Identifying Geologic Features that Lead to Different Mining Paths

Miguel A. Cuba and Clayton V. Deutsch

Paradigms used to evaluate mining sequences that rely on stochastic methodologies are unfeasible to be implemented in real cases due to the problem of having to analyze a large set of realizations. In practice, only a reduced number of mining sequences are presented for evaluation. The SLM paradigm delivers realistic realizations of mining sequences, which account for the periodic evolution of uncertainty in the model of the deposit. However, due to the stochastic nature of the SLM paradigm, the results are difficult to be used in real cases. In this paper, a methodology to identify major trends in the SLM mining sequences is proposed. This methodology reduces the number of SLM mining sequences to evaluate by clustering them into major mining trends. The details of the proposed methodology and an example to illustrate its implementation are discussed in this paper.

Introduction

Mine design is a complex process that involves some aspects that cannot be automated or modeled with mathematical approaches, e.g., management decisions. Due to this complexity, during the design of a mine plan, only a reduced set of the more important mining sequences are evaluated. Conventionally, the variability of the mine plans due to the geologic uncertainty is not accounted for in the simulation paradigms because it would require evaluating a large number of mining sequence alternatives. Hence it becomes impractical to implement Monte Carlo simulation in the evaluation of the deposit (Dominy, Noppé, & Annels, 2002).

In the Simulated Learning Model (SLM) framework, a set of realizations of the mining sequence are simulated. These mining sequences account for the effect of periodical evolution of information updating over time (Cuba, Boisvert, & Deutsch, 2010). To make the SLM framework an alternative that can be implemented in real case situations, it is necessary to present the realizations of the mining sequences in a condensed form. Instead of evaluating each of the realizations of the mine plan independently, major trends in the mining sequence can be identified, thus a reduced group of alternatives can be evaluated instead of the set of realizations. Furthermore, variability patterns in the block model that are the source of these major trends can be identified and can be used as drilling targets to validate specific major trends.

Clustering techniques are widely used in different fields of study like biology, information retrieval, and business to classify data based on specific similarities. Clusters provide a conceptual representation of the elements in which they reside (Tan, Michael, & Vipin, 2006). An approach based on adapted clustering techniques is proposed to find these major trends in the SLM mining sequences. For illustration purpose, a synthetic 3D deposit is evaluated using the SLM framework, where the simulated mining sequences are grouped in major trends to simplify the analysis.

SLM information as elements of a dataset for clustering

The clustering of the mining sequences consists of finding similarities in the planned regions over time. In case these similarities exist, characteristic geologic features that lead to them can be identified. An important part of the implementation of clustering techniques is the comparison of the elements of a dataset to find similarities or dissimilarities that can be used for classification. Conventionally, these elements are represented as points in a feature space. In the case of mine plans, a mining sequence comprises a set of regions positioned in the deposit following a path. Due to the geometric complexity of these regions, for comparison purposes, they cannot be adequately represented by point-like elements such as centroids or medoids. The proposed strategy for comparing the mine plans is based on the comparison of specific volumetric regions of each period.

Implementation of clustering techniques

There is a wide range of clustering techniques that is an unfeasible task to review all of them (Kaufman & Rousseeuw, 2005). Based on a discussion of the three most representative clustering techniques by Tan, Michael, and Vipin (2006), the agglomerative hierarchical clustering (AHC) algorithms are preferred for implementation over the *k*-mean algorithms, and the density based clustering algorithms (DBSCAN). The most important reasons are: *k*-mean algorithms require the number of clusters as a parameter, and DBSCAN algorithms have issues in defining

density of high-dimensional data (Tan, Michael, & Vipin, 2006). AHC algorithms do not require the number of clusters as a parameter, besides they are easy to implement.

The AHC algorithms start assuming that all the elements are clusters of one element. In an iterative process, the two closest clusters are merged to form a new cluster. The iterations stop when all the elements are grouped into one big cluster. There are three conventional approaches for calculating the dissimilarity between clusters: 1) MIN, 2) MAX, and 3) group average (Tan, Michael, & Vipin, 2006). In each of them, the pairwise distances between all the elements are calculated. In the MIN approach, the minimum distance is taken, the maximum for the MAX approach, and the average is calculated in the group average approach. Other methodologies such as the Ward's method or the Lance-Williams formula are also other alternatives.

