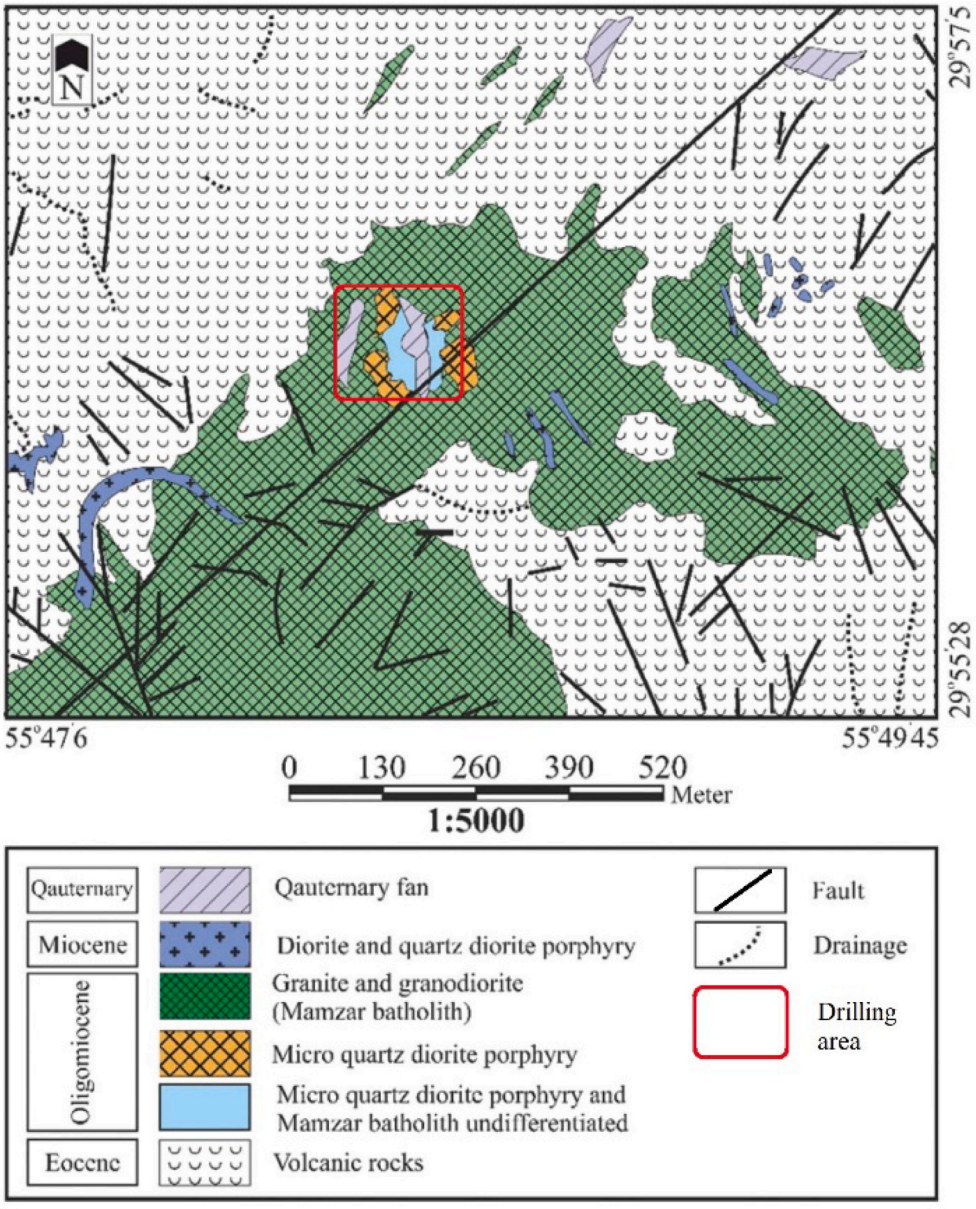

Fig. 3. Simplified map of the Sarkuh Cu deposit (Company, 2008).

identify these geo-domains. Classical clustering algorithms such as hierarchical clustering (Maimon et al., 2005), K-means (Jain, 2010), spectral clustering (Jain et al., 1999), and Gaussian mixture (Scrucca et al. 2016; Madenova and Madani, 2021) can be used for this purpose. For instance, de Sá et al. (2021) used K-means clustering, pluri-Gaussian simulation, and a lithotype rule for the cluster numbers. By using the lithotype rule, a continuous distribution of each cluster was obtained. However, the resulting geo-domains (hereafter, geo-clusters) obtained by the classical clustering algorithms may appear spatially patchy and unstructured. These geo-domains are usually impractical in mining exploitation because, from the mining operational point of view, it is always of interest to plan the underlying domains that are spatially contiguous, connected, and compact.

To solve this issue, several clustering algorithms (Oliver and Webster, 1989; Ambroise et al. 1997; Scrucca, 2005; Romary et al. 2012; Romary et al. 2015; Fouedjio, 2016a; Fouedjio, 2016b; Fouedjio, 2017a; Fouedjio, 2017b; Fouedjio et al. 2018; Martin and Boisvert, 2018; D'Urso and Vitale, 2020) were developed to take into account the spatial dependency of the data so as to establish reasonable geo-clusters suitable for mine planning and exploitation. The application of these spatially dependent clustering methodologies has been successfully applied in different geo-domain characterizations for ore body modeling, as thoroughly discussed in Fouedjio et al. (2018), Martin and Boisvert, (2020), and Moreira et al. (2020). One benefit of these methods is that they produce compact, spatially connected, and contiguous categorical domains. This, however, may lead to the generation of non-stationary geo-domains that display heterogeneity throughout the deposit. This property of the random function model requires the application of advanced geostatistical interpolation tools for modeling these geo-clusters at unsampled locations in order to provide the corresponding volumes to model the continuous variable(s).

These methods can be classified into deterministic and stochastic approaches. Deterministic methods only predict a unique geo-cluster at unsampled locations, and the uncertainty cannot be quantified. Instead, stochastic geostatistical methods are of particular interest for this purpose. Among others, sequential indicator simulation (Alabert, 1987; Journel and Alabert, 1990) is a commonly used method that is extensively available in most commercial software programs. However, this

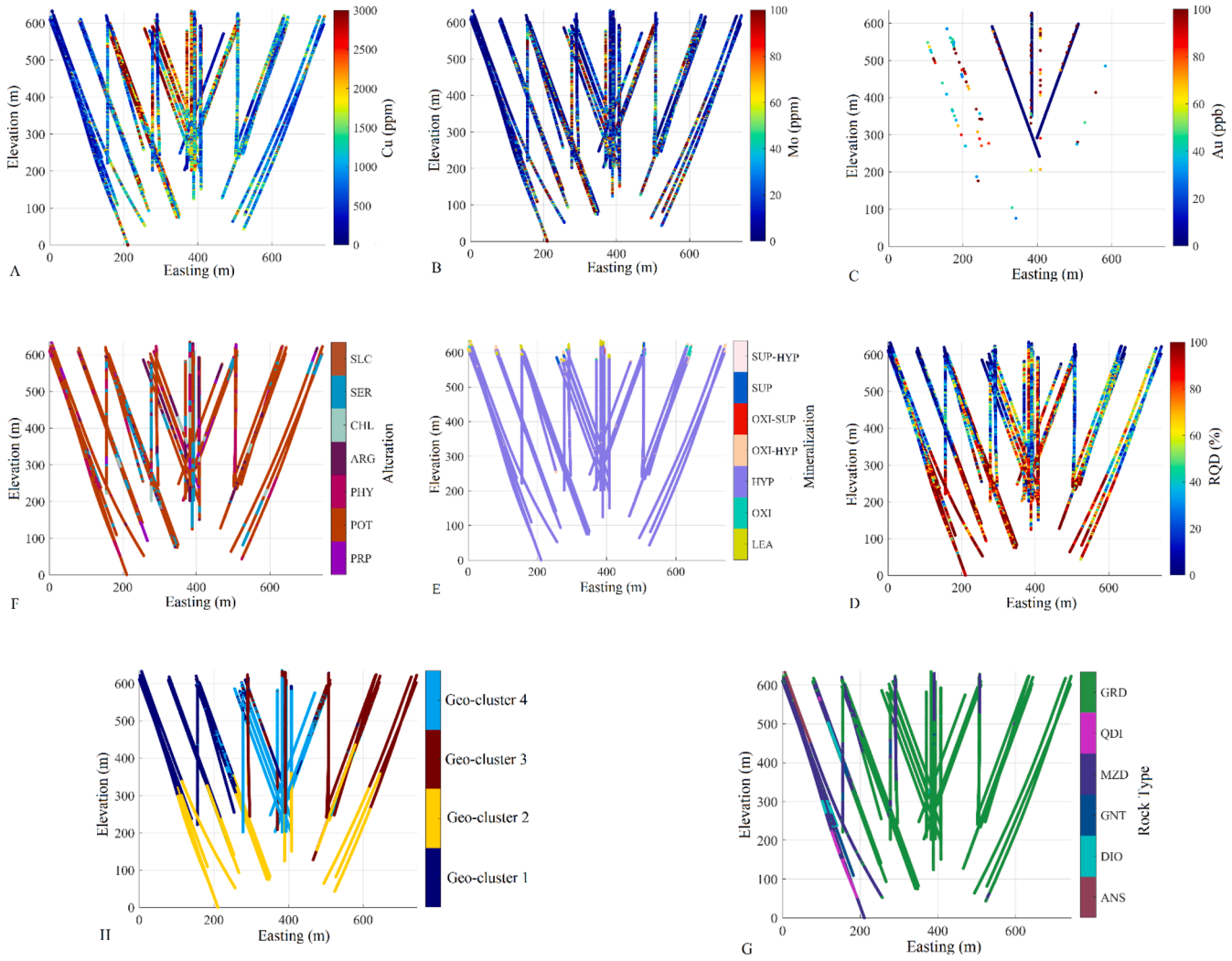

Fig. 4. 2D location maps of boreholes for Cu (ppm) (A), Mo (ppm) (B), RQD (%) (C), Au (ppb) (D), and mineralization zone (E), alteration (F), rock types (G), and geo-clusters (H).

Table 1
Level of relationship between continuous-continuous variables (upper diagonal: Pearson linear correlation, and lower diagonal: Spearman non-linear correlation); categorical-categorical variables (Cramer's V coefficient); and continuous-categorical variables (Cramer's V coefficient).

	Cu	Mo	Au	RQD	Mineralization	Alteration	Rock Type	Geo-clusters
Cu	1	0.122	**0.731**	0.015	W	W	**S**	**S**
Mo	**0.276**	1	0.131	0.027	W	W	**S**	M
Au	**0.507**	0.047	1	−0.129	–	W	W	W
RQD	0.061	0.147	−0.146	1	**S**	M	**S**	**VS**
Mineralization	W	W	–	S	–	**VS**	W	M
Alteration	W	W	W	M	**VS**	–	M	M
Rock Type	**S**	**S**	W	**S**	W	M	–	**VS**

W: weak association, M: Moderate association, S: Strong association, VS: very strong association.

method is suboptimal when the geo-domains are heterogeneous (such as geo-clusters) and show large-scale geological features. In fact, conventional sequential indicator simulation is poor at reproducing the compactness and spatially contiguous geological features, which are desired in modeling the geo-clusters. The reason is that conventional sequential indicator simulation is built based on the stationary property of the random function model and uses only the variogram as two-point statistics. In the case of such complex characteristics as those in geo-clusters, one possibility is to use secondary information (Deutsch, 2006). A methodology is developed in this paper that uses the multinomial logistic regression model to produce such secondary data by integrating it with a non-stationary sequential indicator simulation to model the heterogeneous geo-clusters. The method was tested with a real copper deposit.

The proposed methodology in this study first requires a geo-clustering technique to characterize the geo-clusters at the sample points, then multinomial logistic regression should be used to produce the local probability (secondary information) of each geo-cluster at the sample points and target grid nodes, and then a non-stationary sequential indicator simulation paradigm is used to stochastically model the heterogeneous geo-clusters at the target grid nodes. Therefore, the aim of this paper is threefold: (a) to develop an algorithm that

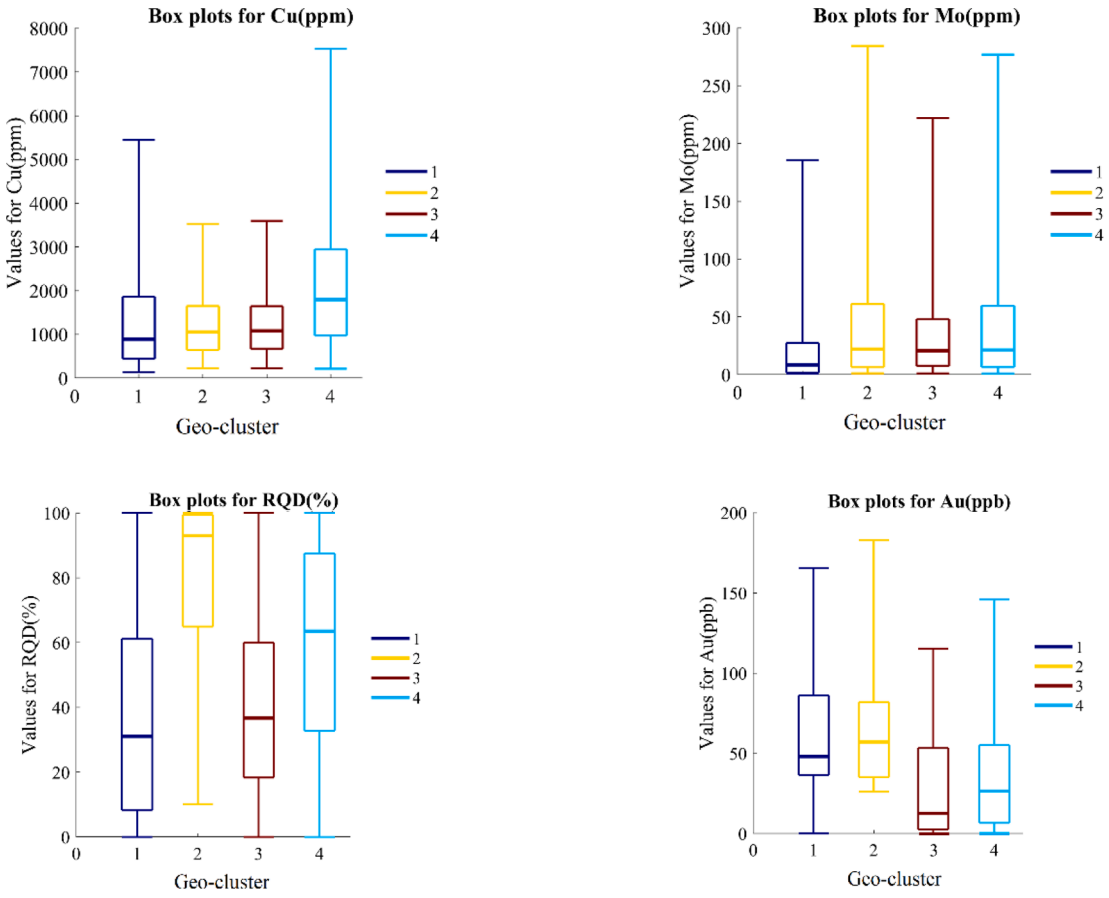

Fig. 5. Boxplots between values of each continuous variable and the resulting geo-clusters.

Table 2
Statistical parameters for Cu(ppm)/Mo(ppm)/Au(ppb)/RQD(%) in each geo-cluster.

