Original Paper

Optimization of Mining–Mineral Processing Integration Using Unsupervised Machine Learning Algorithms

Siyi Li,[1] Yuksel Asli Sari,[1] and Mustafa Kumral [1,2]

Traditional ore-waste discrimination schemes often do not take into consideration the impact of fluctuations of the head grade, which have on the performance of mineral processing facilities. This research introduces the use of target grades for processing destinations as an alternative to cut off grade-based methods and models, in each processing destination, the losses due to deviation from targets via the Taguchi loss function. Three unsupervised learning algorithms, *k*-means clustering, CLARA and *k*-mean-based approximate spectral clustering, are presented to group mine planning blocks into clusters of similar grades with different processing destinations. In addition, a technique considering uncertainties associated with block grades is proposed to generate new sequences that reduce variation in processing capacities across the life of mine (LoM). The case study in this paper involves the treatment of a realistically large mining dataset. The results showed that clustering methods outperform cutoff grade-based method when divergence from target grades is penalized and that reclassification of blocks based on data from geostatistical simulations could achieve smoother capacities for processing streams across the LoM.

KEY WORDS: Mining and mineral processing, Taguchi loss function, Target grades, Robust clustering, CLARA, Spectral clustering.

INTRODUCTION

Block classification is one of the aspects of mine design that has a direct impact on the profitability of the operation. Many critical reviews on ore-waste classification based on estimation and simulation have been presented (Srivastava 1987; Isaaks 1990; Glacken 1997; Verly 2005). However, one important factor that is often ignored in open pit mine planning is the impact on the performance of processing facilities while having inputs with significant fluctuations in grades. Maintaining a consistent input for

processing facilities is imperative as deviations from the target grades of a processing stream lead to unintended losses in recovery, which can be modeled via the Taguchi loss function (Kumral 2015), a quadratic function that penalizes deviation from a certain target (Taguchi 1986). It has been proposed that every processing stream maintains a target grade where blocks with exactly the same grade receive no loss from processing, but those with grades different from the target are penalized based on their deviations. An illustration of this proposal is shown in Figure 1. Hence, minimizing deviations from target grades would lead to a reduced loss in recovery and throughput and, in turn, the increased value of profits from the operation. A more consistent input for processing will also lead to a more uniform recovery and throughput, which tend to be more desirable.

[1]Department of Mining and Materials Engineering, McGill University, 3450 University Street, Montreal, QC H3A 0E8, Canada.
[2]To whom correspondence should be addressed; e-mail: mustafa.kumral@mcgill.ca
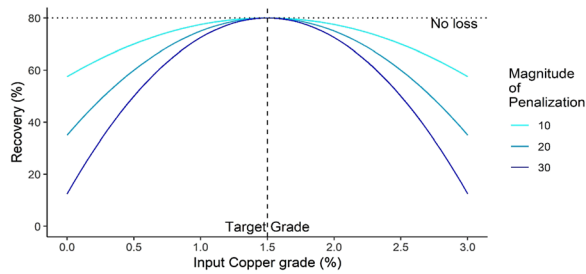
Figure 1. Relationship between input grade and recovery modeled by Taguchi loss function.

Consequently, unsupervised machine learning algorithms such as *k*-means clustering or partitioning around medoids (PAM) could be used to group blocks into different clusters, with each cluster signifying a processing stream with pre-defined target grades. In doing so, the within cluster dissimilarities could be minimized, while target grades of each processing stream could then be set to the grade values of each cluster centroid. In recent years, many machine learning methods have been introduced to optimize geology, mining and mineral processing systems (Zhang et al. 2011; Yanyan et al. 2015; Goodfellow and Dimitrakopoulos 2016; Ruiseco et al. 2016; Li et al. 2019; Nguyen et al. 2019; Nwaila et al. 2019; Rajabinasab and Asghari 2019; Villalba Matamoros and Kumral 2019). However, but few studies have taken into full consideration the penalties that come with deviation from targets in input grade and processing capacity. Performance of the introduced clustering technique will be evaluated with the overall profitability of the operation while taking into account the high costs of constructing additional processing facilities, so that new processing streams are built if and only if the cost more than balances out for the losses in recovery due to deviation from target grades. A group of related research topics has also highlighted the potential applications of clustering techniques in addressing similar research problems (Ruiseco and Kumral 2017; Ruiseco et al. 2016; Sari and Kumral 2018; Sepúlveda et al. 2018).

After deciding the optimal number of processing streams through clustering, the capacities of processing streams could be found by counting the number of data points in each cluster. Nevertheless, planning of a processing stream's capacity during the

life of mine (LoM) is also important and challenging, as companies generally seek to maximize NPV in mine planning and hence blocks of higher values tend to be extracted at the earliest possible period, leaving the overall processing capacity skewed. However, producing below the processing capacity or deviating from the process target grade may also lower the NPV.

Therefore, in this research, the traditional block sequencing is improved by identifying blocks whose processing destination, according to a portion of its simulated grades, differs from that determined by the average expected grades. Switching the processing destination of such blocks reduces variation in processing capacities across the LoM at minimum cost and risk. The original contribution of this research roots from the introduction of target grades in mineral processing streams and the utilization of the Taguchi loss function for modeling penalized recovery. Moreover, CLARA, which is a robust clustering algorithm for large datasets, was used, and total revenues from different scenarios were compared and optimized. In addition, block destinations were tweaked according to sequential Gaussian simulations and capacities of processing streams can be further smoothed across the LoM.