Since the simulated mine plans are not point elements, for implementing an AHC algorithm, it is necessary 1) to find a measure of dissimilarity that can be used for comparing the planned regions, and 2) choose a methodology for defining the clusters. Both aspects are discussed in the next sections.

Measure of dissimilarity

The dissimilarity between two elements is usually obtained by calculating the *Euclidean* or the *Manhattan* distances between them. Both metrics are special cases of the *Minkowski distance* (Kaufman & Rousseeuw, 2005). Other metrics were also discussed by many authors such as Hartigan (1975), Romesburg (1984), and Abonyi and Feil (2000) among others. Dissimilarity metrics have to satisfy four rules:

1. The metric is zero or positive,
2. The magnitude of the metric is invariant regardless of the direction,
3. The metric is zero when two elements are identical,
4. When comparing three elements, the magnitudes of the metrics should behave as the sides of a triangle.

Semi-metrics can be also used under the condition that they have to obey the first three rules (Fielding, 2007). For example, one semi-metric that has been successfully implemented in several clustering applications is the mutual neighbour distance (Abonyi & Feil, 2000).

In the case of mine plans, the elements of the dataset are not points in a certain feature space but 3D solids that are primarily defined by a volume, a shape, and a position in the deposit. A semi-metric called *volumetric distance* (1) is implemented to account for the main features of the mine sequences. This is defined as the weighted average of the *directional volumetric distances* (2) between the two solids. The volumetric distance is defined as:

$$D_{ab} = \frac{n_a D_{a\to b} + n_b D_{b\to a}}{n_a + n_b}, \tag{1}$$

where, D_{ab} is the volumetric distance between solids a and b, n_a and n_b are the respective weights as a function of their total volumes, and $D_{a\to b}$ and $D_{b\to a}$ are the directional volumetric distances between the two solids.

The directional volumetric distances are calculated as the mean of the minimum Euclidean distances from all the discretized point locations from the first solid to the second solid (2). For implementation purposes, the two solids, a and b, are discretized in n_a and n_b points, respectively. The number of discretized points can be used as the weights of the directional volumetric distances in (1). The directional volumetric distance is expressed as:

$$D_{a\to b} = \frac{1}{n_a}\sum_{i=1}^{n_a} \min\left(dist\left(\mathbf{u}_i, S_b\right)\right),\ \forall \mathbf{u}_i \in S_a, \tag{2}$$

where, $D_{a\to b}$ is the directional volumetric distance from solid a to solid b, n_a is the number of discretized points of solid a, $\mathbf{u}_i$ is the i-th discretized point of solid a, and S_a and S_b are the irregular geometric shapes of solids a and b. During implementation, the centre of the mined blocks within each region can be used as discretized points of the volumetric elements.

The dissimilarity between two mining sequences is calculated as the weighted average of the volumetric distances between periods (3). To keep semi-metric consistence, the dissimilarity is calculated for all the common periods between the two mining sequences. The mining sequence distance is calculated as:

$$MD_{xy} = \frac{\sum_{i=1}^{n_p} w_{(i)} D_{ab(i)}}{\sum_{i=1}^{n_p} w_{(i)}}, \quad (3)$$

where, $MDxy$ is the mining sequence distance between mining sequences x and y, $w(i)$ is a distance weight to give more importance to specific periods, and n_p is the number of common periods between the two mining sequences. During implementation, it is also possible to calculate the mining sequence distances between specified intervals of time periods.

Definition of clusters

In the AHC algorithms, the definition of clusters is often estimated by calculating and evaluating the *dendrogram* that is a tree-like diagram where the branches represent the hierarchy and relationship of how the clusters were merged (Tan, Michael, & Vipin, 2006). The degree of dissimilarity between two clusters is related to the height of the branches.

In cases where there is a reduced number of elements, the definition of clusters can be obtained from a visual evaluation of the dendrogram. This becomes more difficult as the number of elements and the dimensionality of the elements grows. Extensive research has been conducted to calculate the definitions of clusters of AHC algorithms. Milligan and Cooper (1985) reviewed 30 approaches for defining clusters, and concluded that the performance of some criteria depends on the nature of data. Statistical approaches like gap statistic (Tibshirani, Walther, & Hastie, 2001) are mentioned to perform well with lots of types of data. However, it is required to simulate referential data under specific conditions, and in the case of the SLM mine plans, this is not practical.

