24-092

Utilizing Big Data Statistical Techniques in Python to Optimize Geometallurgy Workflow for Metallurgical Test Work Sample Selection

Muhammad Usman Siddiqui
Ausenco, Vancouver, BC

Kevin Erwin
Ausenco, Vancouver, BC

Rajiv Chandramohan
Ausenco, Vancouver, BC

Connor Meinke
Ausenco, Vancouver, BC

Shaihroz Khan
Ausenco, Toronto, ON

ABSTRACT

High-quality sample selection for metallurgical test work is essential to a geometallurgy study, but the large multi-dimensional dataset makes sample selection a daunting task, as classifying the dataset while respecting its heterogeneity is difficult. This paper presents a streamlined approach for sample selection, utilizing custom-built tools in Python to standardize the methodology, saving time and costs. This approach uses the cumulative sum method, principal component analysis, and k-means clustering method to elegantly cluster the data and select representative samples. A case study is used to demonstrate the effectiveness of the methodology by selecting 40 samples for flotation test work.

NOMENCLATURE

PCA—Principal component analysis
CuSum—Cumulative sum
PC—Principal component
WCSS—Within-cluster sum of square

INTRODUCTION

Geometallurgy is a compelling methodology in the mining industry to bridge the gap between metallurgy and geology. It aims to reduce technical risk and enhance the economic performance of a mineral processing plant by accounting for the variability in a deposit to strengthen investor confidence, facilitating robust revenue models, and developing scenarios with variable throughput rates to more accurately forecast cash flows for future mining operations. A geometallurgy study looks at the relationship between a deposit's geological characteristics, its variability, and its response to metallurgical processes. Selecting samples that capture the heterogeneity of the deposit for metallurgical test work is an essential component of a geometallurgy study.

High-quality sampling is vital across the entire mine value chain, as sampling errors are additive and generate monetary and intangible losses (Dominy et al., 2018). The goal is to select samples that accurately describe the deposit (Dominy et al., 2018). A geometallurgy database usually consists of sample id, mineral grade, lithology, alteration, and test work data if available. It is the basis for choosing representative samples for metallurgical characterization test work (competency, hardness, recoveries) which is then used in robust flowsheet development and equipment selection for optimum life of mine performance. However, these datasets can be large with multiple columns, making sample selection a daunting task during the analysis.

Michaux et al. (2020) present a framework to develop a geometallurgical program in which they discuss the methodology to cluster a geometallurgy database into similar clusters, and they present the cumulative sum (CuSum)

method to smoothen the noisy assay grade data. However, their work is a general guideline for a geometallurgy program and not a streamlined approach for sample selection. The k-means method is a popular unsupervised machine learning clustering technique which can be applied for geometallurgical clustering, but its sole application is quite limited as it does not capture the outliers properly (Potakey et al., 2022). A way to increase the accuracy of the clustering exercise is by reducing the dimension of the database by applying the principal component analysis (PCA) method before running the k-means algorithm.

Selecting representative samples for metallurgical test work is a challenging task. Therefore, a streamlined and efficient sample selection process utilizing modern-day tools such as Python is valuable. This paper builds on the works done by Michaux et al. (2020) and Potakey et al. (2022) and presents a streamlined process for sample selection used to select 40 samples with a minimum mass of 20 kg for flotation test work for the pre-feasibility study of a copper mine. Validation steps are built into the methodology to ensure high-quality samples are selected, ensuring the representation of the deposit. The sample selection process utilizes Python to efficiently analyze, visualize and apply statistical techniques to the large database. Writing custom Python code for the process allows the process to be standardized and reproducible and enables multiple iterations in a brief time resulting in time and cost savings.

METHODOLOGY

Minerals with the most significant effect on test work are assigned to three buckets: key, deleterious, and rock elements. Key minerals are the most valuable minerals from the mine and are most important in mineral processing to maximize the revenue. Deleterious minerals negatively impact the commercial value of the product or the environment. Rock minerals are the minerals in the host rock that are not valuable and thrown away. Rock minerals are essential for comminution test work samples to test the breakage behavior of the ore but may not be necessary for flotation test work samples aiming to test the ore's chemical behavior. In this study, the samples were selected for flotation test work. Therefore, the key elements considered were copper, iron, sulfur, and zinc and deleterious and rock minerals were not considered.

The CuSum analysis is run on the mineral grade values of the geometallurgy dataset and a column containing the CuSum values, a column containing the difference between each grade value and the column average, and a column (Delta CuSum) containing a value of 1, if the grade value is greater than the average or a value of –1, if the grade value is lower than the average are appended to the dataset in Jupyter Notebook. The method generally involves quantifying changes in lithology and assigning a numerical value to each section with the same lithology. For this study, lithology was not considered upon recommendation of geologists and metallurgists.