METHODOLOGY

Clustering Algorithms for Optimal Process Design

The k-means Clustering Algorithm

The *k*-means algorithm is one of the most commonly used clustering algorithms that scales relatively well with large datasets. It partitions a given dataset into *k* pre-specified number of clusters in such a way that within cluster dissimilarity is minimized and inter-cluster dissimilarity is maximized. Various distance measures exist for defining dissimilarity among data points, including the Euclidean distance, the Manhattan distance and many other correlation-based distances. Euclidean distance is chosen in this case because it considers exactly the spatial distance between points. A brief summary of the *k*-means algorithm is shown in Algorithm 1 (Kassambara 2017).

Algorithm 1. k-means Clustering

Input:

      Data matrix with n observations: $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}]^T, \; \boldsymbol{x_i} \in \mathbb{R}^p$

      Number of clusters k

      Maximum number of iterations N

Output:

      k clusters

1:    Randomly initiate cluster centroids $\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, ..., \boldsymbol{\mu_k} \in \mathbb{R}^p$

Until convergence or iterations $= N$ **do**

2:    Assign each observation (i) to the closest centroid

$$c_{(i)} \leftarrow arg \; \min_j ||\boldsymbol{x_i} - \boldsymbol{\mu_j}||_2^2$$

3:    Update each centroid (j) by taking the average

$$\boldsymbol{\mu_j} \leftarrow \frac{\sum_i \mathbb{1}_{\{x_i \in C_j\}} \boldsymbol{x_i}}{|C_j|}$$

One common metric used to evaluate the goodness of k-means clustering is the total within cluster sum of squares (TWSS); its formula is:

$$\text{TWSS} = \sum_{i=1}^{k} WSS(i) = \sum_{i=1}^{k} \sum_{j \in C_i} \left\| x_i - \mu_j \right\|_2^2 \quad (1)$$

where x_j refers to the jth data point, μ_i is the cluster center of the ith cluster and C_i is the set of all points in the ith cluster. Results of the k-means algorithm are known to be sensitive to the selection of k initial cluster centers. It is common practice, therefore, to start with many different initial allocations and then choose the one that performs best. As the number of clusters, k, has to be specified before the algorithm could be run, the optimal number of clusters can be determined by plotting the TWSS against the number of clusters (James et al. 2013). Ideally, the TWSS value should be minimized. However, as the value will always tend to zero when the number of clusters tends to the number of observations in the dataset, it is important to select the optimal number of clusters (k_{optim}) such that a further increase in the number of clusters will lead to significant diminishing benefit in the reduction in TWSS. Identifying the optimal number of clusters in this manner is also more commonly known as the elbow method.

Partitioning Around Medoids (PAM) and CLARA

The k-means clustering algorithm has numerous drawbacks, including: (1) The number of clusters, k, must be manually chosen by the users; (2) the final output is dependent on the initial random assignment of cluster numbers; and (3) the algorithm shows sensitivity to noise and outliers due to the use of means. The first and second problem can be addressed, respectively, by running the algorithm for a set of plausible values of k, and different initial random cluster assignments (usually from 25 to 50) and selecting the solution with the best performance. While trying to find a better clustering algorithm, it is natural to consider other methods such as hierarchical agglomerative clustering or graph-based spectral clustering, which do not require the prior specification of the number of clusters. Unfortunately, however, such clustering techniques, despite being powerful, do not scale well with large data. When clustering a dataset with n observations into k clusters, the computational complexity of the k-means algorithm per iteration is approximately $O(nk)$, whereas hierarchical and spectral clustering could cost as much as $O(n^3)$, making them almost impossible to be applied in clustering large-scale mining data. Partitioning around medoids (PAM) can be considered similar to a robust form of k-

means clustering. While in k-means, each cluster is represented by the mean of all data that belongs to it, in PAM a cluster is represented by its most central element, named its medoid. The general PAM algorithm is described in Algorithm 2 (Kassambara 2017).

Algorithm 2. Partition Around Medoids (PAM)/

Input:
      Data matrix with n observations: $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}]^T, \ \boldsymbol{x_i} \in \mathbb{R}^p$
      Number of clusters k
      Maximum number of iterations N
Output:
      k clusters
1:    Randomly initiate cluster centroids $\boldsymbol{\mu_1}, \boldsymbol{\mu_2}, ..., \boldsymbol{\mu_k} \in \mathbb{R}^p$
Until convergence or iterations = N **do**
2:    Assign each observation (i) to the closest centroid
3:    Within each cluster, for each pair of medoid and non-medoid, compute the change in TWSS if a switch is made.
4:    Make the optimal switch then go to 2, converge otherwise

---

At the cost of $O\left(k(n-k)^2\right)$, PAM is still too cumbersome to be applied to truly large datasets. Hence, a modified version of PAM based on re-sampling named CLARA (Clustering LARge Applications) was selected to optimize processing options. The CLARA algorithm is a modified version PAM designed for large datasets, and its general idea is to draw multiple samples from the dataset and apply PAM to them. The basic steps of CLARA are displayed in Algorithm 3 (Kaufman and Rousseeuw 2008).