	Geo-cluster			
	1	2	3	4
Number of data	2675/ 2675/ 41/ 2742	2661/ 2661/ 10/2670	2866/ 2866/ 165/2866	2739/ 2739/ 374/2771
Average	1449.10/ 39.57/ 63.48/ 37.08	1269.29/ 53.30/ 66.40/ 79.17	1303.71/ 41.69/ 4.83/ 40.64	2298.1/ 51.33/ 13.88/ 58.49
Variance	2244389.39/ 42770.27/ 1696.25/ 1010.77	854962.46/ 11555.40/ 1918.84/ 728.99	983722.84/ 4668.91/ 542.39/ 772.68	3973801.16/ 9271.89/ 1935.69/ 1023.08
Minimum	14/0/0/0	5/0.67/26/0	14/0.82/0/0	56/0/0/0
Median	890/8.4/48/ 30.02	1054/22/57/ 92.96	1080/20.7/ 15.3/36.67	1795/21.2/ 24.7/63.51
Maximum	24600/4843/ 195/100	13061/1914/ 183/100	15378/1414/ 206/100	17937/1462/ 340/100
COV*	1.03/5.22/ 0.64/0.85	0.72/2.01/ 0.65/0.34	0.76/1.63/ 4.82/ 0.68	0.86/1.87/ 3.16/0.54

*COV: coefficient of variation.

integrates a non-stationary sequential indicator simulation with multinomial logistic regression; (b) to test the algorithm using an actual case study in a porphyry copper deposit; and (c) to compare the results with a conventional sequential indicator simulation and discuss them according to several evaluation criteria.

2. Materials and methods

2.1. Geo-clustering

Any unsupervised spatially dependent machine learning algorithm can be used to infer geo-clusters that are not only compact but also contiguous and well structured. A complete review of these methods are thoroughly discussed in the work by Fouedjio (2020). However, among others, geostatistical hierarchical clustering (Romary et al. 2012; Romary et al. 2015) is of particular interest in this study since it employs both continuous and categorical variables simultaneously for identification of geo-clusters suitable for mineral resource modelling (Madani et al. 2021b). The notion of geostatistical hierarchical clustering is discussed in depth in the work by Romary et al. (2015). As a brief description, the sample points in this method should be first linked using Delaunay triangulation graphs (Green and Sibson, 1978). Then, the agglomerative hierarchical clustering algorithm clusters the sample points based on the established connections in the Delaunay network, which is based on a matrix of distance between two linked data locations. This distance is a function of the measured continuous and categorical variables at the data location associated with the weights assigned to each. Among the continuous variables, the coordinates can also be embedded in this clustering paradigm. Using the coordinates and the established linkage graph, the spatial continuity is taken into account to obtain the clusters. This is the only improvement to the traditional hierarchical clustering technique forming the geostatistical hierarchical clustering algorithm.

In addition, this method is easily accessible in commercial software programs such as Minestis® and Isatis.neo®. Once the geo-clusters are identified thorough the sample locations, the next step is to then model and propagate them into the unsampled target grid nodes using the

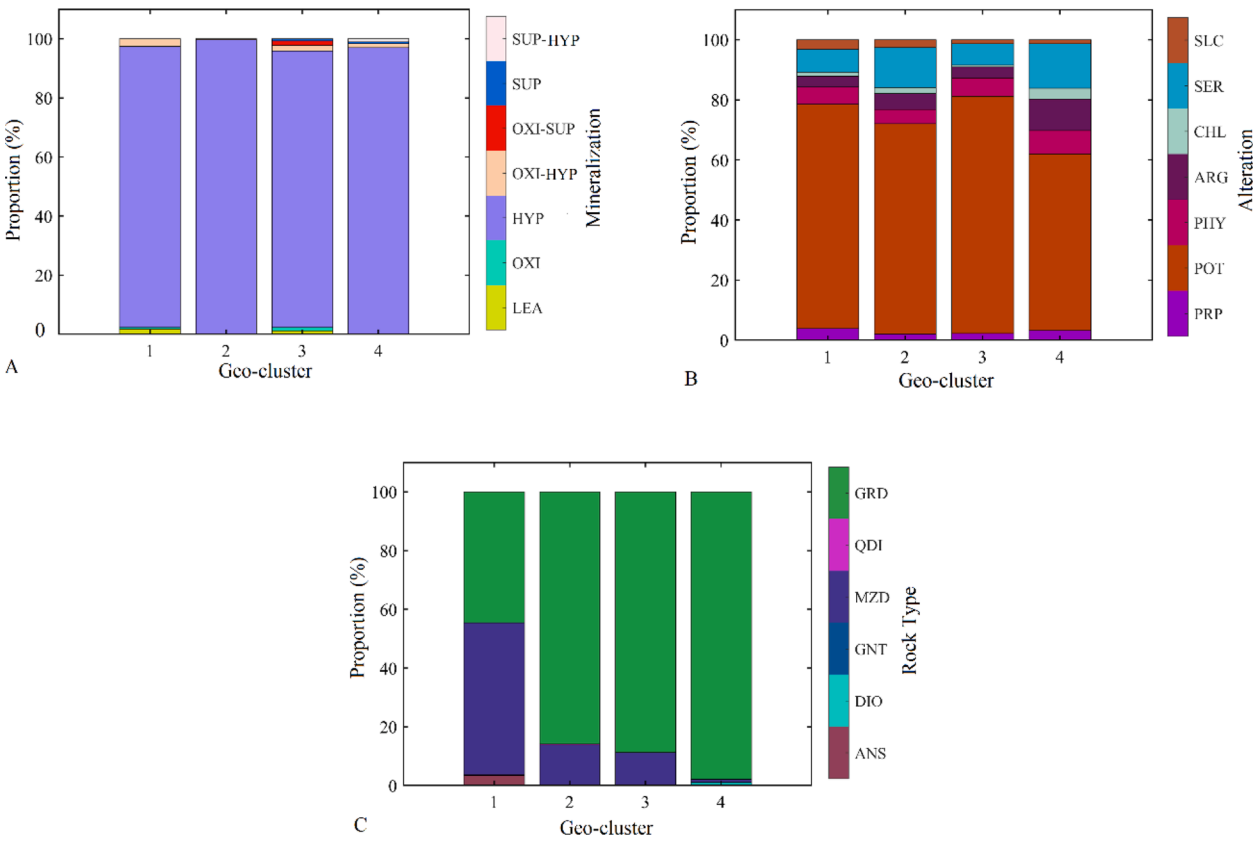

Fig. 6. Proportional stacked bar chart illustrating the association between mineralization, alteration, rock type and geo-clusters. The codes are defined in the text.

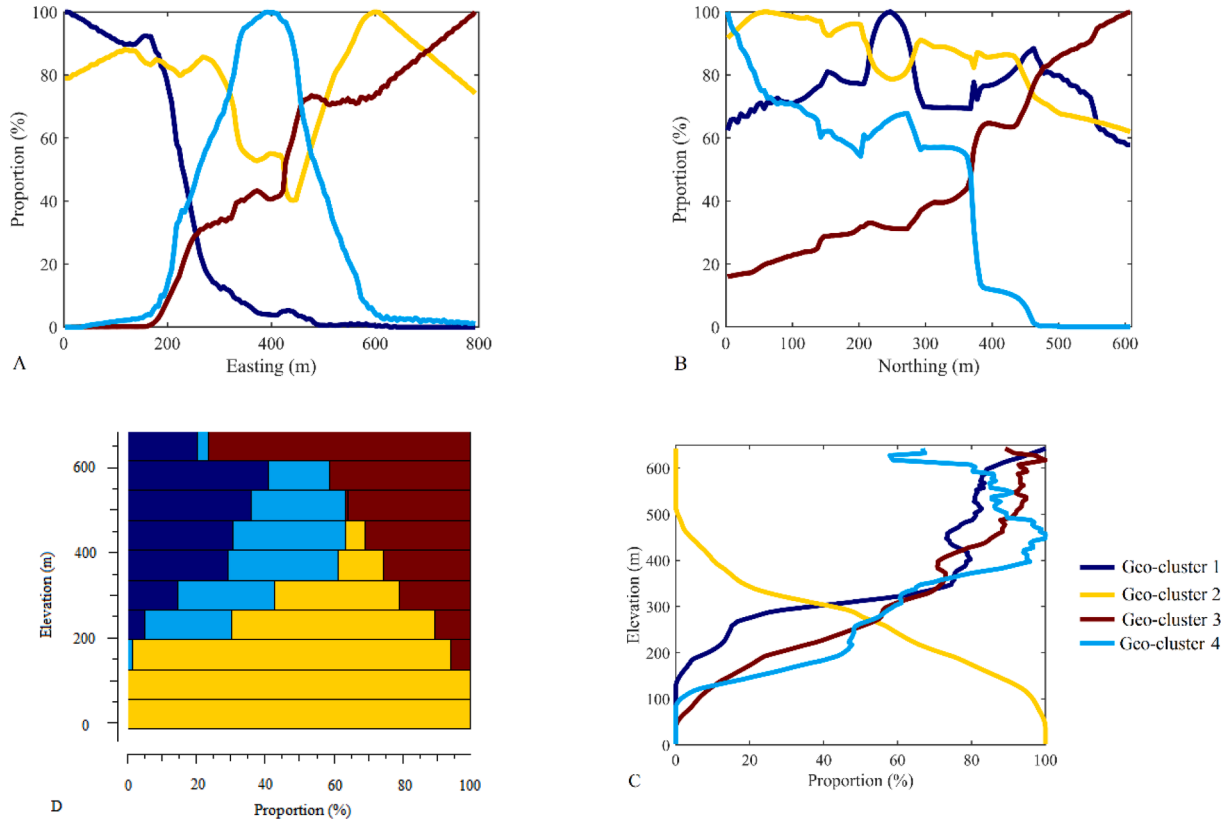

Fig. 7. Trend analysis of proportion of each geo-cluster against geographical coordinates (A-C) and vertical proportion curve (D).

Table 3
Statistical parameters for evaluation of the goodness of fitted multinomial logistic regression model.

	Standard error			
	ρ_p	ρ_{1p}	ρ_{2p}	ρ_{3p}
$p = 1$	0.395	6.93×10^{-4}	4.81×10^{-4}	5.50×10^{-4}
$p = 2$	0.364	4.23×10^{-4}	4.23×10^{-4}	3.98×10^{-4}
$p = 3$	0.392	4.56×10^{-4}	4.55×10^{-4}	5.11×10^{-4}
Mean of raw residuals	-0.23×10^{-14}	-0.34×10^{-14}	0.56×10^{-14}	0.02×10^{-14}
mean of Pearson residuals	0.120	−0.046	0.005	0.023
mean of deviance residuals	1.131			

sequential indicator simulation.

2.2. Sequential indicator simulation

A sequential indicator simulation (Alabert, 1987; Journel and Alabert, 1990) is a stochastic technique used to model the N categories. In this text, categories refer to the geo-clusters obtained from any unsupervised spatially dependent clustering techniques. These geo-clusters, if they are identified deterministically, are mutually exclusive at all the sample locations. However, there are some clustering techniques that produce a probabilistic allocation of the sample points to the clusters sought, but these techniques are not within the scope of this study.

Once the sample points are assigned to each geo-cluster (deterministically), to stochastically model them at unsampled locations, i.e., target grid nodes using conventional sequential indicator simulation, first, the geo-clusters (hard conditioning data) are transformed into a matrix with N columns of hard indicator data:

$$\mathrm{Ind}(K;\mathrm{n}) = \begin{cases} 1, & \textit{if geo} - \textit{cluster n assigned to sample point } K \\ 0, & \textit{otherwise} \end{cases} \quad n = 1, \cdots, N \tag{1}$$

Then, a random path is determined to visit each node of the target grid only once. In the next step, a stationary simple kriging (SSK) paradigm is used to constitute the conditional probability of the occurrence of each geo-cluster n at the target grid node K^*:

$$Ind^*_{SSK}(K^*;n) = \sum_{\beta=1}^{\upsilon} \Lambda_\beta^{SSK}(K^*;n) Ind(K_\beta;n) + \left[1 - \sum_{\beta=1}^{\upsilon} \Lambda_\beta^{SSK}(K^*;n)\right] \times \mu_n \tag{2}$$

where data υ consist of the hard and previously simulated indicator geo-clusters that are trapped in the neighborhood; μ_n is the global declustered prior proportion of each geo-cluster n; $\Lambda_\beta^{SSK}(K^*;n)$ is the weights allocated to the indicator geo-cluster $Ind(K_\beta;n)$ at the $K_\beta(\beta = 1, \cdots, \upsilon)$ of this indicator geo-cluster. The weights Λ are obtained by solving a variance–covariance matrix for each K^*.

The order relation problem is anticipated once the conditional probability of the occurrence of all N geo-clusters are estimated using Eq. (2). This means that the estimated conditional probabilities are not always equal to one and some of the values might show negative values.

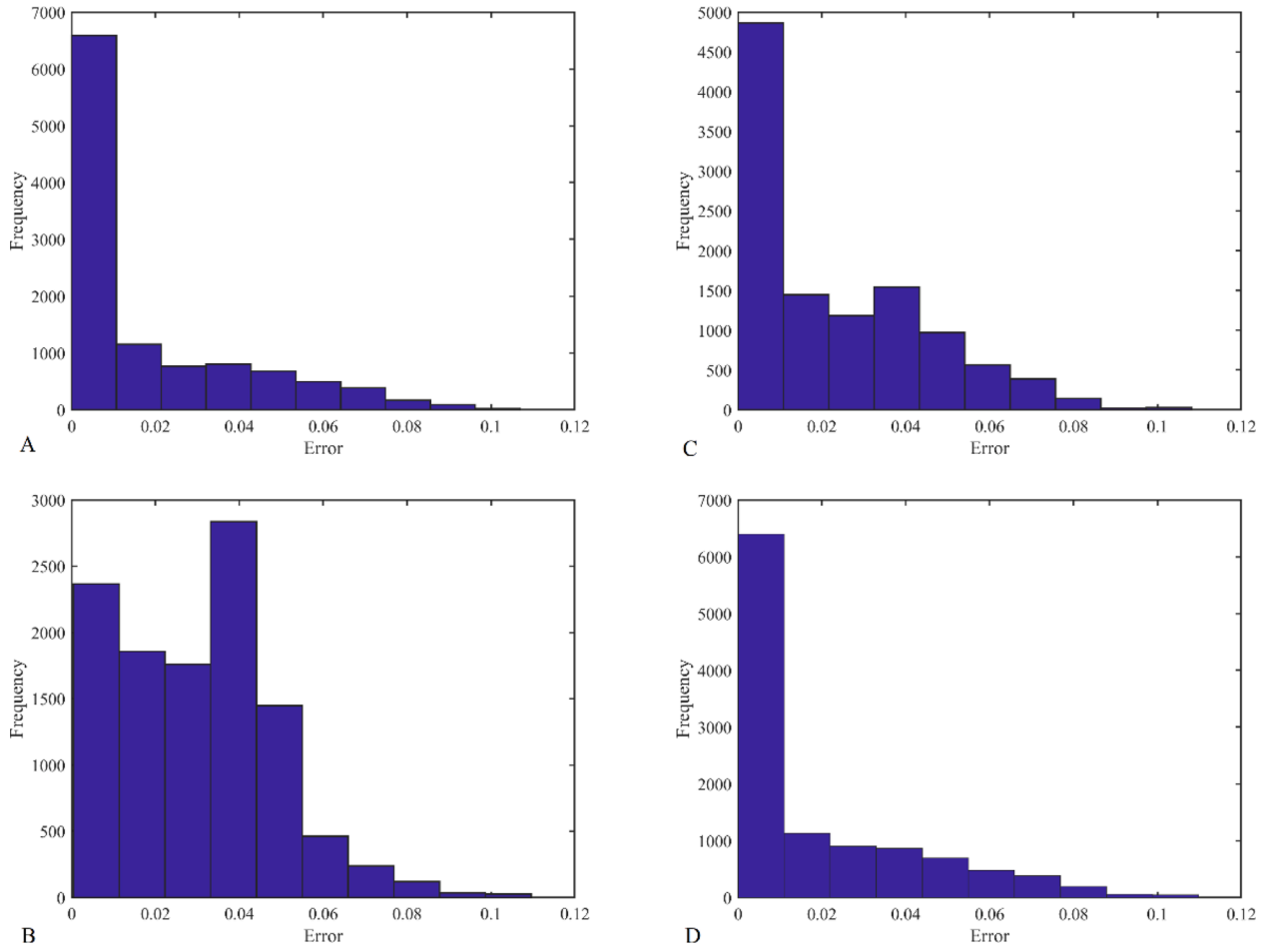

Fig. 8. Error over the confidence bounds calculated at each sampling point for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D).