The data is sectioned into CuSum intervals to quantify the heterogeneity change down the drillholes. For the main mineral copper, the r-squared values of three consecutive grade CuSum data points are calculated across the entire length of the column, and consecutive data points for which the r-squared value stays greater than specified, 0.9 in this study, are grouped as one CuSum interval. This method is susceptible to grade changes and only applies to the main mineral.

For other minerals (iron, sulfur, and zinc), the Delta CuSum values are used to assign CuSum intervals. Grade values that are consecutively greater than the average (assigned a Delta CuSum value of 1) or lower than the average (assigned a Delta CuSum value of –1) are grouped as one CuSum interval. This method is less sensitive to grade changes and, therefore, is used for less important minerals.

The CuSum intervals for each assay are combined into overall CuSum intervals where a change in CuSum interval of one assay dictates the change in the overall CuSum interval, and a numerical value is assigned to each overall CuSum interval. Generally, the change in lithology also dictates the change in the overall CuSum interval, but lithology was disregarded in this study. CuSum intervals are scrutinized by geologists and metallurgists by plotting the CuSum curves of the relevant minerals and assessing the CuSum intervals with regards to each mineral and by plotting the normalized CuSum curves of all appropriate minerals and evaluating the overall CuSum intervals (Figure 1). Python functions were written to automate the tasks, including cleaning and preparing the geometallurgy dataset, performing the CuSum analysis, assigning CuSum intervals, clustering the data and plotting the curves.

After defining the CuSum intervals, a cutoff grade may be applied to filter out the low-grade values. The filtered database is then condensed by averaging all the grade values in each CuSum interval. The PCA technique prepares the database for clustering by reducing its dimensions. Rows in the database with no values are dealt with before performing the PCA. Rows with all null values and null values for non-important minerals are dropped. Suppose there are rows with null values for essential minerals. In that case, the database is divided into sections so that each section contains a combination of mineral grades with no null values and is clustered separately (Table 1).