K-means-Based Approximate Spectral Clustering (KASP)

Spectral clustering is one of the most powerful modern clustering algorithms and is based on the spectral decomposition of the graph Laplacian matrix of the data matrix. As a graph-based method, each observation in the data matrix is viewed as a vertex in the graph, and the dissimilarities between data are viewed as edges between vertices. Spectral clustering functions by identifying the optimal cut to

---

Algorithm 3. CLARA

Input:
      Data matrix with n observations: $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, ..., \boldsymbol{x_n}]^T, \ \boldsymbol{x_i} \in \mathbb{R}^p$;
      Number of clusters k;
      Maximum number of iterations N;
      Sample size m;
Output:
      k clusters;
Until convergence or iterations = N **do**
1:    Randomly draw a sample $\boldsymbol{S} \in \mathbb{R}^{m \times p}$ from $\boldsymbol{X}$
2:    Identify k representative medoids via $PAM(\boldsymbol{S}, k, N)$
3:    Assign each observation (i) in $\boldsymbol{X}$ to the closest centroid, then calculate TWSS
4:    Go back to 1, keep clustering result if TWSS decreases

The computational complexity of CLARA is $O\left(km^2 + k(n-k)\right)$, which is a significant improvement from PAM. The downside of CLARA is that if the best k medoids were not selected in the sampling process, then CLARA would produce a sub-optimal solution. When applying CLARA, the algorithm is run with a large m value for multiple times in order to adjust for sampling bias.

partition the graph such that the sum of the weights of the edges cut in the process is minimized. The basic form of a spectral clustering algorithm is described in Algorithm 4 (von Luxburg 2007).

Algorithm 4. Spectral Clustering

Input:

      Data matrix with n observations: $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}]^T$, $\boldsymbol{x_i} \in \mathbb{R}^p$

      Number of clusters k

Output:

      k clusters

1:     Form adjacency matrix ($\boldsymbol{W} \in \mathbb{R}^{n \times n}$) according to pre-defined dissimilarity measure

2:     Form diagonal degree matrix ($\boldsymbol{D} \in \mathbb{R}^{n \times n}$) such that the diagonal entries of $\boldsymbol{D}$ corresponds to the row sums of $\boldsymbol{W}$

3:     Form graph Laplacian matrix $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$

4:     Compute the spectral decomposition of $\boldsymbol{L}$, $\boldsymbol{L} = \boldsymbol{V\Lambda V}^T$, then find the k eigenvectors ($\boldsymbol{Z} \in \mathbb{R}^{n \times k}$) corresponding to the k smallest eigenvalues of $\boldsymbol{L}$

5:     Use k-means to cluster $\boldsymbol{Z}$ into k clusters, assign the rows of $\boldsymbol{X}$ to the same clusters as rows of $\boldsymbol{Z}$

Unfortunately, the spectral clustering algorithm is computationally expensive at a complexity of $O(n^3)$, largely due to the need to construct explicitly the adjacency matrix $\boldsymbol{W}$ and the spectral decomposition of $\boldsymbol{L}$. With a large dataset, one of the alternatives is to use the k-means-based approximate spectral clustering algorithm (KASP) proposed by Yan et al. (2009), which functions by first compressing the data into l representative observations, then applying spectral clustering to the compressed data. The ratio $\frac{l}{k}$ is referred to as the compression ratio. A brief summary of the KASP algorithm is shown in Algorithm 5.

recovery. Therefore, given k and n, there are in total C_k^n different scenarios for processing. Let the maximum number of clusters be m, then the total number of possible scenarios is given by:

$$\text{Number of scenarios} = \sum_{k=1}^{m} C_k^n \qquad (2)$$

The idea of 'target grade' is applied in this paper, such that grade deviation from the mean will receive a penalized recovery during processing can be modeled with the Taguchi loss function (Kumral 2015).

Algorithm 5. KASP

Input:

      Data matrix with n observations: $\boldsymbol{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}]^T$, $\boldsymbol{x_i} \in \mathbb{R}^p$

      Number of clusters k

      Number of representations l

Output:

      k clusters

1:     Use k-means to partition $\boldsymbol{X}$ into l clusters, record the cluster centroids as landmarks $\boldsymbol{L} \in \mathbb{R}^{l \times p}$

2:     Use spectral clustering to partition $\boldsymbol{L}$ into k clusters. Assign the rows of $\boldsymbol{X}$ to the corresponding clusters of their representations

---

An Economic Evaluation of Processing Scenarios

After using the k-means algorithm to group the data points into k different clusters, the clusters were sorted in ascending order of average grades. Then, different k processing streams were sampled from n available processing streams without replacement, also in ascending order of recovery, to match the k clusters. Ordering the clusters as well as the processing streams ensures that clusters with higher average grades are sent to processing streams designed to have higher

$$L\left(x_j^\gamma\right) = c\left(x_j^\gamma - \mu_i^\gamma\right)^2 \forall x_j \in C_i \qquad (3)$$

where x_j^γ is the value of attribute γ (in the polymetallic case) of the jth block in the ith cluster, which is denoted by C_i, with $\mu_{i_i}^\gamma$ being the value of attribute γ of its center. $L\left(x_j^\gamma\right)$ represents the loss in the recovery of attribute γ and c is a constant that magnifies the penalization. Revenue and cost calcu-

lations were performed on each scenario and the one that maximizes profit was deemed as optimal. The formulas for calculations of revenue and cost are shown in Eqs. 3 and 4. The meanings of the parameters are shown in Table 1.