In the case of the mining sequences, due to the complexity of the mining sequence information, the clusters are defined visually by considering the degree of variability of the mining periods in the mining sequence clusters. The mining sequence trends are represented by cluster prototypes, which are calculated based on the frequency of occurrence of the mining sequence elements in each cluster. The occurrence of major trends in the mining sequences is closely related to the amount of information sampled from the deposit. On one hand, in presence of scarce information, the evolution of the model of the deposit throughout the lifetime of the project is too variable, and that variability is transferred to the corresponding mining sequences. On the other hand, as more information is added to the exploratory drilling campaign, the variability of the model due to the evolution of uncertainty tends to reduce. The occurrence of major trends tend converge into one generic trend.

Proposed Approach

The proposed approach consists of two steps: 1) calculating the dissimilarity between mining sequences and 2) classify the elements by implementing a hierarchical clustering technique. Since the elements to classify are a set volumetric regions placed one after another at a specified time period, conventional point data dissimilarity metrics cannot be implemented. The measure of dissimilarity between two mining sequences is the mining sequence distance (3). The steps to calculate this semi-metric are:

- Select mining regions with similar number of periods. This is important to preserve the properties of the semi-metric. In this approach, two mining sequences with different number of periods are not comparable.
- Calculate the volumetric distances (2) between the regions planned grouped by periods.
- For each pair of mining sequences, combine the calculated volumetric distances to calculate the mining sequence distances (3). The importance of specific periods can be adjusted by assigning weights to the volumetric distances by period.

In the second part of the classification, an agglomerative hierarchical clustering technique is implemented. The steps are as follow:

- Calculate the distance matrix by calculating the mining sequence distance for all the combinations of the mining sequences in the dataset.
- Based on the distance matrix calculate the corresponding dendrogram of the mining sequences.

- Define either visually or calculate the clustering cut-off value. Due to the complexity of the mining sequence elements, it is preferred a visual tuning of the most appropriate clustering cut-off.

For presentation purpose, the clustered sequences can be represented as prototypes, which are element like objects that represent a cluster of elements. In the case of mining sequences, the prototypes can be calculated based on the frequency of the regions planned per period within each cluster. The prototypes have to provide a simplified representation of the mining sequences in each cluster.

Example

The example presented in this section consists of 25 mining sequence realizations of an artificially generated deposit. The implemented infill campaign consists of one infill drill-hole per period. The average number of periods of the mining sequences is eight and the minimum number of periods is five, which means all the realizations at least have five periods (see Figure 1).

The clustering of the mining sequences is done considering the five first periods to include all the realizations in the example. The volumetric distances per period (2) are calculated for all the mining sequences within the selected period interval. In this example, the dissimilarities between mining sequences are calculated as the simple average of the dissimilarities between their respective periods (3). The simple average is used to weight equally the dissimilarity between periods. In this example, the clustering cut-off of the dendrogram is set to 6.5 and four clusters are identified (see Figure 2). The four identified clusters represent the 72% of the total dataset; the remaining 18% are classified as outliers. The proportion between clustered data and outliers as well as the number of clusters depend on the cut-off clustering value. The selection of clustering cut-off is subjective and is based on balancing the reduction of the dispersion of the volumetric distances per period, and reducing the total number of clusters. Smaller cut-off values tend to reduce the dispersion of the volumetric distances at the cost of increasing the total number of clusters. Having a large number of clusters is not very efficient from the perspective of evaluating the mining sequences.

The comparison of the initial state of the dispersion of the volumetric distances (see Figure 3) versus the dispersion within the clusters can be used as a measure of performance of the selection of the clustering parameters. The list of indexes of the SLM realizations within each cluster is:

- **Cluster 1**: 19, 6, 12, 13
- **Cluster 2**: 2, 20, 17, 22, 5, 11, 21, 18
- **Cluster 3**: 3, 1, 10
- **Cluster 4**: 4, 9, 15

The dispersion of the volumetric distances within the four clusters is presented in Figure 4. The reduction of the variability of the volumetric distances is considerable with respect to the initial state. In the initial state, the major source of dispersion occurs in the third period. In the four clusters, this large dispersion in the third period still persists but at a much smaller scale, meaning that the mining sequences within the clusters tend to be more alike.