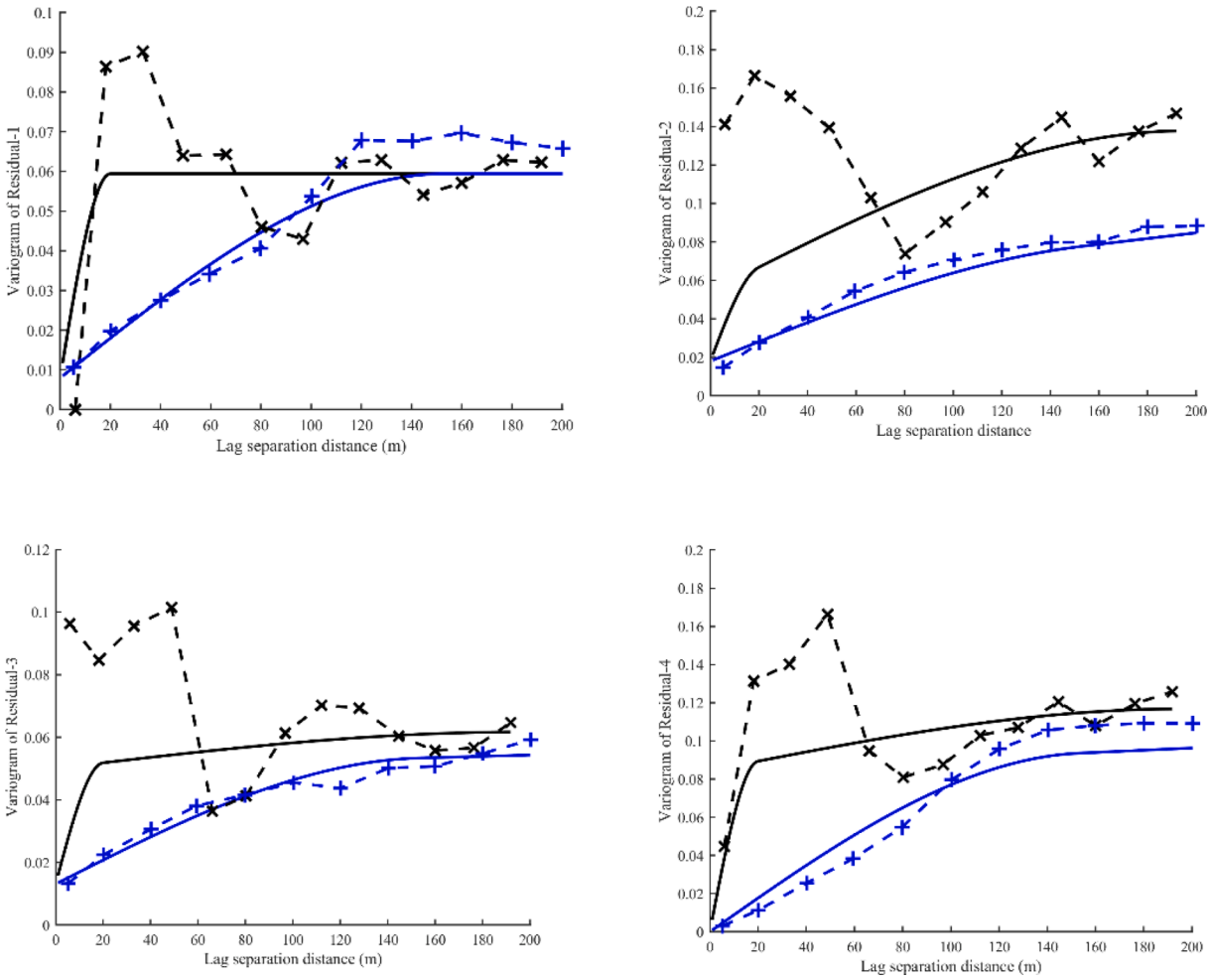

Fig. 9. Variogram analysis of residuals for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Blue points and line: vertical variogram, and black points and line: horizontal variogram. Dotted and solid lines are experimental and theoretical variogram, respectively.

Then, this deviation should be corrected. The last step involves simulating a geo-cluster at the target grid node using a random number drawn from a uniform distribution in [0, 1] by a Monte Carlo simulation. After these steps, the simulated value is added to the hard conditioning data and then the algorithm proceeds to the following target node, the order of which is already identified by a random path. To generate another realization at location K^*, one needs to repeat the previous steps to draw another random value from the conditional distribution, leading to the production of another simulated geo-cluster. However, this traditional paradigm of a sequential indicator simulation is suitable for strictly stationary scenarios.

In the following, the proposed methodology is discussed where the stationary simple kriging is replaced with non-stationary simple kriging in a conventional sequential indicator simulation, and an updated variant of this algorithm is given in conjunction with multinomial logistic regression.

2.3. Multinomial logistic regression

Multinomial logistic regression is a supervised classification machine learning algorithm that is a generalization of logistic regression to multiclass problems (Long, 1997; Long and Freese, 2006). In logistic regression, one deals with binary dependent variables, whereas in multinomial logistic regression, there are N (more than two) possible outcome for dependent variables. In this multiclass logistic regression algorithm, each sample point $\beta = 1 \cdots \tau$ (τ total number of observations) consists of a categorical variable Y with N possible outcomes where its variation depends on a set of K independent variables, which uses the multinomial logit model. In this model, to obtain the multinomial logit model for each geo-cluster n, one geo-cluster was selected as reference category, usually the last one N, and then the other $N-1$ geo-clusters were separately regressed against the reference geo-cluster. This can be formulated as follows if geo-cluster N is chosen as the reference category:

$$\begin{cases} Ln\dfrac{\mu(Y_\beta = 1)}{\mu(Y_\beta = N)} = \rho_1 + \rho_{11}K_1 + \rho_{12}K_2 + \cdots + \rho_{1p}K_p \\ Ln\dfrac{\mu(Y_\beta = 2)}{\mu(Y_\beta = N)} = \rho_2 + \rho_{21}K_1 + \rho_{22}K_2 + \cdots + \rho_{2p}K_p \\ \cdots \\ Ln\dfrac{\mu(Y_\beta = N-1)}{\mu(Y_\beta = N)} = \rho_{N-1} + \rho_{(N-1)1}K_1 + \rho_{(N-1)2}K_2 + \cdots + \rho_{(N-1)p}K_p \end{cases} \quad (3)$$

where $\mu(.)$ is the probability of the corresponding category; $p = 1, \cdots P$ is the number of independent variables and ρ_{np} is a regression coefficient associated with the *nth* category and the p independent variables. All of the $N-1$ equations are solved simultaneously to estimate the coefficients ρ_{np}. One possible solution is to use the maximum likelihood approach.

Based on the nominal dependent model, and the assumption that the coefficients of the last category are zero, then the probability of being in each category at each sample point β and the probability of the *Nth* category are:

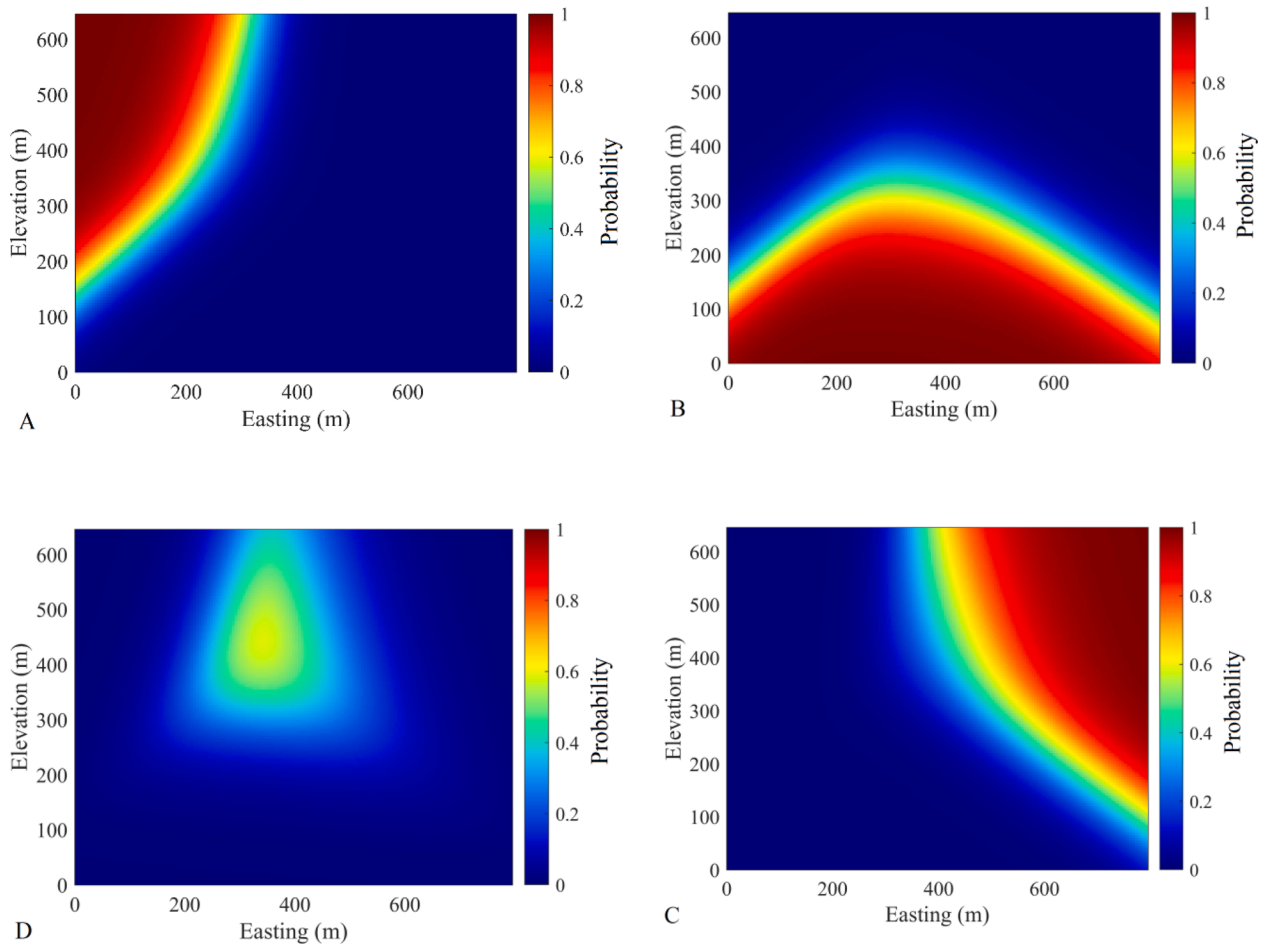

Fig. 10. Estimated probability maps for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D) obtained from multinomial logistic regression function.

$$\begin{cases} \mu(Y_\beta = n) = \dfrac{e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}{1 + \sum_{n=1}^{N-1} e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}, \\ \mu(Y_\beta = N) = \dfrac{1}{1 + \sum_{n=1}^{N-1} e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}, \end{cases} \quad n = 1, \cdots, N-1 \tag{4}$$

The probability of being in each category at each target grid node and the probability of the *Nth* category are:

$$\begin{cases} \mu(Y^* = n) = \dfrac{e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}{1 + \sum_{n=1}^{N-1} e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}, \\ \mu(Y^* = N) = \dfrac{1}{1 + \sum_{n=1}^{N-1} e^{\rho_n + \sum_{l=1}^{p} \rho_{nl} K_l}}, \end{cases} \quad n = 1, \cdots, N-1 \tag{5}$$

In order to evaluate the goodness of fit over the regression formula obtained from Eqs. (4) and (5), different statistical parameters related to the regression coefficients ρ_{np} and the probability of being in each category μ can be quantified. For the former, the standard error of the coefficient estimate ρ_{np} is the estimated standard deviation of the error in measuring it. The evaluation of these coefficients can also be accompanied with examination of the residuals. In this matter, the raw residual is the observed category minus the fitted category at the corresponding location; the Pearson residual is the raw residual scaled by the estimated standard deviation, and the Deviance residual can be computed as the square root of twice the difference between the loglikelihood of the *nth* observation in the actual data and the loglikelihood of the *nth* observation in the fitted model. For the latter, the error bounds on the predicated probabilities obtained from Eq. (5) can be calculated. This can be obtained by using the statistical parameters calculated over Eq. (4) to compute the lower and upper confidence bounds. The difference between these two values can return the error over the confidence bound. Lower values of these parameters (errors) in the regression model signify better estimated coefficients associated with lower errors.

2.4. Proposed simulation approach

In the case of geo-clusters with heterogeneous behavior, using stationary simple kriging in the conventional sequential indicator simulation is suboptimal. To circumvent this difficulty, an alternative can be used to replace the simple kriging paradigm in Eq. (2) by a non-stationary simple kriging (*N*-SSK), using the residuals from locally varying mean probabilities (Deutsch, 2006):

$$Ind^*_{N-SSK}(K^*; n) = \sum_{\beta=1}^{v} \Lambda_\beta^{N-SSK}(K^*; n)[Ind(K_\beta; n) - \mu(Y_\beta = n)] + \mu(Y^* = n) \tag{6}$$

with the same notation as above except $\mu(Y_\beta = n)$ and $\mu(Y^* = n)$, which are estimated probabilities at the conditioning data points and locally varying mean probabilities at the target grid nodes, respectively. The former helps calculate the residuals, and the latter is the trend component. The conditioning data points consist of hard and previously simulated indicator geo-clusters. In fact, the final estimated value $Ind^*_{N-SSK}(K^*; n)$ is obtained by adding the estimated residuals $Ind(K_\beta; n) - \mu(Y_\beta = n)$ to the trend component $\mu(Y^* = n)$. To obtain the

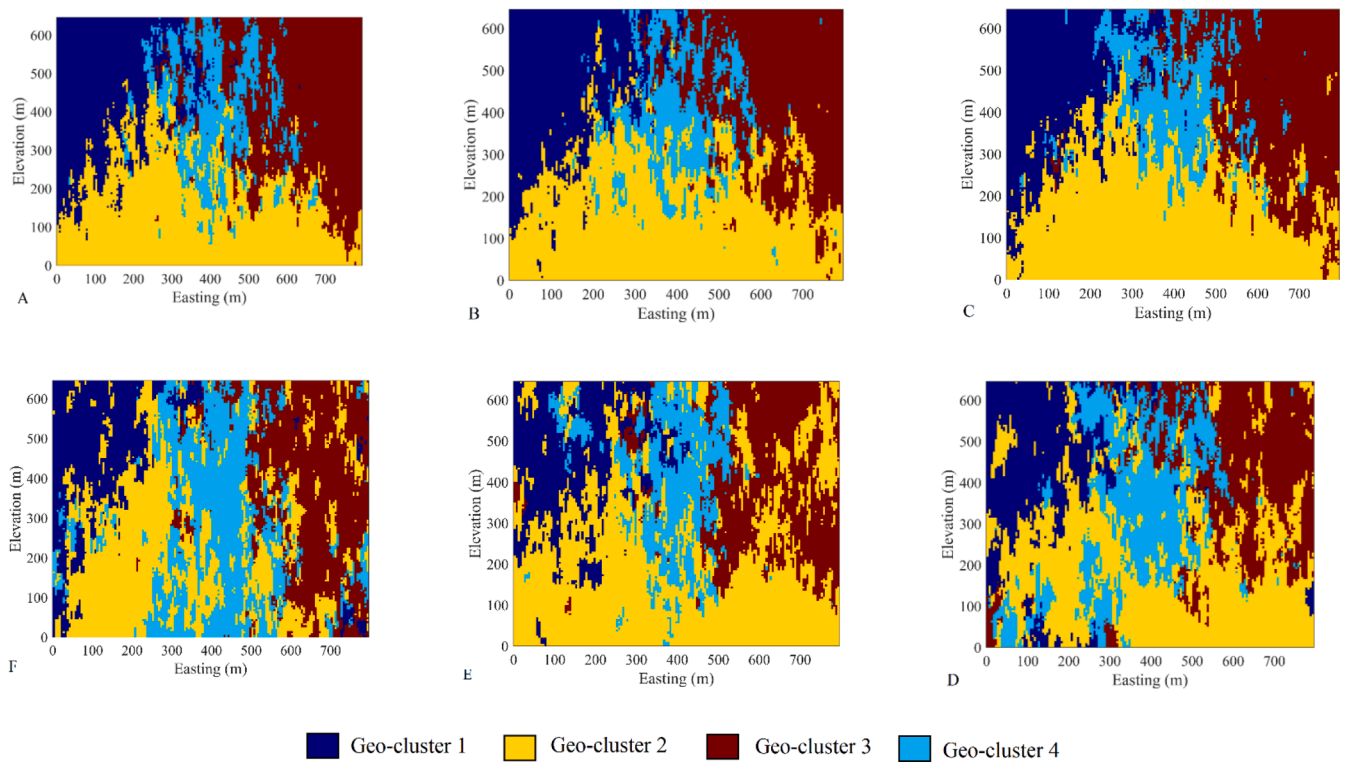

Fig. 11. Three random realizations obtained from SIS-N-SSK (A-B-C), SIS-SSK (D-E-F). Left: realization #17, middle: realization #39, and right: realization #71.

residuals for each geo-cluster $n = 1,..N$, a regression function must be fitted over the sample data points by using Eq. (3). Since the spatial variation of each category depends on the geographical coordinates of the sample points in heterogeneous geo-clusters, then a regression function can be fitted using the coordinates as independent variables and the geo-cluster as a dependent variable. Linear regression in this case is ineffective since the dependent variable includes categories that are characterized by integers.