$$\text{Total revenue} = \sum_{j=1}^{N} R(x_j)$$

$$= \sum_{j=1}^{N} \sum_{\gamma \in \boldsymbol{a}} x_j^{\gamma} \times \rho \times V \times P_{\gamma}$$

$$\times \left[r_i^{\gamma} - L\left(x_j^{\gamma}\right) \right] \quad (4)$$

$$\text{Total cost} = \text{Total construction cost for processing streams}$$
$$+ \text{Total mining cost}$$
$$+ \text{Total processing cost}$$

$$(5)$$

$$\text{Total cost} = \sum_{i=1}^{m} M_{\boldsymbol{i}} + N \times m_c \times \rho \times V + \sum_{j=1}^{N} \sum_{i=1}^{m} y_{ji}$$
$$\times \rho \times V \times p_i.$$

$$(6)$$

Processing Capacity Tuning Based on Simulation

In the previous step, an optimal processing scheme was selected via a clustering algorithm such that for every mineral processing stream, deviation from target grade is minimized. Having found the most suitable processing options, block sequencing and scheduling were completed in a commercial mine production scheduling software with the mean of the simulated block grades as input, the corresponding sequence output was exported and the number of processed blocks for each processing stream in each period was found. In order to have the processing capacities of the streams as uniform as possible, the blocks were analyzed based on their grades in the 15 different simulations, so that different probable grade scenarios of blocks can be studied, and a subset of blocks could be sent to alternative destinations if their grades correspond to different destinations in different scenarios. Such blocks are named 'marginal' blocks and are defined as blocks whose most likely destination according to a subset of the simulated grades differs from the one computed from the average expected case. After identifying the marginal blocks, in each period, depending on the situation, marginal blocks were sent to their most likely destination to reduce variation in processing capacities. When the high-grade processing was over the mean capacity and the low-grade processing was under the mean capacity, marginal low-grade blocks currently sent to high-grade processing were switched to low-grade processing to fill the gap, and if there were insufficient blocks, then marginal low blocks currently sent to waste were also switched to low processing. When low-grade processing was over the mean capacity and the high-grade processing was under, then marginal low-grade blocks were sent from waste to low-grade processing and the marginal high-grade blocks from low-grade processing to high-grade processing as well. Similarly, if both processing streams were over or under the mean capacity, then the marginal waste blocks currently in low and high processing were sent to waste or marginal low and

Table 1. List of parameters for revenue and cost calculations

Parameter	Representation	Unit
a	Vector of attributes	
ρ	Block bulk density	ton/m^3
V	Block volume	m^3
x_j^{γ}	jth block grade of γth attribute	%
N	Total number of blocks	
P_{γ}	Price of γth attribute	\$
r_i^{γ}	Recovery from ith processing stream of γth attribute	%
$L\left(x_j^{\gamma}\right)$	Loss of recovery from ith processing stream of γth attribute	%
y_{ji}	Binary variable (1 if jth block sent to ith processing, 0 otherwise)	
p_i	The processing cost of ith processing stream	\$/ton
$M_{\boldsymbol{i}}$	Cost of constructing the ith processing stream	\$
m	Total number of clusters/processing streams	
m_c	Mining cost	\$/ton

the high blocks were sent, respectively, to low and high processing. While switching destinations, marginal blocks were ranked according to the descending order of likelihood; hence, blocks with highest likelihoods were switched first. A schematic of the process is shown in Figure 2. By switching the destination of the marginal blocks to their corresponding most likely destinations, variation in processing capacities across the LoM can be effectively reduced at a minimum level of risk.

CASE STUDY

Determination of Optimal Processing Scenario

A relatively large dataset related to a Cu deposit was used in this study. The dataset contains 145,800 blocks with 15 equally likely geostatistical simulations generated with sequential Gaussian simulations (Journel and Alabert 1989; Chilès and Delfiner 2012). The simulations were realized on the nodes or locations of a random grid. In this simulation, conditioning data were converted to equivalent normal values, and the variography of the converted values was computed. Using conditioning and previously simulated values, the value was then estimated (kriged) at the simulation location of the grid. A random sample was finally taken from the distribution characterized by the estimated kriged value and its variance at the simulation location on the grid. This process was repeated for all locations on

Figure 2. Identification and changing destination of marginal blocks.

the grid. In addition to generating multiple realizations of grade uncertainty, sequential Gaussian simulation was also used to reproduce variability in various engineering phenomena such as soil water content (Delbari et al. 2009), the standard penetration tests to characterize soil exploration (Basarir et al. 2010), Ni contamination (Qu et al. 2013) and appraising geochemical anomaly (Liu et al. 2018). As expressed by Dowd (1993), geostatistical simulation must meet the following criteria: (i) simulation and actual values agree with each other at all sample locations; (ii) each simulation must exhibit the same spatial dispersion; (iii) each simulation and the true values must exhibit the same distribution; and (iv) if there are multiple attributes, their simulations must co-regionalize each other in the same manner as the true values. These criteria were tested for the simulations and verified that the criteria are satisfied. Thus, a series of simulations complying the criteria given above was reproduced. An important speculative aspect is the number of simulations required in mine planning works. Goovaerts (1999) discussed the effect of the number of simulations on transfer functions and concluded that sequential Gaussian simulation produced outcomes that are more accurate. He also emphasized that having more than 20 simulations has not much effect on accuracy.

To compensate for computational complexities, relatively few possible processing stream options were considered. Detailed information regarding those processing options is shown in Table 2. A list of profitability parameters used in this case study is detailed in Table 3. A histogram depicting the expected average of the 15 simulations is shown in Figure 3.