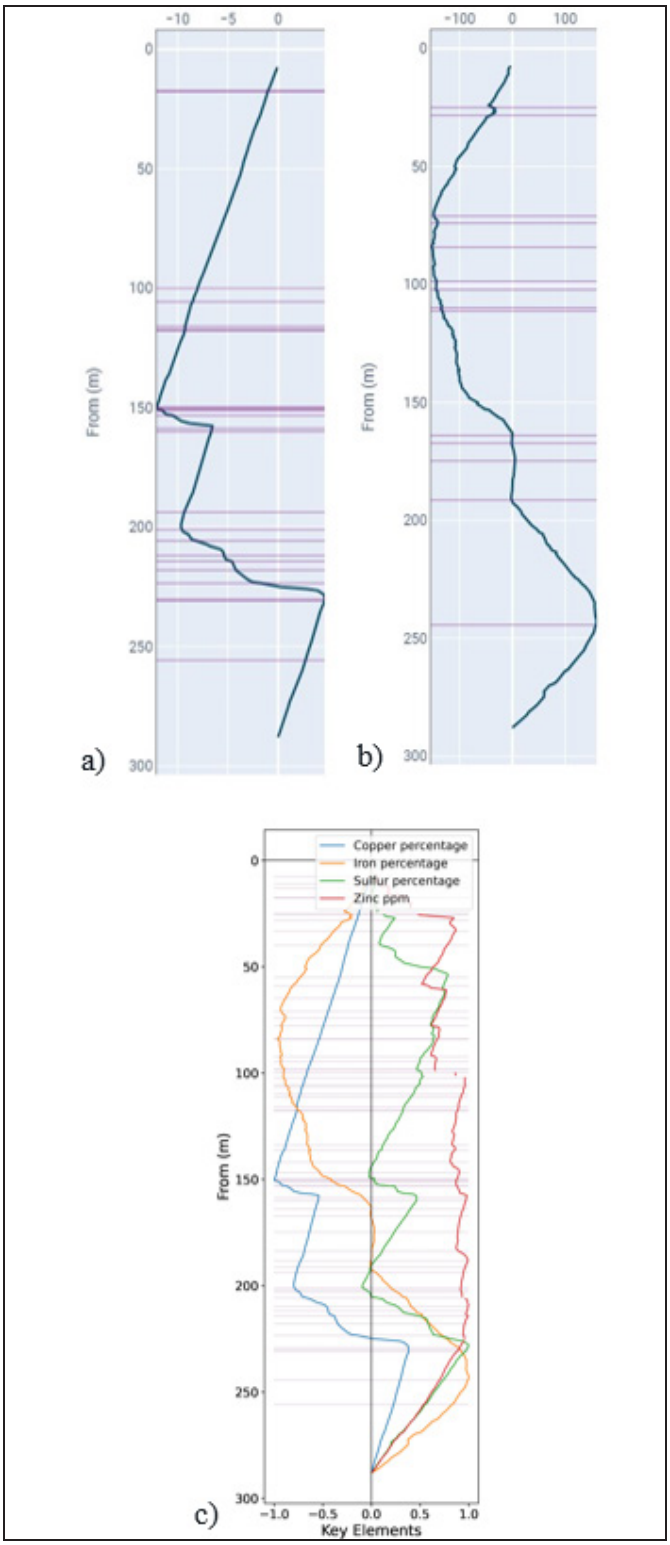

Figure 1. Plotting CuSum intervals on mineral CuSum curves, a) higher sensitivity CuSum intervals overlayed on copper percentage CuSum curve, b) lower sensitivity CuSum intervals overlayed on iron percentage CuSum curve, c) overall CuSum intervals overlayed on copper, iron, sulfur, and zinc normalized CuSum curves

The PCA technique is performed on each database section using the PCA function in the scikit learn library in Python. The number of principal components (PCs) required is determined using the percentage variability each PC captures, and the number of PCs which cumulatively capture greater than eighty percent of the data's variability is chosen.

The k-means clustering function requires the number of k-means clusters the analyst wants as an input. The number of suitable k-means clusters for the dataset is determined using the elbow method. The within-cluster sum of square (WCSS) vs number of clusters curve is plotted for each section of the database (Figure 2). The number of clusters after which the curve becomes linear is chosen, which is eight in this case (Figure 2). The PCs data frame, which is output from the PCA function, is fed into the k-means function and the number of clusters is specified to obtain the clustered database. The function appends a column to the database and assigns a numerical value to all the rows in one cluster. The clustered data frames for all the different sections are concatenated into one database.

Table 1. Combination of columns with null values in their rows. False represents when the column's rows do not contain null values, and True is when they have null values. Each combination is clustered separately.

Copper Grade is Null	Iron Grade is Null	Sulfur Grade is Null	Zinc Grade is Null
False	False	False	False
False	True	False	False
False	True	False	True
True	False	False	False

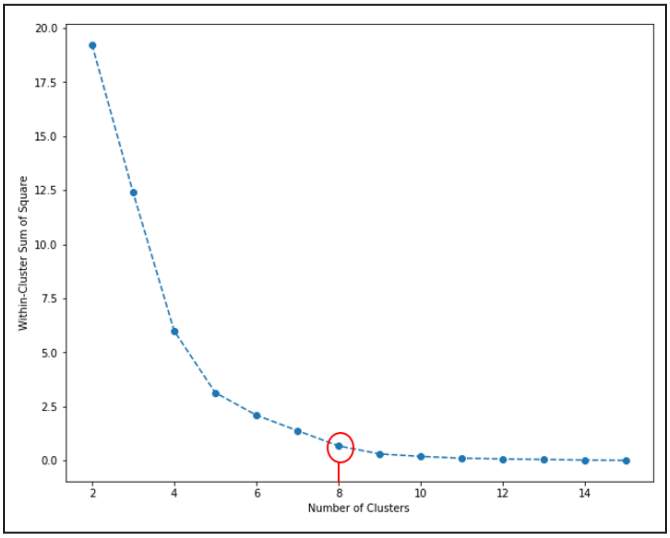

Figure 2. The elbow method to determine the number of suitable k-means clusters, 8 in this case

The mass of each length interval of a drillhole is calculated based on the diameter of drill core used for drilling and is important to ensure the samples selected meet the required mass for test work. The number of required samples are based on the recommendation of metallurgists and geologists. The percentage proportion of mass of each cluster is calculated and multiplied by the total number of samples required to determine the number of samples needed from each cluster. Some clusters may have an insignificant mass proportion, so no samples are selected from them. The required number of samples are selected from each cluster ensuring each sample has sufficient mass for test work.

An ideal sample will have sufficient mass, belong to the same CuSum interval, and belong to the same k-means cluster. After checking for similarity, samples from different CuSum intervals may be combined to meet the mass requirement. It is not advisable to combine samples from other clusters. The selected samples are compiled in another database and scrutinized by metallurgists and geologists to ensure the heterogeneity of the deposit is sufficiently captured. In this study, 40 samples with a minimum mass of 20 kg were elegantly selected using this methodology. Each sample was part of a single CuSum interval and cluster, and merging samples between CuSum intervals was not necessary.

Inspecting the Clusters

The clusters are inspected to ensure the number of clusters chosen is accurate and outliers do not significantly affect the clusters. Scatter plots showing all the clusters with different features on the axes are used to visually assess the cluster boundaries, ensuring similar data is clustered together (Figures 3, 4, and 5). Plotting with different features on the axes is important because multiple features were considered in the clustering exercise, and it is important to see the cluster boundaries with respect to each feature.