To simplify the presentation of the mining clusters, the periods of the mining sequences are represented as a condensed cloud of points called prototypes (see Figure 5). The prototypes can be followed to have a simplified representation of the mining path of each cluster. In Figure 5, the prototypes of the four identified clusters or major mining trends are presented.

As expected of the SLM mining sequences, the first period is similar in each of the SLM realizations. It can be seen in the dispersion of volumetric distances between the regions planned in the first period (see Figure 3). In this particular example, there is not much spatial variability in the regions planned in the second period. In Figure 3, the variability of the second period appears relatively low. This can be also seen in the identified cluster prototype maps in Figure 5. The second period seems to be positioned at the same location in the four prototype maps. In Figure 6 - left, the dendrogram of the second period shows the majority of the regions planned are clustered at one location in the deposit. The spatial variability of the third period is larger than the second period. The regions in the third period appear to be clustered in two large clusters (see Figure 6 - right). This can be confirmed in the four identified clusters in Figure 5, where one group of similar regions of the third period are present in clusters 1 and 4, and another group of similar regions can be found in clusters 2 and 3. The reason why the dispersion of the regions is the highest is because the regions are positioned at the two extremes of the

deposit. For the rest of the periods, the regions planned in the fourth and fifth period tend to be positioned at different positions in the deposit (see Figure 5).

Concluding remarks

In this paper, an approach is proposed for presenting the simulated information of SLM mine plans in a condensed form. This condensed form is more suitable to be used during the design stage of a mining project instead of the complete set of realizations. The results are presented in the form of prototypes of the mining trends.

The proposed methodology is easy to implement. Since it relies on agglomerative heretical algorithms for classifying information, it is flexible for different types of comparisons. The most difficult part is to find a proper metric of dissimilarity of the elements. This is solved in this paper by using the volumetric distance and the mining sequence distance. More comparison features, such as material type, block grade, etc. can be added easily to the analysis. In the form the proposed approach is implemented dimensionality is not a problem.

References

Abonyi, J., & Feil, B. (2000). *Cluster Analysis for Data Mining and System Identification.* Boston, Berlin: Birkhauser.

Cuba, M. A., Boisvert, J. B., & Deutsch, C. V. (2010). A New Approach to Forecast Long Term Recoverable Reserves: The Simulation Learning Model. *Centre for Computational Geostatistics - Report Twelve*, p. 415.

Dominy, S. C., Noppé, M. A., & Annels, A. E. (2002, October). Errors and Uncertainty in Mineral Resource and Ore Reserve Estimation: The Importance of Getting it Right. *Exploration and Mining Geology, 11*(1-4), pp. 77-98.

Fielding, A. H. (2007). *Cluster and Classification Techniques for Biosciences.* New York: Cambridge University Press.

Hartigan, J. A. (1975). *Clustering Algorithms.* Toronto: John Wiley & Sons.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data.* Hoboken, New Jersey: John Wiley & Sons.

Milligan, G. W., & Cooper, M. C. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, 159-179.

Romesburg, H. C. (1984). *Cluster Analysis for Researchers.* Belmont, California: Lifetime Learning Publications.

Tan, P. N., Michael, S., & Vipin, K. (2006). *Introduction to Data Mining.* Addison Wesley.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Royal Statistical Society*, 411-423.