In this case study, blocks with grades lower than the lowest possible cutoff grade (in this case, 0.84%) determined from the processing stream option with the lowest processing cost and recovery were not included in the clustering algorithm, such that only blocks classified as ore were partitioned into clusters. The optimal number of clusters was decided by plotting the TWSS against the number of clusters and selecting the cluster number where the next increment in the number of clusters results from a significantly lower decrease in TWSS than the previous number. The results from the clustering methods are shown in Figure 4. Due to limited computational power available, the maximum compression ratio of KASP used was 2%. KASP with 1% compression ratio was also performed to identify
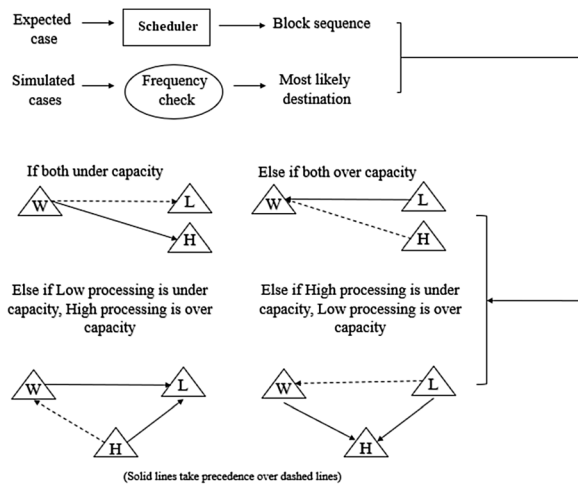
Table 2. List of processing stream options

Processing stream	Processing cost ($/ton)	Recovery (%)	Construction cost ($M)
1	20	40	10
2	35	65	10
3	45.5	80	12.5
4	57.25	95	15

Table 3. List of profitability parameters

Parameter	Representation	Unit
P_{copper}	Price of Cu per ton	$5939.1
m_c	Mining cost per ton	$1.75
c	Magnitude of penalization	30
V	Block volume (Block size 5m × 5m × 10m)	250 m^3
ρ	Block bulk density	4 ton/m^3
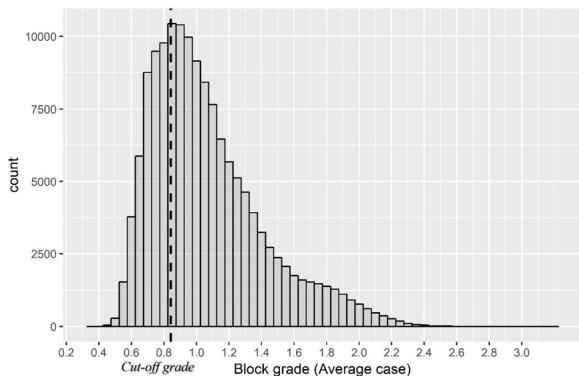
Figure 3. Simulated average block grades.
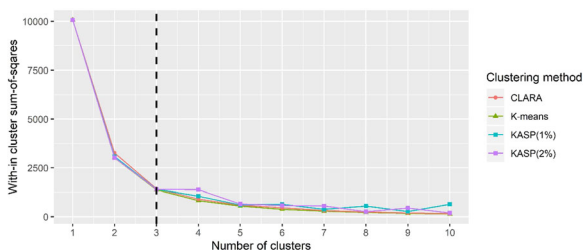
Figure 4. TWSS plot for different clustering techniques.

the impact of compression ratio on the overall performance of the clustering algorithm. The optimal number of clusters from both clustering methods was found to be 3. It can be observed that k-means has only marginally better TWSS when compared

with CLARA, even when medoids were used as cluster centers instead of means in CLARA, while KASP performed similarly to K-means before three clusters, but fluctuated with more clusters, possibly due to the small compression ratio used.
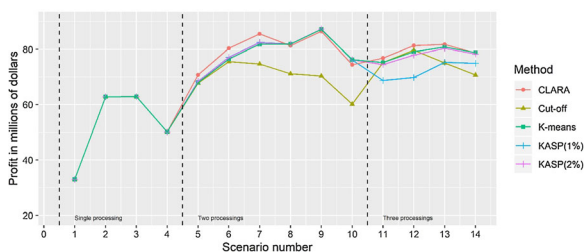
The total number of processing scenarios was calculated to be 14. Economic evaluations were performed on all scenarios, according to k-means clustering, CLARA, KASP with 1% and 2% compression ratio and marginal cutoff grade, respectively. Table 4 details the possible processing scenarios, where, in each scenario, the clusters were mapped with different corresponding processing destinations. For instance, in processing scenario 7, the dataset was partitioned into two clusters ranked by average grade values, with cluster 1 mapped with processing 1 and cluster 2 with processing 4. The profits for different processing scenarios and clustering methods were computed exhaustively and are shown in Figure 5.