Since $Ind(K_\beta; n)$ is known at the sample points, as an alternative, multinomial logistic regression can provide these estimated probabilities $\mu(Y_\beta = n)$ at these locations so as to calculate the corresponding residuals. In addition, by using the derived parameters of the multinomial regression function, one can computing the locally varying mean probability or trend component $\mu(Y^* = n)$ at the target grid nodes. To obtain such a model in our proposed study, the observations, dependent categorical variable, N possible outcomes, and K independent variables in Eqs. (4 and 5) refer to the sample points, geo-clusters obtained from unsupervised machine learning,N possible number of geo-clusters, and geographical coordinates, respectively.

In Eq. (6), since the simple kriging is based on the residuals over the conditioning data $Ind(K_\beta; n) - \mu(Y_\beta = n)$, a variogram analysis should be implemented over the residual values at the sample data locations for each n geo-cluster. The estimated value is then added to the locally varying mean probability or trend component of geo-cluster n. As long as the conditional probability $Ind^*_{N-SSK}(K^*; n)$ for each geo-cluster n is estimated at the target locations, the subsequent simulation steps of the geo-clusters will be similar to those of the conventional sequential indicator simulation.

The workflow for modelling non-stationary geo-clusters is as follows:

1- Identification of geo-clusters at the sample data locations using an unsupervised machine learning algorithm.
2- Trend analysis to check for trend identification of the categorical variable. Nearest neighborhood can be an option for this purpose.
3- Fit multinomial logistic regression models to each geo-cluster, derive the coefficients, and compute the estimated probabilities at the sample points.
4- Calculate the residuals over the sample points using the fitted multinomial logistic regression models and infer the variogram models of the residuals.
5- Calculate the locally varying mean probability over the target grid nodes for each geo-cluster to obtain the trend component using the fitted multinomial logistic regression models.
6- Implement the sequential indicator simulation using non-stationary simple kriging and generate the realizations.

A brief flowchart also is provided in Fig. 1.

In order to test the proposed algorithm, an actual porphyry copper deposit was selected, and the results were compared to the conventional sequential indicator simulation where a stationary simple kriging was used.

2.5. Case study

2.5.1. Geological setting

The Urumieh–Dokhtar magmatic arc (UDMA) of Iran, which is formed due to the subduction of the Neotethys Ocean underneath Eurasia (Berberian and King, 1981; Berberian and King, 1981; Berberian et al., 1982; Ghasemi and Talbot, 2006; Hosseini et al., 2017), stretches along the length of the Zagros Orogenic Belt and is made up of a linear magmatic belt. The main types of plutonic rocks within this zone are diorite, granodiorite, gabbro, and granitoids. The volcanic rocks are generally trachy-basaltic (nearly shoshonitic), andesitic, and dacitic in composition. Additionally, agglomerates, ignimbrites, and tuffs are other volcanic products (Alavi, 1994). The UDMA covers an area about 1700 km long × 4–50 km wide (StScklin, 1968). The arc contains most of the copper deposits of Iran (Fig. 2) (Boomeri et al., 2010; Aghazadeh et al., 2015).

The Sarkuh porphyry Cu ± Mo deposit is one of the copper deposits that is located in the southern part of the UDMA. It is located about 10

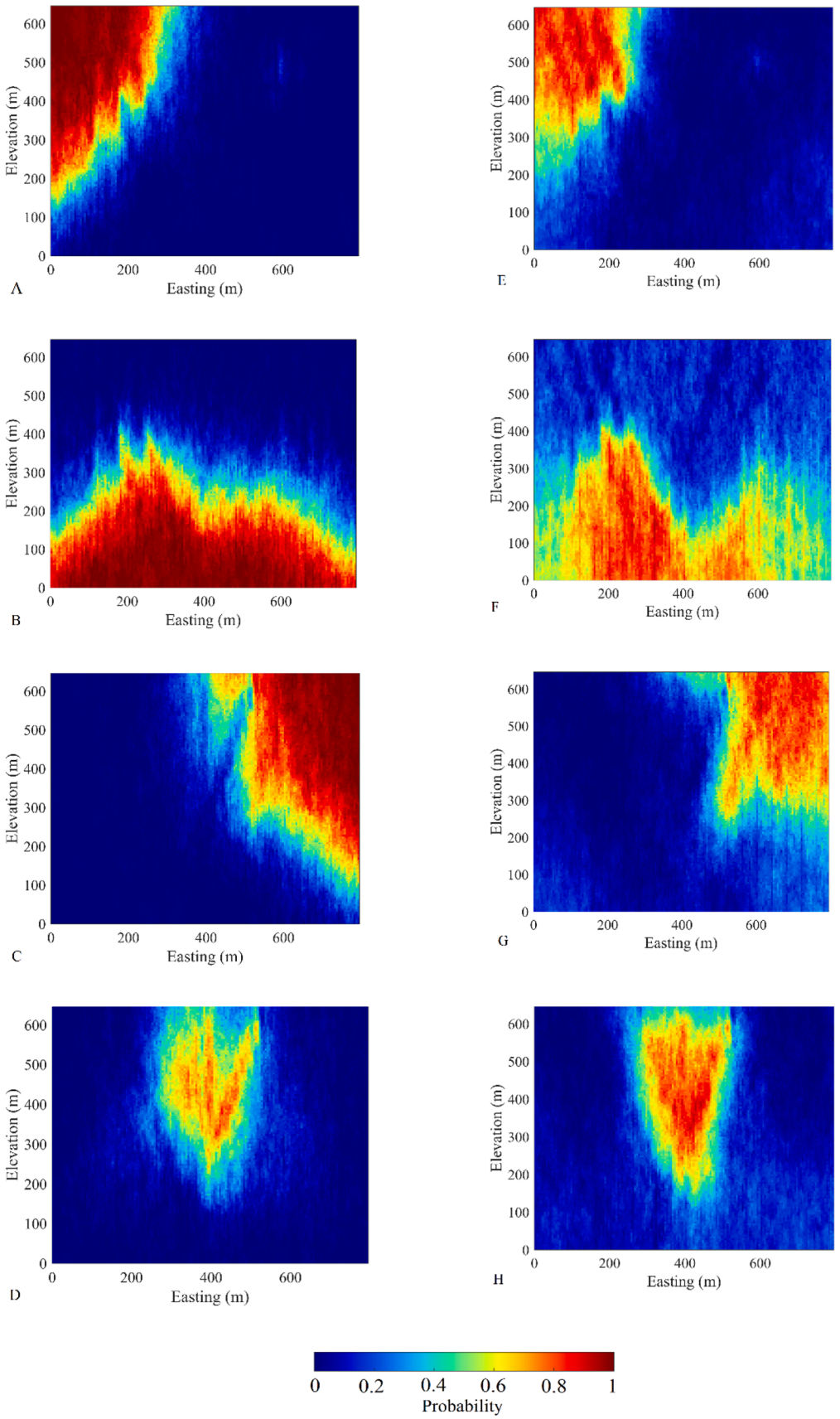

Fig. 12. Probability maps obtained with 100 realizations for SIS-N-SSK (A-B-C-D) and SIS-SSK (E-F-G-H); Geo-cluster1 (A & E), Geo-cluster 2 (B & F), Geo-cluster 3 (C & G), Geo-cluster 4 (D & H).

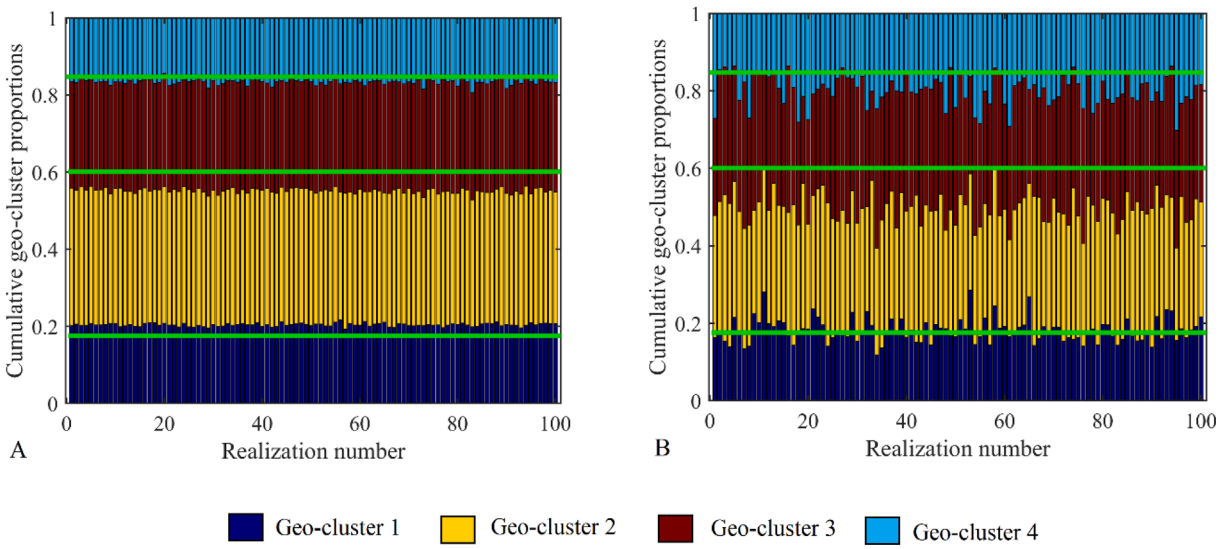

Fig. 13. Cumulative geo-cluster proportions over 100 realizations obtained with SIS-N-SSK (A) and SIS-SSK (B). Solid green lines indicate the declustered proportions calculated from the borehole (sample points).

Table 4
Error calculated by using indicator variograms of the original data and the resulting simulations.

	Geo-cluster 1		Geo-cluster 2		Geo-cluster 3		Geo-cluster 4	
	Horizontal	vertical	Horizontal	vertical	Horizontal	vertical	Horizontal	vertical
SIS-N-SSK	0.0006	−0.0162	−0.0246	0.0091	0.0265	0.0134	−0.0610	0.0195
SIS-SSK	0.006	0.0076	0.0552	0.0522	0.0298	0.0530	0.6690	0.0340

km northeast of Pariz city, Iran. A distance of six kilometers separates it from the largest copper deposit in Iran, the giant Sarcheshmeh deposit. The Sarkuh region is composed mostly of Eocene volcanic sedimentary assemblages, which are the oldest rock types in the region (Fig. 3). The volcanic units in the area mainly consist of tuffs, andesite, andesite–basalt, and pyroclastic breccias. There are a number of intrusive bodies that protrude in the SarkKuh deposit. The oldest intrusive body among them is the Oligocene Band-e-Mamzar pluton (Nourali and Mirnejad, 2012; Mirnejad et al., 2013; Aghazadeh et al., 2015). In the southern part of the deposit, an outcrop of pluton can be observed. It consists of diorite, tonalite, granodiorite, and monzogranite rock types (Nazarinia et al., 2019). Shallow-level intrusive bodies also intrude into the older rocks. These intrusive bodies played the most important role in the mineralization and formation of alteration zones (Malekshahi et al., 2018). Their outcrops can be observed in the central part of the deposit (Aghazadeh et al., 2015) and consist of granodiorite (SarkKuh porphyry) and monzodiorite porphyry stocks and granite, monzodiorite, and granodiorite porphyry dykes. Among all the above-mentioned intrusive rocks, the SarkKuh porphyry is the most widespread and is primarily responsible for mineralization (Malekshahi et al., 2018). Additionally, several dioritic, granitic, and granodioritic post-mineralization dykes cut into older intrusive and volcanic units (Aghazadeh et al., 2015). In the Sarkuh deposit, five different types of veins can be distinguished based on their mineralogy and their crosscutting relationships. They can be categorized into three classes: (1) pre-mineralization barren veins including quartz + K-feldspar and quartz + biotite veins, (2) mineralized veins including quartz + biotite + chlorite ± magnetite ± pyrite ± chalcopyrite quartz + pyrite + chalcopyrite ± bornite + chalcocite-sericite, and (3) post-Mineralization barren quartz ± calcite veins (Zarasvandi et al., 2019).

2.6. Presentation of the dataset

The dataset belongs to the SarkKuh copper porphyry drilling campaign consisting of 50 boreholes distributed in a semi-regular sampling pattern covering an approximate volume of 283 km^3. The borehole assaying reported Cu (ppm), Mo (ppm), Au (ppb), and Rock Quality Designation (RQD) (%) as the main continuous variables, for which Cu and Mo act as co-products and Au acts as a by-product in this deposit. RQD is a measure of the quality of rock core taken from a borehole. RQD signifies the degree of jointing or fracture in a rock mass measured in percentage, where an RQD of 75 % or more shows good-quality hard rock and an RQD of less than 50 % shows low-quality weathered rocks. Based on this definition, the rock mass can be classified into excellent, good, fair, poor, and very poor when the RQD is between 90 % and 100 %, 75 % and 90 %, 50 % and 75 %, 25 % and 50 %, 0 % and 25 %, respectively.

To preserve the confidentiality, the continuous variables were multiplied by a scale factor and the local coordinates were reported on the maps. This calc-alkaline porphyry copper deposit is characterized by three categorical variables (recognized by core logging), composed of seven mineralization zones: leached (LEA), oxidized (OXI), hypogene (HYPO), oxidized hypogene (OXI-HYP), oxidized supergene (OXI-SUP), supergene (SUP), and supergene hypogene (SUP-HYP); seven alteration types: propylitic (PRP), potassic (POT), phyllic (PHY), argillic (ARG), chlorotic (CHL), sericitic (SER), and siliciclastic (SLC); and six rock types: andesite (ANS), diorite (DIO), granite (GNT), monzodiorite (MZD), quartz diorite (QDI), and granodiorite (GRD). Fig. 4A-G show the location maps of the ore grades, the RQD, and the geo-domain variables. From the geological perspective, the mineralization in this deposit is mainly associated with GRD and the POT zone. In other words, HYP, POT, and GRD are the dominant mineralization, alteration zone, and rock types in this deposit.