Figures

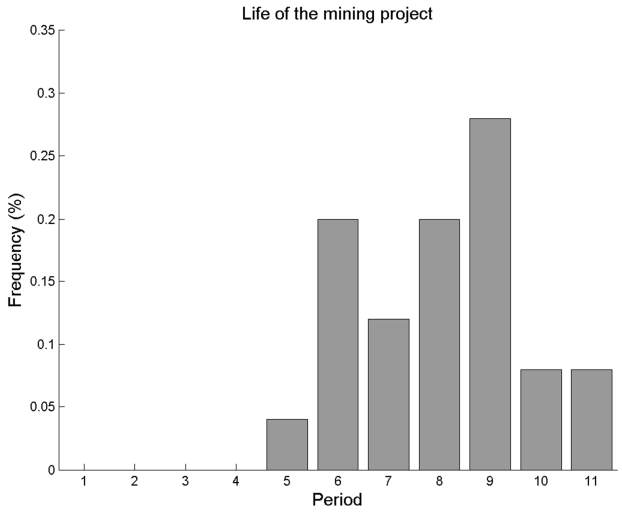

Figure 1: histogram of number of periods of simulated SLM mining sequences

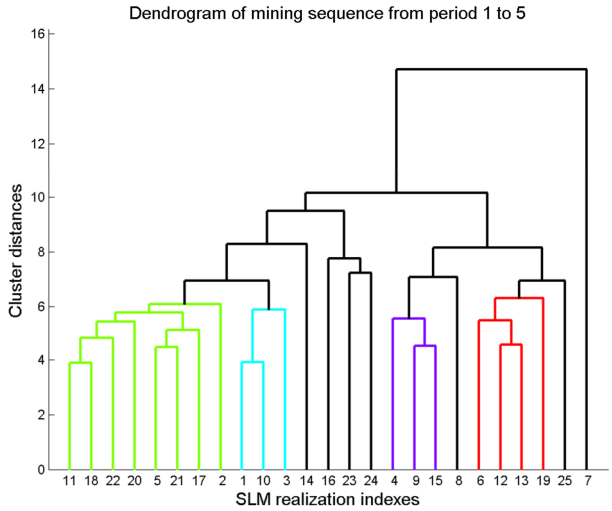

Figure 2: dendrogram of SLM realizations between periods 1 to 5