As can be seen from Figure 5, for this particular dataset and parameters, the maximum profit was generated by the KASP with 2% compression ratio at processing scenario 9 with a value of $87.25 M. In general, when the deviations from target grades were penalized in mineral processing, determining block destinations via clustering algorithms generated higher profits when compared to using marginal cutoff grade. In this particular case, CLARA generated higher profits than results from other clustering algorithms in most scenarios, but KASP with 2% compression ratio performed marginally better

Table 4. List of processing scenarios

Scenario number	Cluster 1	Cluster 2	Cluster 3
1	1		
2	2		
3	3		
4	4		
5	1	2	
6	1	3	
7	1	4	
8	2	3	
9	2	4	
10	3	4	
11	1	2	3
12	1	2	4
13	1	3	4
14	2	3	4

Table 5. Processing capacities for old and new sequencing

	Mean	Variance
Low processing		
Old sequencing (before switching)	7948	420622
New sequencing (after switching)	7904	289407
High processing		
Old sequencing (before switching)	2891	1442632
New sequencing (after switching)	2779	1187592
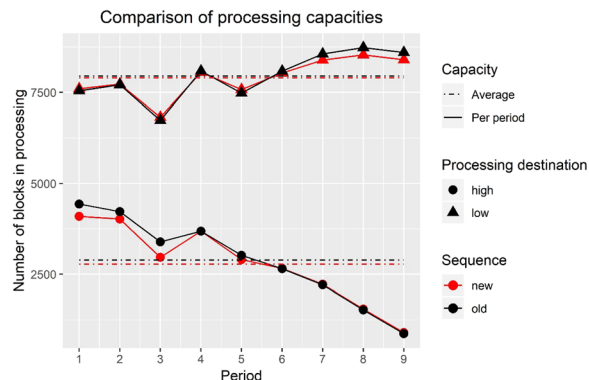
Figure 6. Processing capacity across LoM for old and new sequencing.

Figure 5. Profit comparison.

than CLARA at its scenario with maximum profit, while CLARA at processing scenario 9 resulted in a profit of $86.47 M. It can also be seen that KASP performed better in all scenarios when the compression ratio was increased. If higher computational power were available, KASP can be projected to yield even better results.

Capacity Tuning of Processing Streams Based on Geostatistical Simulations

From the previous section, the processing scenario with the highest profit was identified to be scenario 9 with processing streams 2 and 4 selected for low-grade and high-grade processing, respectively. At a mining capacity of 15,000 blocks per period, the commercial software outputs a total LoM of 10 periods (years). Marginal blocks were identified from the geostatistical simulations and were then sent to their most likely destination to reduce variation in processing capacities. The pro-

cessing capacities of processing streams across the LoM before and after the switching are shown in Table 5. The final year of LoM was intentionally left out as most of the valuable ore has been mined out and there is insufficient material left for mining. Details of the mean and variances of processing capacities across the LoM are shown in Figure 6. The mean for both processing streams was lowered to a small extent due to switching blocks from low- and high-grade processing to waste. The new sequencing generated by the switching of marginal blocks managed to lower the variance in low-grade processing capacity by 31% and that of high-grade processing by 17%. Because of the reclassification of blocks, a smoothing effect on the processing volumes throughout the periods can be observed.