It can be observed that Cu is mostly concentrated in the center of the deposit and is abundant in shallow elevations, but it tends to decrease with the increasing depth of this deposit. The spatial behavior of Mo, however, is somewhat different. It shows a high variance with an irregular disseminated pattern in the entire deposit. Au is undersampled

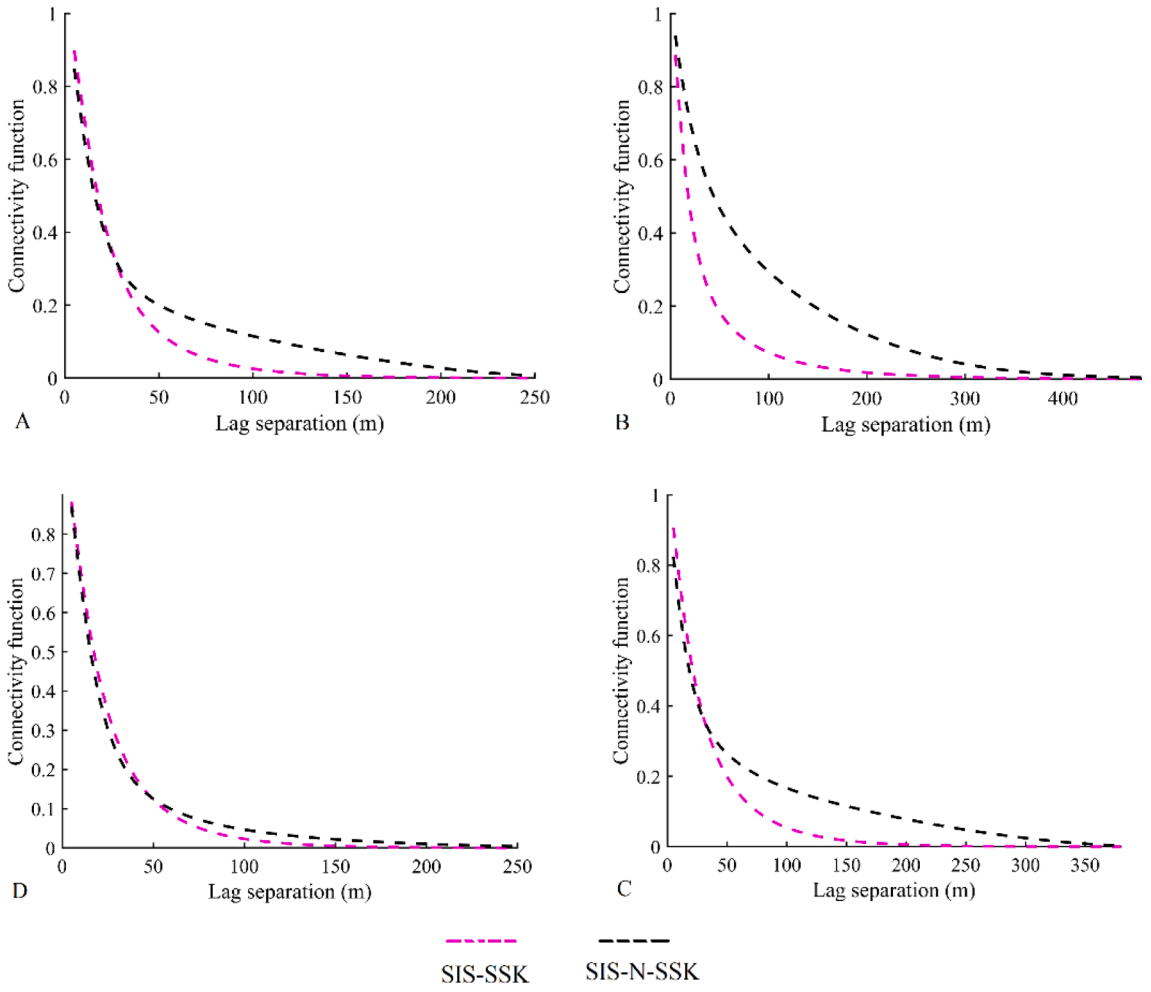

Fig. 14. Connectivity measures as a function of lag separation along vertical direction for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Dashed lines: average of connectivity functions obtained with 100 realizations.

as compared with other variables, with only 590 sample points assayed for this element. The Au grade is measured only through a few boreholes and, as can be seen, it is abundant in the west part of the deposit, and its variation is in agreement with the spatial distribution of Cu. The RQD is high in the center and increases by depth. This means that the rocks close to surface (mostly in the corners) are poorer compared to those at a large depth.

In order to quantitatively investigate the relationship among the continuous variables, the correlation coefficient was calculated. For categorical data, however, the linear correlation coefficient is impractical. An alternative is to use Cramer's V coefficient (Cramér, 2016), which is a measure of the association between such discrete variables, interpreted as a measure of dependency. This coefficient can be obtained as follows (Cramér, 2016):

$$V = \sqrt{\frac{\chi^2}{n(q-1)}} \tag{7}$$

where χ^2 is the Chi-squared test statistic computed from the contingency table; q is the smaller number of rows and columns in the contingency table; and n is the total number of sample locations. This coefficient ranges from 0 (poor association) to 1 (perfect association): no association ($V < 0.05$), weak association ($0.05 \leq V < 0.10$), moderate association ($0.10 \leq V < 0.15$), strong association ($0.15 \leq V < 0.25$), and very strong association ($0.25 \leq V < 1$). To infer the level of the association between continuous and categorical variables in this deposit, the continuous variables are converted into categorical data using the thresholds equivalent to their quartiles. This allows one to use Cramer's V coefficient to identify the potential strength of the association between categorical-converted continuous and other categorical variables. Table 1 reports the levels of the interdependencies and associations.

As can be seen, Cu is strongly correlated with Au, and there is a moderate correlation between Cu and Mo. Mineralizations are associated very strongly with alteration but weakly with rock type. The reason is that the HYP zone is highly associated with POT alteration and occurs as quartz-sulfide veinlets, dissemination, and stockwork. Alteration is moderately associated with rock type. It also can be observed that Cu, Mo, and RQD are associated strongly with rock type, while Au does not show any significant association with these three geo-domains. Therefore, following a common practice in resource modelling of copper deposits, rock type appears to be a remarkable variable to be identified as the estimation domains in this deposit to model the continuous variables inside them. However, the problem in this technique is that one may ignore the influence of mineralization zones and alteration in definition of the target estimation domains. To circumvent this problem, machine learning-based algorithms can be used to determine such domains, which are resulted by incorporating more than one variable. To do so, a machine learning clustering-based approach was used in this study to determine the estimation domains, and it not only considers the rock type but also takes into account the influence of Cu, Au, Mo, RQD, mineralizations, and alteration zones simultaneously. Three ore grades were important because they directly impact the mine plan for this deposit. The RQD was also included so that the rock quality of the geo-clusters could be evaluated. This is important for mine planning when

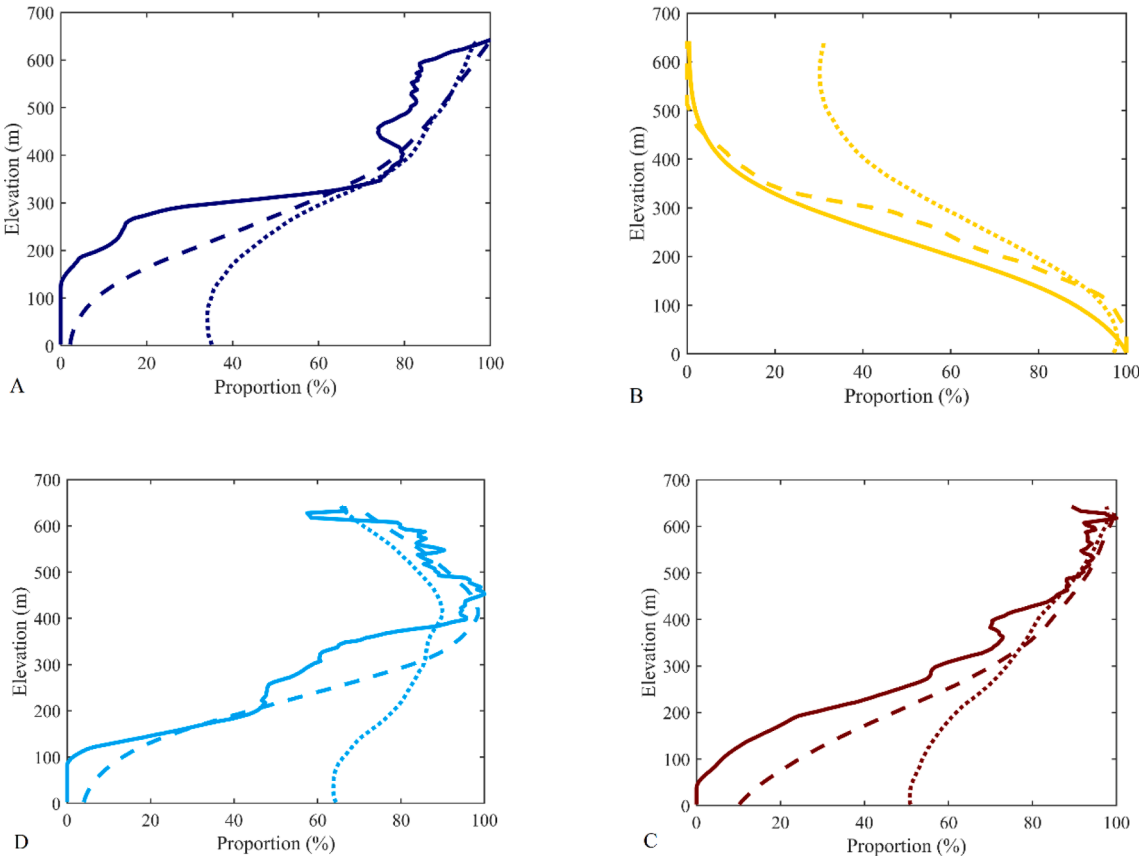

Fig. 15. Trend analysis reproduction along elevation over the simulation results for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Solid line: original trend; dashed line: average of trends over 100 realizations obtained with SIS-N-SSK; and dotted line: average of trends over 100 realizations obtained with SIS-SSK.

dealing with bench design in open-cast mining or stope stability analyses in underground mining activities. In addition, this approach can provide indirect insights regarding the quality of the rock for crushing and grinding to optimize the mineral processing plants. In practice, the RQD can be considered as indirect geo-metallurgical parameter.

The geostatistical hierarchical clustering algorithm method is available from Isatis.neo, and it was also used to obtain the target geo-clusters in this study. The Cu, Mo, Au, RQD, mineralization, alteration, and rock types associated with their geographical coordinates were then inputted into this algorithm. A different number of clusters were tested, and, finally, four geo-clusters were obtained, as shown in Fig. 4H. The reason for choosing four as the optimum number of clusters for this dataset was because this number allowed us to better partition the deposit into sub-domains with better interpretability of the continuous variables inside each, which is applicable for better mine planning. The associations between the resulting geo-clusters and all seven continuous and categorical variables showed that there were very strong, strong, medium, and weak associations between the RQD and rock type; Cu; mineralization, alteration, and Mo; and Au geo-clusters, respectively (Table 1). Fig. 5 shows the boxplot of the continuous variables and the resulting geo-clusters, where the statistical parameters of each continuous variable are reported (Table 2) and can help interpret the obtained domains. A stacked bar chart (Fig. 6) is also presented showing the coverage of each geo-cluster by each mineralization zone, alteration, and rock type. The geo-clusters can be interpreted as follows:

- Geo-cluster 1: This cluster is located in the north-west of the deposit, has a relatively high concentration of Cu, a medium concentration of Mo, and a high concentration of Au, and mostly includes rocks of poor quality based on an average RQD of 37.08 %. Most of this cluster is dominated by GRD and MZD, with POT and SER being the main rock types and alteration types, respectively, associated with the HYP zone. Therefore, this cluster can be targeted for high extraction of Cu and Au, which are contained in soft and poor rocks.
- Geo-cluster 2: This cluster covers the bottom of the deposit, where it shows high RQD values, signifying the presence of rocks with good and excellent qualities. This domain possess the lowest concentration of Cu and a high concentration of Mo and Au. This domain is covered mostly by GRD, and POT–SER–PHY as the main rock types and alteration zones, associated fully with the HYP zone. Therefore, this cluster can be a significant source of Mo and Au, which are mainly embedded in rocks with strong and good qualities.
- Geo-cluster 3: This cluster offers a medium concentration of Cu and Mo and a low concentration of Au. Most of the rocks in this domain have poor qualities, which dominate in the north-east part of the region and, to a small extent, in the center of the deposit. GRD and POT are the main rock types and alterations in this group and are mostly trapped in the HYP zone. This cluster possess the maximum coverage of POT. Therefore, this domain, while possessing rocks with poor qualities, can be investigated as a target for extracting Cu and Mo.
- Geo-cluster 4: This cluster is located in the center of the deposit somewhat close to the surface, where there is a very high concentration of Cu, a relatively high concentration of Mo, and a medium concentration of Au with fair rock qualities. This cluster is highly dominated by GRD and POT as the main rock types and alterations and is associated with the HYP zone. This cluster indicates the maximum coverage of GRD and the least coverage of POT. The characteristics of this cluster are most likely related to the secondary-enriched zone formed by oxidation of initial sulfide minerals, which

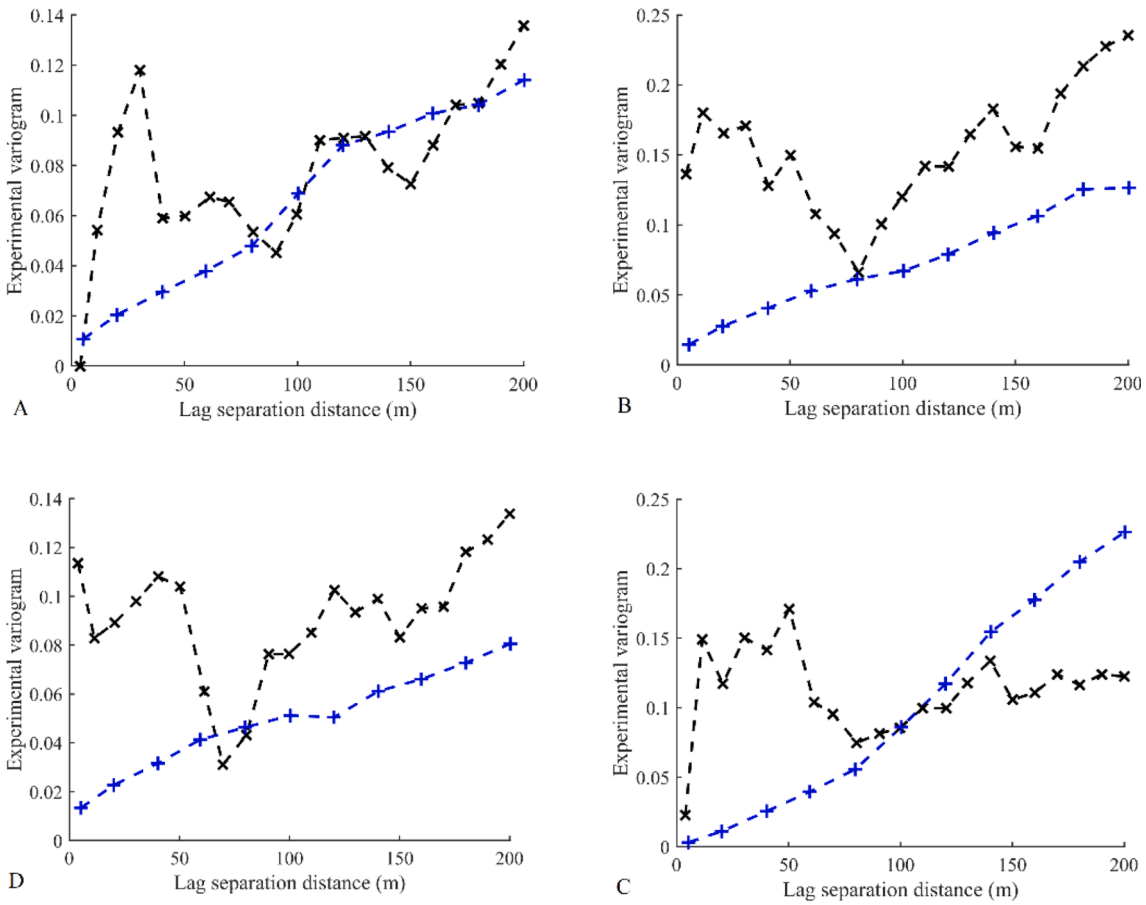

Fig. 16. Experimental variogram analysis of indicators for geo-cluster 1(A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Blue dashed line: vertical variogram, and black dashed line: horizontal variogram.

is commonly seen in porphyry copper deposits. Therefore, this cluster can be a significant source of Cu and Mo (Au) with fair rock qualities.