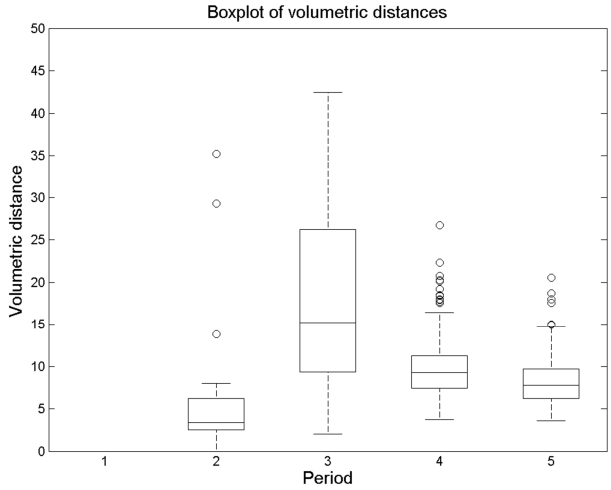

Figure 3: boxplot of volumetric distances per period

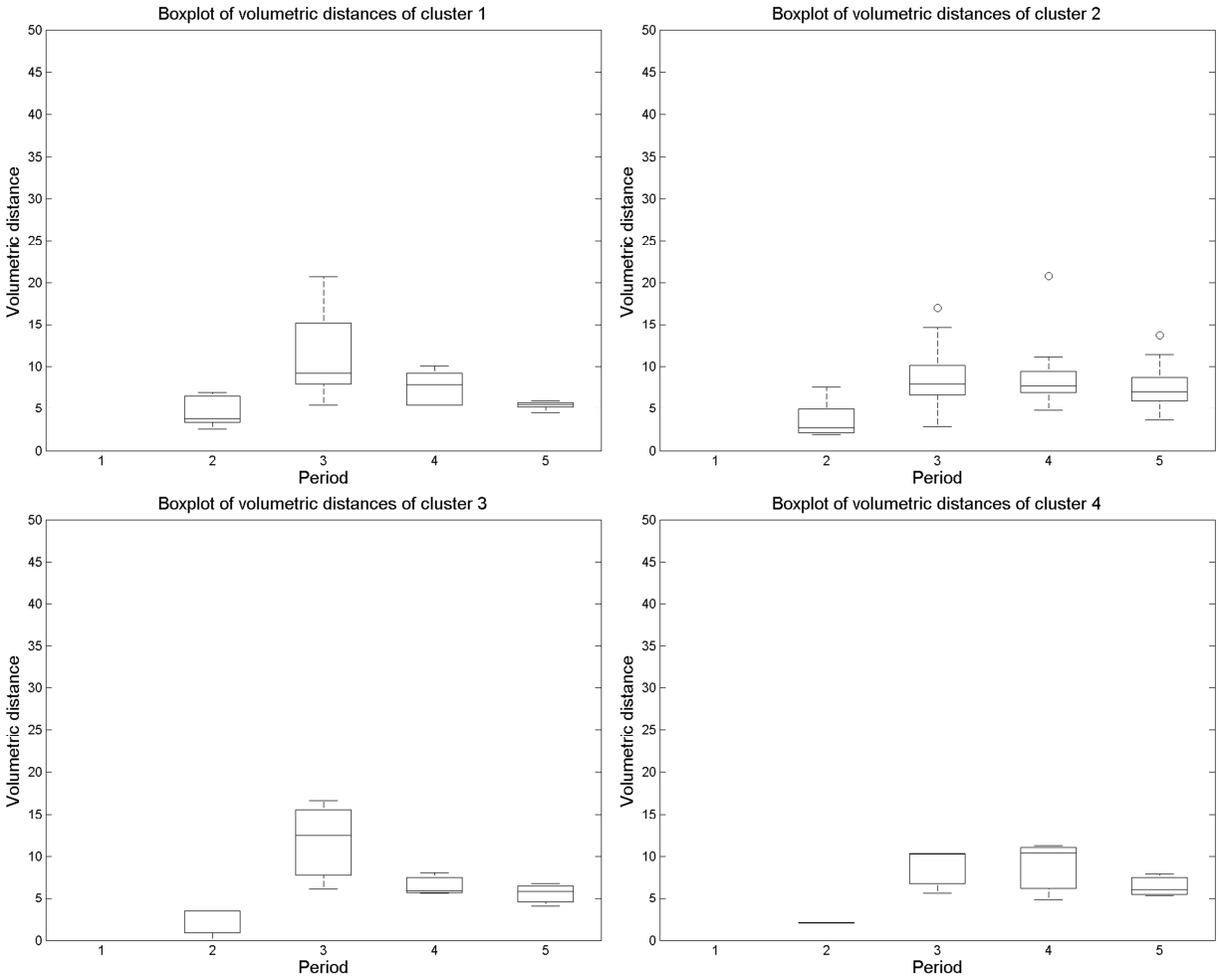

Figure 4: boxplots of the four identified mining sequence clusters