Figure 7 shows in situ grades and the outcomes of CLARA and KASP for three destinations. In this figure, the blocks shown in navy blue, green and claret red are routed to waste dump, low-grade and high-grade processing, respectively. The consistency between the grades and block destinations can be seen easily in the figure. As also seen from the figure, the number of blocks to be sent to high-grade
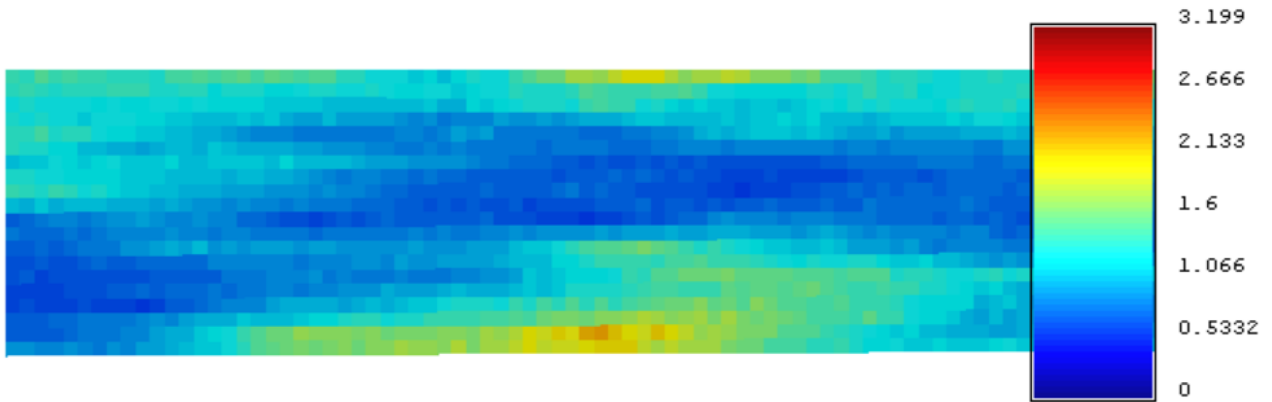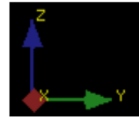
(a) Block grades (%)
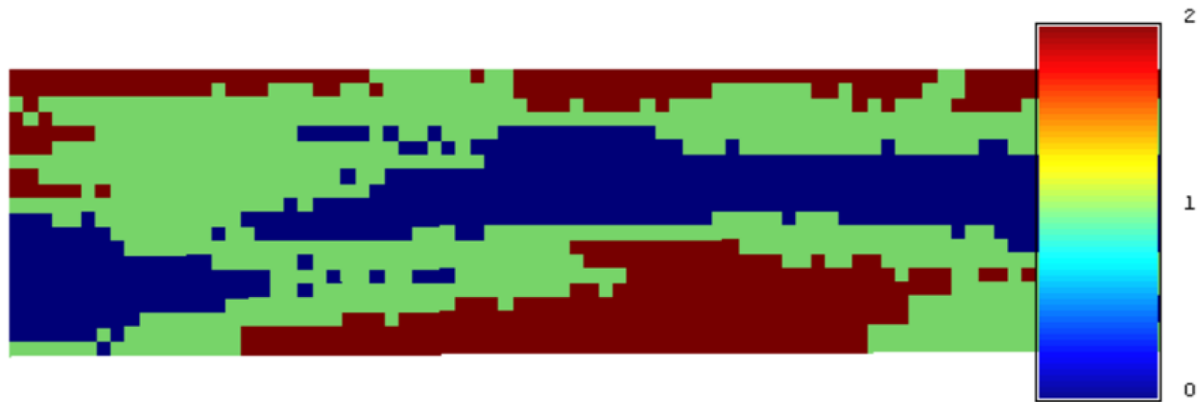
(b) Results of CLARA
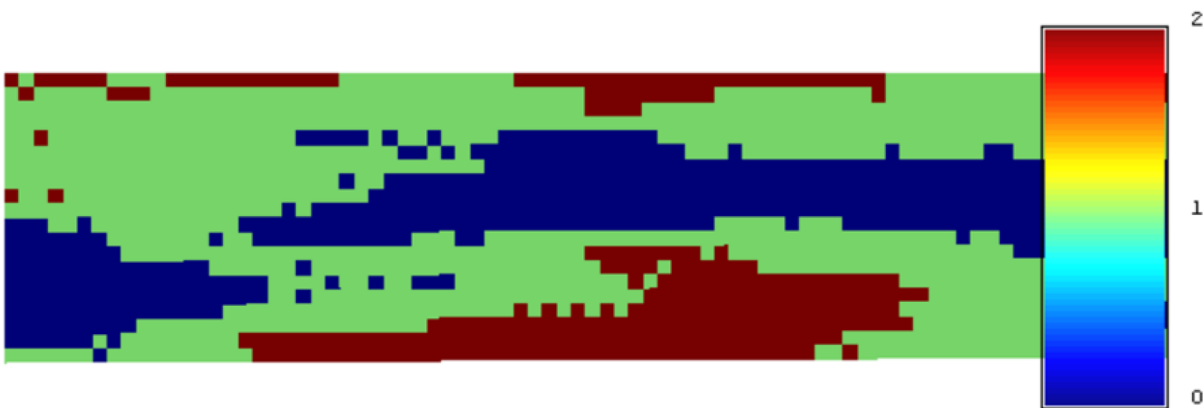
(c) Results of KASP

Figure 7. Block grades **a** and process destinations of blocks **b** CLARA and **c** KASP.

processing is more in CLARA's results compared to KASP's results.

CONCLUSIONS

This research introduces the use of clustering algorithms to generate clusters of selective mining units with similar grades that correspond to different processing destinations while minimizing the within cluster dissimilarities in mineral grades. Realistic concerns, including deviation from target grades and capacities in processing facilities, are also taken into consideration via the penalization of recovery via the Taguchi loss function and calculating the number of data points in each grouped cluster. One of the important factors in the determination of the profit from the clustering algorithms is the magnitude of penalization of the Taguchi loss function, with better results expected from the clustering methods when a high degree of penalization is present. Another influential factor is the overall scale and profitability of the mining operation, with smaller operations being unlikely to balance out the high amount of additional costs of constructing extra processing facilities. A more sophisticated clustering algorithm than k-means, CLARA, is based on performing PAM on random samples of the original dataset and is considered more robust than k-means. In this particular setting of the study, clustering with respect to CLARA generated more profit than k-means in almost all scenarios, despite k-means performing slightly better in scenario 9, the scenario with the highest profit. KASP, which provides a computationally efficient solution approximate to spectral clustering, was the top-performing clustering algorithm and generated higher profit than k-means in the optimal scenario. Increasing the compression ratio of KASP also had an impact on generating better results. In future studies, when the dataset is large, all three clustering methods should be considered in grouping blocks with similar grades. Furthermore, by identifying marginal blocks judging from the simulated block grades, it is possible to tune the processing capacities by changing their destinations. In doing so, variation in processing capacities across the LoM can be reduced at minimum risk and cost. Although the simulated grades may differ from the actual grades and this may result in potential economic loss, the aim of the proposed methodology is to provide an efficient capacity installation approach in which the mine

production schedule is considered. If the grade heterogeneity increases in a deposit, the number of simulations should be increased to capture the deposit uncertainty more accurately. After the settling of the processing capacities, the mine schedule can be generated with the new parameters. The other extension will be an incorporation of rock and metallurgical characteristics affecting processing performance into the process design.

REFERENCES

Basarir, H., Kumral, M., Karpuz, C., & Tutluoglu, L. (2010). Geostatistical modeling of spatial variability of SPT data for a borax stockpile site. *Engineering Geology, 114*(3), 154–163.

Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: Modeling spatial uncertainty.*, Wiley series in probability and statistics New Jersey: Wiley. https://doi.org/10.1002/9781118136188.

Delbari, M., Afrasiab, P., & Loiskandl, W. (2009). Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content. *CATENA, 79*(2), 163–169.