As can be seen from Fig. 4H, the spatial distribution of the geo-clusters is likely to be non-stationary. This is clear when inspecting the approximate locations of each geo-cluster in this map. As already mentioned, geo-cluster 1, geo-cluster 2, geo-cluster 3, and geo-cluster 4 approximately cover the north-west, bottom, north-east, and center of the deposit, respectively. In order to examine this concept quantitatively, a trend analysis was implemented. To do so, first, the deposit was discretized into the rectangular blocks with a mesh size of 5m × 5m × 5m to construct 2,558,400 blocks in total for the entire deposit. Second, the geo-clusters were allocated to the nearest neighbor block in order to establish a deterministic visualization of the geo-clusters in this deposit. The obtained categorical model was then transformed into the indicators. A lag of 5 m was selected to calculate the proportion of each indicator along the easting, northing, and elevation (Fig. 7A-C). The proportion of geo-cluster 1 was found to be high in the west and close to the surface, but substantially decreased toward the east and at depth, while along the northing, it was almost homogenous. Geo-cluster 2 showed a sharp trend along the elevation, implying that this geo-cluster is only present at higher depths and it is quite unlikely for this domain to be displayed close to the surface, while not much significant non-stationary trend variability is expected along the easting. Geo-cluster 3 seems to be the most heterogeneous geo-domain in this deposit according to the sharp trend analysis along all three directions. Geo-cluster 4 also shows a very sharp trend along the northing and elevation, meaning that the presence of this domain at higher depths and in the north of the region is unexpected, and it mostly dominated the center and upper part of the deposit. A vertical proportion curve (Matheron et al. 1987) was also computed. This simple tool is suitable for quantifying the variability of geo-clusters based on the proportion of each geo-cluster as a function of depth. Fig. 7D shows the proportion of each geo-cluster at each level, where it reflects the progression obtained from the clustering approach along elevation. By linking this vertical progression to the interpretation of geo-clusters (above), one may deduce that the concentrations of Cu (main source: geo-clusters 1 and 4), Mo (main source: geo-clusters 2 and 4), and Au (main source: geo-clusters 1 and 2) decrease, increase, and increase at higher depths, respectively. The same trend was observed for rock strength (main source: geo-cluster 2) in that its quality increased with the increasing depth, changing from poor and fair rocks in the surface into good and excellent rocks at higher depths. In brief, the quality of the rocks improves with increasing depth.

3. Results

3.1. Geostatistical modelling of geo-clusters

The method proposed in this study, a non-stationary sequential indicator simulation using the residuals from locally varying mean probabilities, was then applied to stochastically model the geo-clusters in the entire copper deposit using the same block model identified earlier for nearest neighborhood prediction. Following the steps of our proposed approach, once the geo-clusters are identified at the sample points (boreholes) using the geostatistical hierarchical clustering method (Fig. 4H), as extensively explained, the next step is to fit multinomial logistic models over these points for each geo-cluster and derive the coefficients.

In this model, there is one dependent variable (i.e., four geo-clusters)

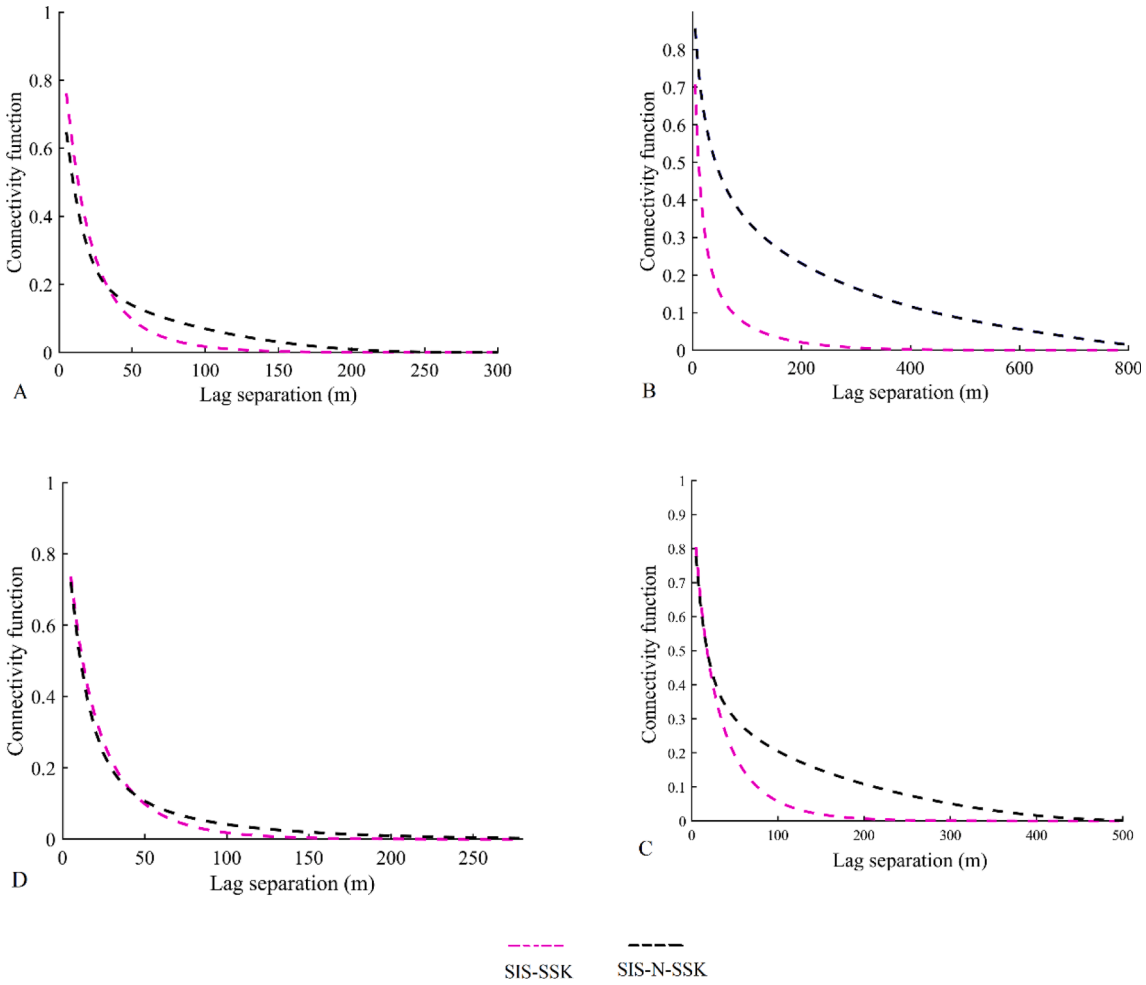

Fig. 17a. Connectivity measures as a function of lag separation along easting direction for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Dashed lines: average of connectivity functions obained with 100 realizations.

and three independent variables as geographical coordinates. Generally speaking, we are interested in inferring multinomial logistic regression models to predict the geo-clusters Y as the function of geographical coordinates K. To do so, the last geo-cluster (geo-cluster 4) was selected as the reference category, and the other geo-clusters were separately regressed against geo-cluster 4. The sub-routine "mnrfit" in MATLAB R2021b was used to derive these functions:

$$\begin{cases} Ln\dfrac{\mu(Y_\beta=1)}{\mu(Y_\beta=4)} = 14.690 - 0.035K_1 - 0.013K_2 + 0.003K_3 \\ Ln\dfrac{\mu(Y_\beta=2)}{\mu(Y_\beta=4)} = 12.998 - 0.013K_1 - 0.017K_2 - 0.004K_3 \\ ln\dfrac{\mu(Y_\beta=3)}{\mu(Y_\beta=4)} = 17.936 - 0.051K_1 - 0.012K_2 - 0.021K_3 \end{cases} \quad (8)$$

where $\beta = 1, \cdots \tau$ (β is the number of sample points), and K_1, K_2, and K_3 correspond to the geographical coordinates of the sample points: easting, northing, and elevation, respectively. This fitting procedure assumes that the coefficients of geo-cluster 4 (acting as the reference category) are all zero. To infer the coefficients, "mnrfit" uses the iteratively weighted least squares algorithm to find the maximum likelihood estimates. In order to evaluate the goodness of fit, standard errors of the regression coefficient ρ, the mean of the raw residuals, the mean of the Pearson residuals, and the mean of the deviance residuals were calculated and are reported in Table 3. As can be seen, they were comparatively small and reasonable for the fitted model.

The next step is to estimate the probability of each geo-cluster at sample points $Y_\beta, \beta = 1, \cdots \tau$:

$$\begin{cases} \mu(Y_\beta=1) = \dfrac{14.690 - 0.035K_1 - 0.013K_2 + 0.003K_3}{1 + 45.625 - 0.064K_1 - 0.043K_2 - 0.022K_3} \\ \mu(Y_\beta=2) = \dfrac{12.998 - 0.013K_1 - 0.017K_2 - 0.004K_3}{1 + 45.625 - 0.064K_1 - 0.043K_2 - 0.022K_3} \\ \mu(Y_\beta=3) = \dfrac{17.936 - 0.051K_1 - 0.012K_2 - 0.021K_3}{1 + 45.625 - 0.064K_1 - 0.043K_2 - 0.022K_3} \\ \mu(Y_\beta=4) = \dfrac{1}{1 + 45.625 - 0.064K_1 - 0.043K_2 - 0.022K_3} \end{cases} \quad (9)$$

The coefficients in the model express the effects of the geo-clusters on the relative risk or the log odds of being in geo-cluster 1, 2, and 3 versus geo-cluster 4. For instance, the coefficient 0.003 indicates that the probability of the geo-cluster being in category 1, as compared to the probability of being in category 4, increases $e^{0.003}$ times for each unit increase along the elevation. The same interpretation is used for the other directions and parameters.

To assess the precision of the obtained probabilities deduced from

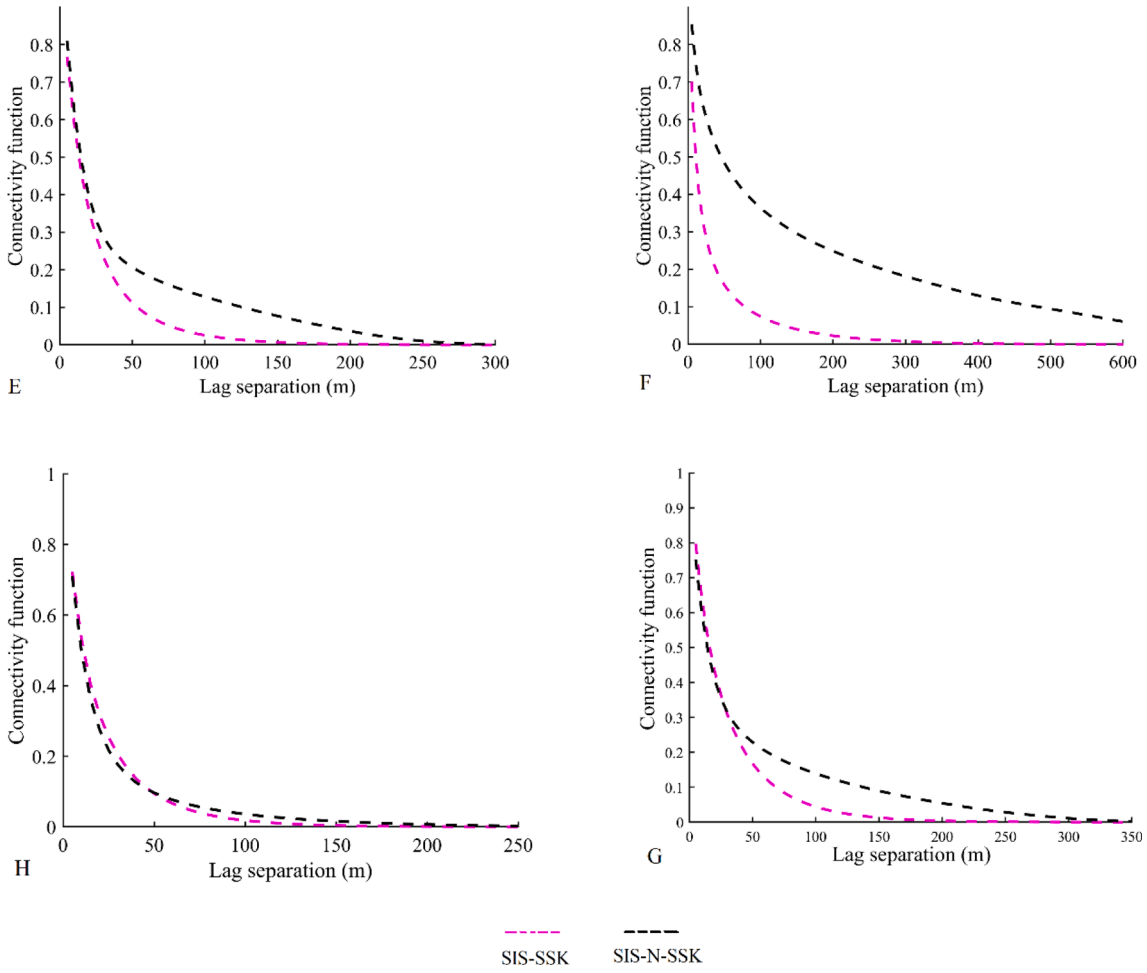

Fig. 17b. Connectivity measures as a function of lag separation along northing direction for geo-cluster 1 (E), geo-cluster 2 (F), geo-cluster 3 (G), and geo-cluster 4 (H). Dashed lines: average of connectivity functions obtained with 100 realizations.

Eq. (9), a histogram of the error confidence bound for each geo-cluster was calculated (Fig. 8). As can be observed, the errors were low and mostly frequent around zero. Small values of these statistical parameters signify that the regression formulas of the proposed multinomial logistic regression model are reliable and can be utilized in further analyses.