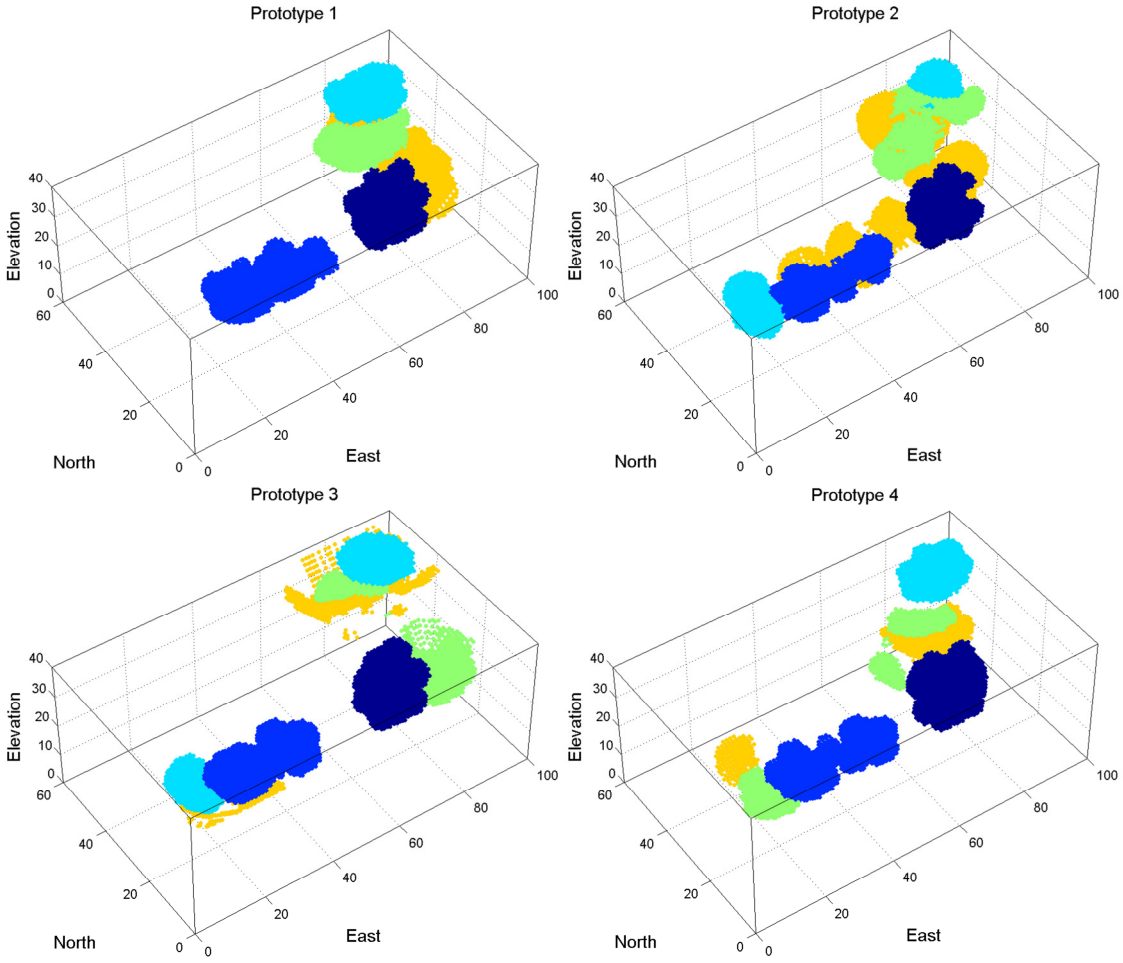

Figure 5: prototypes of four mining sequence clusters from periods 1 to 5

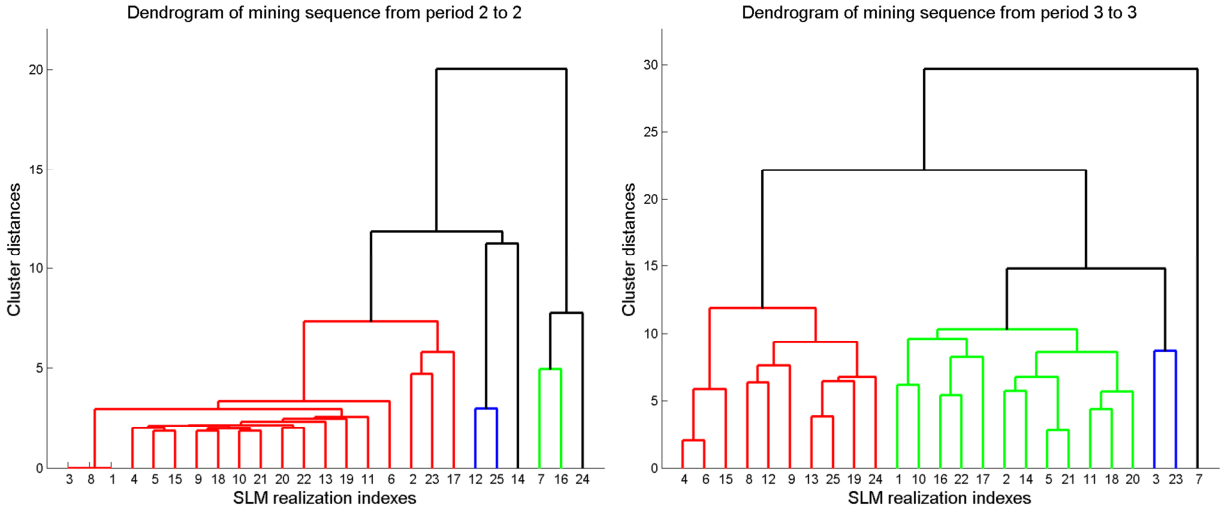

Figure 6: dendrogram of the second (left) and third (right) period of 25 realizations