Dowd, P. A. (1993). Geostatistical simulation. Course notes for the MSc in mineral resources and environmental geostatisticsLeeds.

Glacken, I. M. (1997). Change of support and use of economic parameters for block selection. In E. Y. Baafi & N. A. Schofield (Eds.), *Geostatistics Wollongong '96*. Berlin: Springer.

Goodfellow, R. C., & Dimitrakopoulos, R. (2016). Global optimization of open pit mining complexes with uncertainty. *Applied Soft Computing, 40,* 292–304.

Goovaerts, P. (1999). Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. *Stochastic Environmental Research and Risk Assessment, 13*(3), 161–182.

Isaaks, E. H. (1990). *The application of Monte Carlo methods to the analysis of spatially correlated data.* Stanford University. Retrieved January 21, 2019 from https://books.google.ca/books?id=say4QwAACAAJ.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R.* New York: Springer.

Journel, A. G., & Alabert, F. (1989). Non-Gaussian data expansion in the earth sciences. *Terra Nova, 1*(2), 123–134.

Kassambara, A. (2017). *Unsupervised machine learning: Practical guide to cluster analysis in R: Unsupervised machine learning.* STHDA. S.l.: CreateSpace Independent Publishing Platform.

Kaufman, L., & Rousseeuw, P. J. (2008). Clustering large applications (Program CLARA). In L. Kaufman & P. J. Rous-

seeuw (Eds.), *Finding groups in data: An introduction to cluster analysis* (pp. 126–163). https://doi.org/10.1002/9780470316801.ch3.

Kumral, M. (2015). Grade control in multi-variable ore deposits as a quality management problem under uncertainty. *International Journal of Quality & Reliability Management, 32*(4), 334–345.

Li, S., de Werk, M., St-Pierre, L., & Kumral, M. (2019). Dimensioning a stockpile operation using principal component analysis. *International Journal of Minerals, Metallurgy and Materials, 26*(12), 1485–1494.

Liu, Y., Cheng, Q., Carranza, E. J., & Zhou, K. (2018). Assessment of geochemical anomaly uncertainty through geostatistical simulation and singularity analysis. *Natural Resources Research, 28*(1), 199–212.

Nguyen, H., Drebenstedt, C., Bui, X.-N., & Bui, D. T. (2019). Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network. *Natural Resources Research*. https://doi.org/10.1007/s11053-019-09470-z.

Nwaila, G. T., Zhang, S. E., Frimmel, H. E., Manzi, M. S. D., Dohm, C., Durrheim, R. J., et al. (2019). *Local and target exploration of conglomerate-hosted gold deposits using machine learning algorithms: A case study of the Witwatersrand gold ores*. South Africa: Natural Resources Research. https://doi.org/10.1007/s11053-019-09498-1.

Qu, M., Li, W., & Zhang, C. (2013). Assessing the risk costs in delineating soil nickel contamination using sequential Gaussian simulation and transfer functions. *Ecological Informatics, 13*, 99–105.

Rajabinasab, B., & Asghari, O. (2019). Geometallurgical domaining by cluster analysis: Iron ore deposit case study. *Natural Resources Research, 28*(3), 665–684.

Ruiseco, J., & Kumral, M. (2017). A practical approach to mine equipment sizing in relation to dig-limit optimization in complex orebodies: Multi-Rock Type, Multi-Process, and Multi-Metal Case. *Natural Resources Research, 26*(1), 23–35.

Ruiseco, J. R., Williams, J., & Kumral, M. (2016). Optimizing ore–waste dig-limits as part of operational mine planning through genetic algorithms. *Natural Resources Research, 25*(4), 473–485.

Sari, Y. A., & Kumral, M. (2018). Dig-limits optimization through mixed-integer linear programming in open-pit mines. *Journal of the Operational Research Society, 69*(2), 171–182.

Sepúlveda, E., Dowd, P., & Xu, C. (2018). Fuzzy clustering with spatial correction and its application to geometallurgical domaining. *Mathematical Geosciences*. https://doi.org/10.1007/s11004-018-9751-0.

Srivastava, R. M. (1987). Minimum variance or maximum profitability. *CIM Bulletin, 80*(901), 63–98.

Taguchi, G. (1986). *Introduction to quality engineering: Designing quality into products and processes*. The Organization: Tokyo.

Verly, G. (2005). Grade control classification of ore and waste: A critical review of estimation and simulation based procedures. *Mathematical Geology, 37*(5), 451–475.

Villalba Matamoros, M. E., & Kumral, M. (2019). Calibration of genetic algorithm parameters for mining-related optimization problems. *Natural Resources Research, 28*(2), 443–456.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing, 17*(4), 395–416.

Yan, D., Huang, L., & Jordan, M. I. (2009). Fast approximate spectral clustering. In *Paper presented at the proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, Paris.

Yanyan, M., Ferrie, F., & Dimitrakopoulos, R. (2015). Sparse image reconstruction by two phase RBM learning: application to mine planning. In *Paper presented at the 2015 14th IAPR International Conference on Machine Vision Applications* (MVA), 18–22 May 2015, Piscataway.

Zhang, G.-Y., Liu, G.-Z., & Zhu, H. (2011). Segmentation algorithm of complex ore images based on templates transformation and reconstruction. *International Journal of Minerals, Metallurgy, and Materials, 18*(4), 385–389.