The estimated probabilities at the sample points enable one to calculate the residuals $Ind(K_\beta;n) - \mu(Y_\beta = n)$. For this purpose, the geo-clusters were converted into the indicators $Ind(K_\beta;n)$ using Eq. (1), then they were subtracted from the estimated probabilities $\mu(Y_\beta = n)$. Since the input (hard conditioning data) to the proposed algorithm should be the residuals sought, the variogram analysis needs to be inferred. After quantifying the anisotropy of each geo-cluster in the region, we determined two directions of anisotropies over the horizontal and vertical directions (Appendix, Fig. 16). Therefore, spherical variogram models with a proper nugget effect considering the maximum and minimum continuities along the vertical and horizontal directions, respectively, were fitted to the experimental variograms of the residuals (Fig. 9):

$$\gamma_{Res-1} = 0.007nugget + 0.051Sph(153\text{m}, 20\text{m}, 20\text{m})$$

$$\begin{aligned}\gamma_{Res-2} &= 0.017nugget + 0.036Sph(153\text{m}, 20\text{m}, 20\text{m}) + 0.083Sph(801\text{m}, 801\text{m}, 195\text{m})\\ \gamma_{Res-3} &= 0.084Sph(153\text{m}, 20\text{m}, 20\text{m}) + 0.032Sph(801\text{m}, 801\text{m}, 195\text{m})\end{aligned} \quad (10)$$

$$\gamma_{Res-4} = 0.013nugget + 0.036Sph(153\text{m}, 20\text{m}, 20\text{m}) + 0.011Sph(801\text{m}, 801\text{m}, 195\text{m})$$

As can be seen, the variograms of the residuals showed a finite sill, implying a stationary hypothesis for these variables. However, as can be observed, the horizontal variogram did not show a satisfactory structure as compared to the vertical variogram, particularly for geo-clusters 2 and 3. A possibility for this is the existence of many inclined boreholes that caused the experimental variogram in a horizontal direction to be

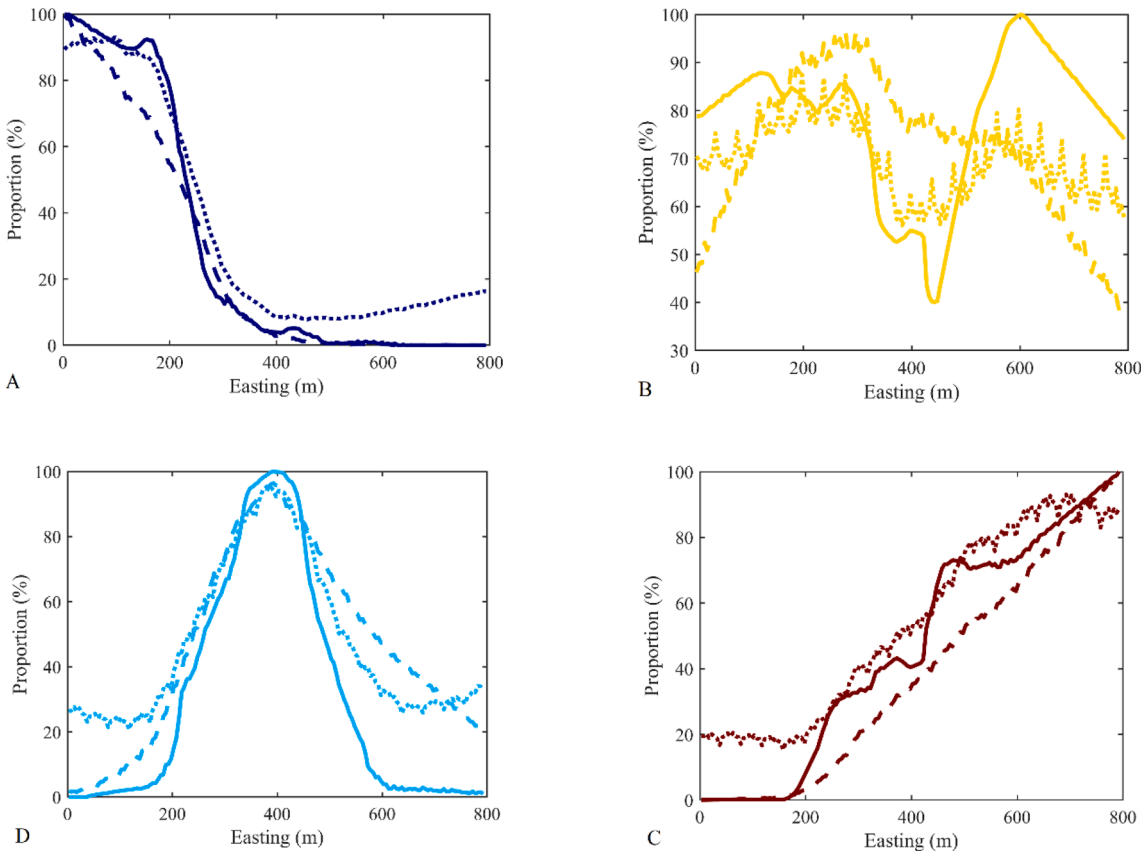

Fig. 18a. Trend analysis reproduction along easting over the simulation results for geo-cluster 1 (A), geo-cluster 2 (B), geo-cluster 3 (C), and geo-cluster 4 (D). Solid line: original trend; dashed line: average of trends over 100 realizations obtained with SIS-N-SSK; and dotted line: average of trends over 100 realizations obtained with SIS-SSK.

rather more irregular.

The next step is to compute the locally varying mean probabilities or trend components $\mu(Y^* = n) n = 1, \cdots 4$ that are obtained using the fitted multinomial regression function (Eq. (8)), but, this time, the independent variables are the geographical coordinates of the target grid nodes Y^*:

$$\begin{cases} \mu(Y^* = 1) = \dfrac{14.690 - 0.035K_1^* - 0.013K_2^* + 0.003K_3^*}{1 + 45.625 - 0.064K_1^* - 0.043K_2^* - 0.022K_3^*} \\ \mu(Y^* = 2) = \dfrac{12.998 - 0.013K_1^* - 0.017K_2^* - 0.004K_3^*}{1 + 45.625 - 0.064K_1^* - 0.043K_2^* - 0.022K_3^*} \\ \mu(Y^* = 3) = \dfrac{17.936 - 0.051K_1^* - 0.012K_2^* - 0.021K_3^*}{1 + 45.625 - 0.064K_1^* - 0.043K_2^* - 0.022K_3^*} \\ \mu(Y^* = 4) = \dfrac{1}{1 + 45.625 - 0.064K_1^* - 0.043K_2^* - 0.022K_3^*} \end{cases} \quad (11)$$

where K_1^*, K_2^*, and K_3^* correspond to the geographical coordinates of the target grid nodes: the easting, northing, and elevation, respectively. Therefore, the estimated trend component $\mu(Y^* = n)$ at each node can be used in Eq. (6). These maps are shown in Fig. 10. As can be observed, the estimated probable area of each geo-cluster was compatible with their spatial distribution over the borehole dataset as explained in Fig. 4H. Geo-clusters 1, 2, 3, and 4 had a high probability of being found in the west, bottom, east, and the upper-central part of the deposit, respectively. This is perfect secondary information that can be used as a trend component for simulations using the proposed approach by adding conditional probabilities of the residuals to the estimation to produce the final estimated geo-clusters based on Eq. (6).

The residuals were calculated over the sample points, and the local mean probability for each geo-cluster was calculated. These two, in addition to the derived model of the variograms for each residual, were inputted into the proposed sequential indicator simulation algorithm, and the results were compared with those of the conventional sequential indicator simulation. Therefore, two cases are compared hereafter:

1- **SIS-N-SSK**: The proposed sequential indicator simulation using non-stationary simple kriging integrated with multinomial logistic regression;
2- **SIS-SSK**: The conventional sequential indicator simulation using stationary simple kriging.

The modelling was implemented over the same constructed block model with the same size and number of blocks as discussed earlier. An identical moving neighborhood was also considered in both cases for original data and previously simulated grid nodes. This was an ellipsoid with maximum elongation along the vertical direction with 40 data in total. No octant was used, and a similar random path sequence was taken into account for both SIS-SSK and SIS-N-SSK to generate 100 realizations for each. Fig. 11 shows three random realizations of each method. As can be seen in this figure, the proposed approach, SIS-N-SSK, was able to reproduce the potential heterogeneity of the geo-clusters, while the traditional approach, SIS-SSK, failed to reproduce the desired compact and non-stationary characteristics of the geo-clusters. SIS-SSK produced patchy and unstructured results, as was expected. Therefore, the results of SIS-*N*-SIS are more compatible with the spatial distribution of geo-clusters as discussed in Fig. 4H.

The uncertainty in the geo-clusters' grid node by grid node can be computed using probability maps. These maps can be created by computing the proportion of each geo-cluster over the 100 realizations

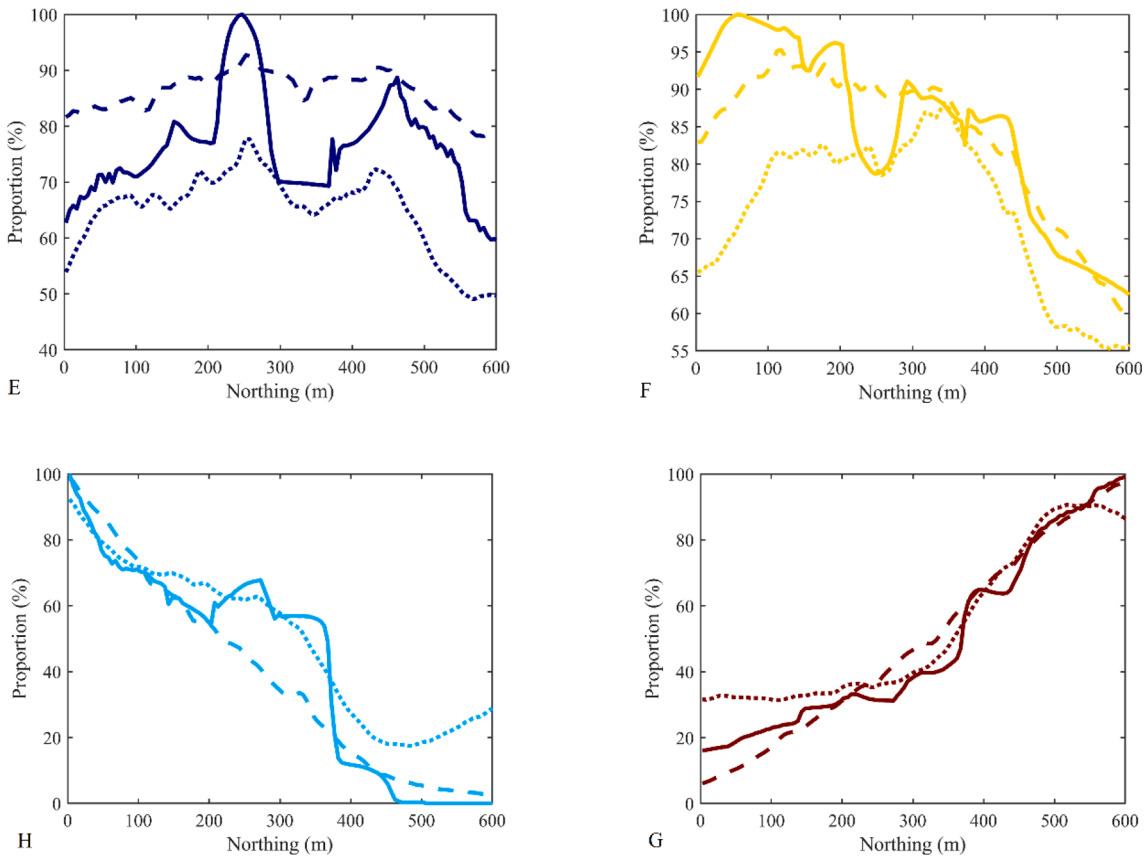

Fig. 18b. Trend analysis reproduction along northing over the simulation results for geo-cluster 1 (E), geo-cluster 2 (F), geo-cluster 3 (G), and geo-cluster 4 (H). Solid line: original trend; dashed line: average of trends over 100 realizations obtained with SIS-N-SSK; and dotted line: average of trends over 100 realizations obtained with SIS-SSK.

(Fig. 12). The areas showing low uncertainty are linked to a given geo-cluster painted in red, signifying that there is a small risk of not locating this geo-cluster, and the areas linked to a very low probability indicate that one can be sure of not being able to locate this geo-cluster, while other areas, painted in light blue, green, or yellow, are more uncertain. The SIS-N-SSK provides more reasonable results, which are more in agreement with the reproduction of trends that are strikingly similar to those in the conceptual model of the spatial distribution of geo-clusters in this deposit.

3.2. Statistical validation

In Fig. 13, the cumulative reproduction of the proportions of the geo-clusters are provided together with the declustered proportions calculated from the boreholes. The experimental proportions are equivalent to 0.170, 0.438, 0.259, and 0.132 for geo-clusters 1, 2, 3, and 4, respectively. The proportions of geo-clusters 1 and 2 are slightly underestimated in SIS-N-SSK, while in SIS-SSK, the underestimation of the proportion of geo-cluster 2 is more significant. Concerning geo-cluster 3, SIS-N-SSK outperformed SIS-SSK. In general, the reproduction of the proportion of geo-clusters is better in SIS-N-SSK than in SIS-SSK.

Another criterion by which to evaluate the performance of the proposed approach is to check the reproducibility of the original variograms for each geo-cluster. For this purpose, the indicator variogram was calculated over the indicators linked to each geo-cluster along the horizontal and vertical directions. The experimental indicator variograms of the indicators over the boreholes were compared with those obtained from individual realizations resulting from SIS-N-SSK and SIS-SSK. To do so, the average variograms were computed over the individual variograms of the realizations, and then an error was calculated at each lag separation distance by subtracting the indicator variogram of the corresponding average variogram from the indicator variogram of the original data. The errors are reported in Table 4. As can be seen, the reproduction of the original indicator variograms resulted in smaller errors for SIS-N-SSK in the majority of the geo-clusters.

It also might be of interest to examine the spatially contiguous characteristics of the resulting geo-clusters within the deposit. This is an important aspect because it significantly impacts the downstream activities of a mining practice from the perspective of exploitation. Indeed, the chosen approach of geostatistical modelling should be able to model the geo-clusters with better interconnectivity or better spatially contiguity. This leads to having more compact and contiguous simulated clusters at the target grid nodes, which are better suited to mine planning and exploitation. To test this, an indicator correlogram can be used (Moreira et al. 2020). In this study, the connectivity function (Renard and Allard, 2013), as a multiple-point statistic, was taken into account to quantify this as a measure of the probability that a grid node is connected with another grid node located at a lag separation distance (h) in such a way that both belong to the same geo-cluster. Therefore, the probability that any two grid nodes are connected can be quantified as a function of h. As an example, this is shown for the corresponding geo-clusters as the function of the separation of h along the vertical direction (Fig. 14). The probability of connectivity within the geo-clusters for the simulation results obtained with SIS-N-SSK was higher, particularly at a large h. This means that the clusters obtained using the proposed approach, SIS-N-SSK, produces more probable contiguous and compact domains. Connectivity measures along the east and north directions are provided in the Appendix (Figs. 17a, b). The reproducibility of ideal connectivity for geo-clusters 1, 2, and 3 was remarkably better in SIS-N-

SSK along all the directions. However, the difference in the reproduced connectivity for geo-cluster 4 was slightly better in SIS-N-SSK. In fact, this geo-cluster is not as contiguous and compact as the other geo-clusters. This is the reason why the difference between the two methods is small.

The next step is to check the reproduction of original trends for each geo-cluster. For this purpose, the trend is computed for each realization obtained with each method, and then their averages are plotted against the coordinates. As an example, Fig. 15 shows these trends when considering this reproducibility versus elevation. The trend reproduced by SIS-K-SSK was more in agreement with the original trend as compared to that reproduced by SIS-SSK. The trend analysis for the east and north directions is presented in the Appendix (Figs. 18a, b). In the majority of the trend analysis results, SIS-K-SSK was superior and it showed that with a stronger trend component, better reproducibility of the trend along the coordinate can be expected. The reason for SIS-K-SSK's better performance is connected to the incorporation of the trend component in the simulation algorithm that is informed by multinomial logistic regression.

4. Discussion

In this study, four geo-clusters were deduced using the geostatistical hierarchical clustering technique. From a geological perspective, the relatively high concentration of Cu in geo-cluster 1 is most likely characterized by micro quartz diorite porphyry and veinlets within quartz veins. The low grade of Cu in geo-cluster 2 is most likely related to the emplacement of the Mamzar stock that leads to the low dissemination of vein stockworks in this area. In addition, the presence of a PHY zone in this geo-cluster might be another reason for the low concentration of Cu since PHY is essentially not a significant alteration zone for ore bearing in this deposit. Geo-cluster 3 is mostly associated with the POT zone in north-east part of the region with a medium concentration of Cu and poor rock qualities. This geo-cluster should be the next target for further resource estimation. The SER, ARG, and PRP are three most important alterations in geo-cluster 4. The characteristics of this cluster are most likely related to the secondary-enriched zone formed by oxidation of initial sulfide minerals, which is commonly seen in porphyry copper deposits.

The geostatistical simulation method proposed in this study can also be applied to modelling any heterogeneous geo-domains, particularly those obtained from recently developed approaches resulting in the inference of geo-clusters. For instance, the geo-clusters modelled using this method can be incorporated for the purpose of resource estimation, mine planning, and geo-metallurgical domaining. The proper identification of domains and the use of suitable modelling approaches can help better determine the downstream analyses of a mining project, which can lead to the production of more achievable business plan for a mine. For instance, in resource estimation, these geo-clusters can be used to model the continuous variables (in this study, Cu, Mo, Au, and RQD) separately in each. Attention must be paid to considering the condition of continuous variable variation across the boundaries of two adjacent geo-clusters as a hard boundary. The estimated grades can be used to calculate the metal quantity in each part of the deposit where their RQD evaluations (one of the advantageous descriptions of rock in a general sense) can provide useful insights for preliminary slop design in the case of early open-pit mine planning. In the case of the early planning of an underground mine, the RQD measures can also be utilized to locate the structures in the best rock area away from weak zones that may result in savings of millions of dollars in construction costs.

The quality of rocks according to the RQD within each geo-cluster can also be a major indirect indication of rock hardness for the optimization of mineral processing plants and energy consumption for rock crushing, grinding, and blastability. Nevertheless, this needs to be accompanied by other geo-metallurgical parameters such as the Work Index and further micro texture analyses of the mineralogical assemblage configurations in each geo-cluster. However, further investigation is needed to verify whether low RQD values are related to poor drilling technique or core breakage upon handling. This is useful to corroborate whether RQD values really dictate the quality of the rock.

Since the core of this algorithm uses the sequential indicator simulation, difficulties related to the order relation problem and neighborhood (Emery, 2004; Deutsch, 2006) still persist. A possible future research direction is to develop other non-stationary approaches that use plurigaussian simulation or multiple-point statistics. There is still room for further improvement of the proposed approach. The existence of plenty of hard data makes the simulation process very slow. To solve this issue, an alternative is to use parallel computing. The proposed method can also be tested in other deposit types.

5. Conclusion

The method presented in this study can be used to model heterogeneous geo-domains that result from spatially dependent clustering machine learning algorithms. Multinomial logistic regression was used to model the secondary information to instruct the modelling of the trend component in a non-stationary sequential indicator simulation. The results were then compared with those of a conventional sequential indicator simulation, and it was observed that the proposed method is superior based on the evaluation criteria in which the visual representation of the geo-clusters in the resulting maps and the reproduction of geo-cluster proportions, indicator variograms, connectivity measures, and trend components were extensively examined. The produced realizations showed that the proposed approach is able to produce compact geo-clusters as expected, compared with traditional method that produced patchy and unstructured domains of the geo-clusters. The local uncertainty calculated in the form of probability maps also verified the reproduction of the desired trend of geo-clusters in the region. It also showed that the reproduction of geo-cluster proportions in the proposed approach is generally better and slightly closer to the original declustered proportion of geo-clusters computed over the borehole data. The indicator variogram reproduction of geo-clusters in the proposed approach is also slightly better in the proposed approach. However, the differences between the reproduction of geo-cluster proportions and original indicator variograms in both simulation methods are small. This signifies that both simulation algorithms produce similar results for the reproduction of original proportions and original variograms. Connectivity measures from the results produced by the proposed method showed more continuous and compact domains, which are desired for the purpose of better resource estimation and mine planning.

The realizations obtained by the proposed method were able to produce the expected non-stationary characteristics of the geo-clusters, which were quantitatively investigated using trend analyses. This is trivial as the proposed approach incorporates the trend component in the simulation algorithm.

The proposed method is capable of modelling any geo-domains including geo-clusters with a trend component, and it is intended to produce compact and spatially contiguous domains. However, the realizations might appear slightly patchy across the border of two adjacent geo-domains, or some tiny spots of geo-clusters may appear in the maps (Fig. 11). To solve this issue, image cleaning based on a maximum a posteriori selection (Deutsch, 1998) can be applied. The results in this study are presented without the use of any cleaning algorithms so as to allow for a better comparison of the results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

The first author is grateful to Nazarbayev University for funding this work via Faculty Development Competitive Research Grants for 2021–2023 under Contract No. 021220FD4951. The authors also acknowledge the constructive comments received by the editor, Dr. Guoxiong Chen and two anonymous reviewers who substantially improved the quality of paper.

Appendix

The following figures show the experimental variogram analysis of indicators; connectivity measures along northing and easting; trend analysis along easting and northing for geo-clusters.

References

Abzalov, M., 2016. Applied Mining Geology, Vol. 12. Springer, Cham.
Adeli, A., Emery, X., 2017. A geostatistical approach to measure the consistency between geological logs and quantitative covariates. Ore Geol. Rev. 82, 160–169.
Aghazadeh, M., Hou, Z., Badrzadeh, Z., Zhou, L., 2015. Temporal–spatial distribution and tectonic setting of porphyry copper deposits in Iran: constraints from zircon U-Pb and molybdenite Re–Os geochronology. Ore Geol. Rev. 70, 385–406.
Alabert, F., 1987. Stochastic imaging of spatial distributions using hard and soft information. Master's thesis, Department of Applied Earth Sciences, Stanford University, Stanford, pp. 332.
Alavi, M., 1994. Tectonics of the Zagros orogenic belt of Iran: new data and interpretations. Tectonophysics 229 (3–4), 211–238.
Ambroise, C., Dang, M., Govaert, G., 1997. Clustering of spatial data by the EM algorithm. Proceedings of GEOEV I - Geostatistivs for Environmental Applications. 493–504. https://doi.org/10.1007/978-94-017-1675-8_40.
Berberian, M. and King, G.C.P., 1981. Towards a Paleogeography and Tectonic Evolution of Iran. Canadian Journal of Earth Science, 18, 210-265. doi: 10.1139/e81-019.
Berberian, F., Muir, I.D., Pankhurst, R.J., Berberian, M., 1982. Late Cretaceous and early Miocene Andean-type plutonic activity in northern Makran and Central Iran. J. Geol. Society 139 (5), 605–614.
Boomeri, M., Nakashima, K., Lentz, D.R., 2010. The Sarcheshmeh porphyry copper deposit, Kerman, Iran: a mineralogical analysis of the igneous rocks and alteration zones including halogen element systematics related to Cu mineralization processes. Ore Geol. Rev. 38 (4), 367–381.
Company, K.C., 2008. Final report on geology and alteration of Sarkuh region in 1:5000 scale.
Cramér, H., 2016. Mathematical Methods of Statistics (PMS-9), Volume 9. In Mathematical Methods of Statistics (PMS-9), Volume 9. Princeton university press.
D'Urso, P., Vitale, V., 2020. A robust hierarchical clustering for georeferenced data. Spat Stat. 35. doi: 100407.
de Sá, V.R., Koike, K., Goto, T.N., Nozaki, T., Takaya, Y., Yamasaki, T., 2021. 3D geostatistical modeling of metal contents and lithofacies for mineralization mechanism determination of a seafloor hydrothermal deposit in the middle Okinawa Trough, Izena Hole. Ore Geol. Rev. 135, 104194.
Deutsch, C.V., 1998. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. Comput. Geosci. 24 (6), 551–562.
Deutsch, C.V., 2006. A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. Comput. Geosci. 32 (10), 1669–1681.
Emery, X., 2004. Properties and limitations of sequential indicator simulation. Stoch. Env. Res. Risk Assess. 18 (6), 414–424.
Emery, X., Ortiz, J.M., 2005. Estimation of mineral resources using grade domains: critical analysis and a suggested methodology. J. South Afr. Inst. Min. Metall. 105 (4), 247–255.
Emery, X., Séguret, S.A., 2020. Geostatistics for the Mining Industry: Applications to Porphyry Copper Deposits. CRC Press.
Fouedjio, F., 2016. A hierarchical clustering method for multivariate geostatistical data. Spat Stat. 18, 333–351. https://doi.org/10.1016/j.spasta.2016.07.003.
Fouedjio, F., 2017. A spectral clustering approach for multivariate geostatistical data. Int J Data Sci Anal. 4 (4), 301–312. https://doi.org/10.1007/s41060-017-0069-7.
Fouedjio, F., 2020. Clustering of multivariate geostatistical data. Wiley Interdiscip Rev: Comput Stat. (March):1–13. doi: 10.1002/wics.1510.
Fouedjio, F., Hill, E.J., Laukamp, C., 2018. Geostatistical clustering as an aid for ore body domaining: case study at the Rocklea Dome channel iron ore deposit, Western Australia. Appl Earth Sci: Transe Inst Min Metall. 127 (1), 15–29. https://doi.org/10.1080/03717453.2017.1415114.
Fouedjio, F., 2016b. Discovering spatially contiguous clusters in multivariate geostatistical data through spectral clustering. Proceedings of the Thirteenth International Conference in Machine Learning and Data Mining in Pattern Recognition. Vol. 10086 LNAI, p. 547–557. doi: 10.1007/978-3-319-49586-6_38.
Fouedjio, F., 2017b. A spectral clustering method for large-scale geostatistical datasets. Proceedings of the Thirteenth International Conference in Machine Learning and Data Mining in Pattern Recognition. Vol. 10358 LNAI, p. 248–261. doi: 10.1007/978-3-319-62416-7_18.
Ghasemi, A., Talbot, C.J., 2006. A new tectonic scenario for the Sanandaj-Sirjan Zone (Iran). J. Asian Earth Sci. 26 (6), 683–693.
Green, P.J., Sibson, R., 1978. Computing Dirichlet tessellations in the plane. Comput. J. 21 (2), 168–173.
Haldar, S. K., 2018. Mineral exploration: principles and applications. Elsevier, Oxford (2013), p. 334.
Hosseini, S.T., Asghari, O., Emery, X., 2021. An enhanced direct sampling (DS) approach to model the geological domain with locally varying proportions: Application to Golgohar iron ore mine, Iran. Ore Geol. Rev. 139, 104452.
Hosseini, M.R., Hassanzadeh, J., Alirezaei, S., Sun, W., Li, C.Y., 2017. Age revision of the Neotethyan arc migration into the southeast Urumieh-Dokhtar belt of Iran: Geochemistry and U-Pb zircon geochronology. Lithos 284, 296–309.
Iliyas, N., Madani, N., 2021. An enhanced co-simulation technique for resource modelling using grade domaining: a case study from an iron ore deposit. Appl. Earth Sci. 130 (2), 81–106.
Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recogn. Lett. 31 (8), 651–666.
Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Computing Surveys (CSUR) 31 (3), 264–323.
Journel, A.G., Alabert, F.G., 1990. New method for reservoir mapping. J. Petrol. Technol. 42 (2), 212–218.
Koike, K., Kiriyama, T., Lu, L., Kubo, T., Heriawan, M.N., Yamada, R., 2022. Incorporation of geological constraints and semivariogram scaling law into geostatistical modeling of metal contents in hydrothermal deposits for improved accuracy. J. Geochem. Explor. 233, 106901.
Long, J. S., 1997. Regression Models for Categorical and Limited Dependent Variables. Sage Publications.
Long, J.S., Freese, J., 2006. Regression Models for Categorical and Limited Dependent Variables Using Stata, second ed. Stata Press, College Station, Texas.
Madani, N., Maleki, M., Sepidbar, F., 2021a. Integration of dual border effects in resource estimation: a cokriging practice on a Porphyry copper deposit. Minerals 11 (7), 660.
Madani, N., Maleki, M., Sepidbar, F., 2021b. Application of geostatistical hierarchical clustering for geochemical population identification in Bondar Hanza porphyry copper deposit. Geochemistry 81 (4), 125794.
Madenova, Y., Madani, N., 2021. Application of Gaussian mixture model and geostatistical co-simulation for resource modeling of geometallurgical variables. Nat. Resour. Res. 30 (2), 1199–1228.
Maimon, O., Rokach, L., (Eds.)., 2005. Data mining and knowledge discovery handbook.
Maleki, M., Jélvez, E., Emery, X., Morales, N., 2020. Stochastic open-pit mine production scheduling: a case study of an iron deposit. Minerals 10 (7), 585.
Maleki, M., Madani, N., Jélvez, E., 2021. Geostatistical algorithm selection for mineral resources assessment and its impact on open-pit production planning considering metal grade boundary effect. Nat. Resour. Res. 30 (6), 4079–4094. https://doi.org/10.1007/s11053-021-09928-z.
Malekshahi, S., Rassa, I., Omran, R.N., Rashid Nezhad Omran, N., Lotfi, M., 2018. Comparison of the results of satellite image processing for extraction of alterations with mineralogy and field studies in Sarkuh Porphyry Copper Deposit. Iranian J. Remote Sens. GIS 10 (4), 1–26.
Martin, R., Boisvert, J., 2018. Towards justifying unsupervised stationary decisions for geostatistical modeling: ensemble spatial and multivariate clustering with geomodeling specific clustering metrics. Comput Geosci. 120 (August), 82–96. https://doi.org/10.1016/j.cageo.2018.08.
Martin, R., Boisvert, J., 2020. Performance of clustering for the decision of stationarity; a case study with a nickel laterite deposit. Comput. Geosci. 144, 104565.
Matheron, G., Beucher, H., Galli, A., Guérillot, D., Ravenne, C., 1987. Conditional simulation of the geometry of fluvio-deltaic reservoirs. In: 62nd Annual Technical Conference and Exhibition of the Society of petroleum Engineers, pp. 591–599. SPE Paper 16753, Dallas.
Mirnejad, H., Mathur, R., Hassanzadeh, J., Shafie, B., Nourali, S., 2013. Linking Cu mineralization to host porphyry emplacement: Re-Os ages of molybdenites versus U-Pb ages of zircons and sulfur isotope compositions of pyrite and chalcopyrite from the Iju and Sarkuh porphyry deposits in Southeast Iran. Econ. Geol. 108 (4), 861–870.
Moon, C., Whateley, M., Evans, A., 2005. Introduction to mineral exploration. Blackwell Scientific Publications, Oxford (2006), p. 481.
Moreira, G. D. C., Coimbra Leite Costa, J. F., Marques, D. M., 2020. Defining geologic domains using cluster analysis and indicator correlograms: a phosphate-titanium case study. Appl. Earth Sci., 129(4), 176-190.
Nazarinia, A., Arvin, M., Poosti, M., 2019. Thermobarometry of Mamzar granitoid body, and its tectonomagmatic implication. Iranian J. Crystallography Mineral. 27 (1), 123–134.
Nourali, S., Mirnejad, H., 2012. Hydrothermal evolution of the Sar-Kuh porphyry copper deposit, Kerman, Iran: A fluid inclusion and sulfur isotope investigation.
Oliver, M.A., Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. Math Geol. 21 (1), 15–35. https://doi.org/10.1007/BF00897238.
Renard, P., Allard, D., 2013. Connectivity metrics for subsurface flow and transport. Adv. Water Resour. 51, 168–196.
Romary, T., Rivoirard, J., Deraisme, J., Quinones, C., Freulon, X., 2012. Domaining by clustering multivariate geostatistical data. Geostatistics Oslo 455–466. https://doi.org/10.1007/978-94-007-4153-9_37.

Romary, T., Ors, F., Rivoirard, J., Deraisme, J., 2015. Unsupervised classification of multivariate geostatistical data: two algorithms. Comput Geosci. 85 (December), 96–103. https://doi.org/10.1016/j.cageo.2015.05.019.

Rossi, M. E., Deutsch, C. V., 2014. Mineral resource estimation, Mineral resource estimation. Springer, Netherlands. doi: 10.1007/978-1-4020-5717-5.

Scrucca, L., 2005. Clustering multivariate spatial data based on local measures of spatial autocorrelation. Quaderni del Dipartimento di Economia, Finanza e Statistica. 20 (1), 1–25. http://www.ec.unipg.it/DEFS/uploads/spatcluster.pdf.

Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8 (1), 289.

Sinclair, A.J., Blackwell, G.H., 2006. Applied Mineral Inventory Estimation. Cambridge University Press, p. 400.

Soltani, S., Hezarkhani, A., 2011. Determination of realistic and statistical value of the information gathered from exploratory drilling. Nat. Resour. Res. 20 (4), 207–216.

Yunsel, T.Y., Ersoy, A., 2011. Geological modeling of gold deposit based on grade domaining using plurigaussian simulation technique. Nat. Resour. Res. 20 (4), 231–249.

Zarasvandi, A., Fereydouni, Z., Pourkaseb, H., Sadeghi, M., Mokhtari, B., Alizadeh, B., 2019. Geochemistry of trace elements and their relations with organic matter in Kuh-e-Sefid phosphorite mineralization, Zagros Mountain, Iran. Ore Geol. Rev. 104, 72–87.