Box and whisker plots are used to assess the distributions of each mineral in a cluster (Figure 6). Box and whisker plots are a good way to visualize the outliers in each cluster and see if there are clusters which may be combined or broken down into more clusters.

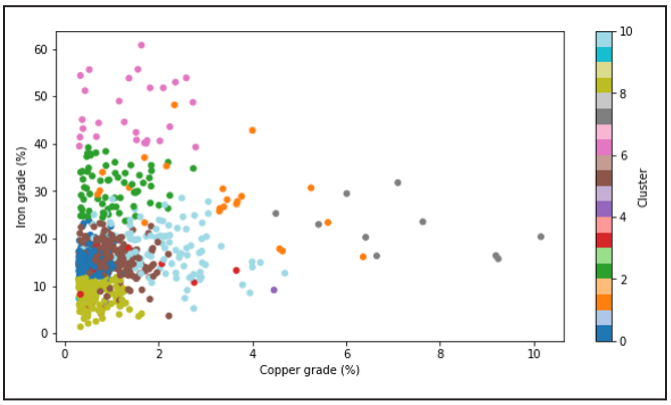

Figure 3. Scatter plot displaying the clustering results with respect to the iron grade and copper grade.

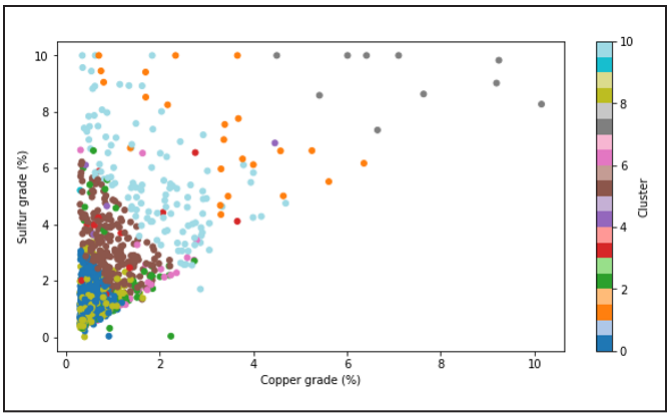

Figure 4 . Scatter plot displaying the clustering results with respect to the sulfur grade and copper grade.

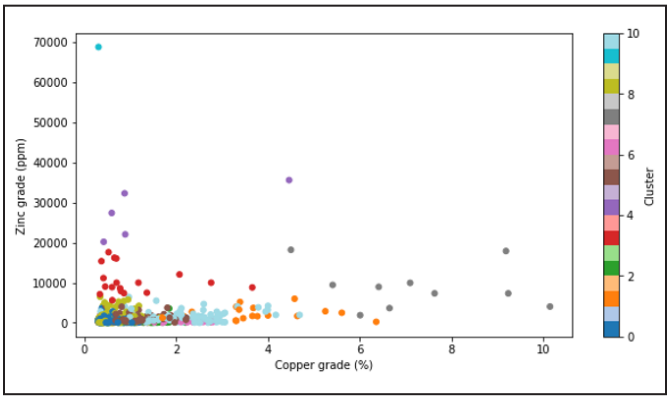

Figure 5. Scatter plot displaying the clustering results with respect to the copper grade and zinc grade.

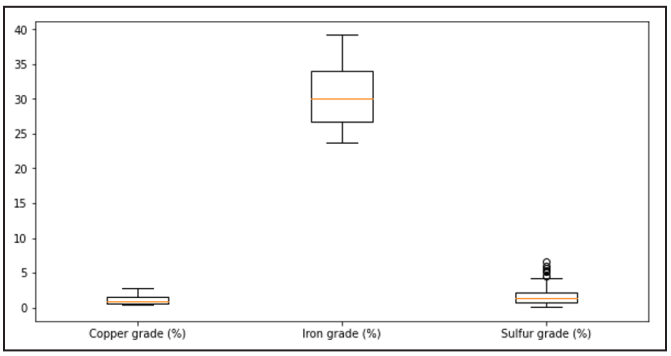

Figure 6. Box and whisker plot for one cluster displaying the distribution of data with regards to each mineral.

In this study, the clusters were deemed suitable for sample selection on the first pass of the clustering exercise. However, some clusters only contained extreme grade values and, therefore, were not used for sample selection to prevent skewing test work results. There is potential to improve the clustering methodology to capture the outliers better using supervised machine learning techniques. However, using supervised machine learning techniques comes with cost and skill limitations, as training the models requires time and specialized skills.

Validating Samples Selected

The variation in copper grade in the raw data set was compared to the variation in copper grade in the samples selected by comparing their histograms. The two histograms (Figures 7 and 8) have a similar trend as most of the samples chosen have a copper grade lower than 2.71% similar to the raw drillhole data, and there is a low number of high-grade values in both histograms.

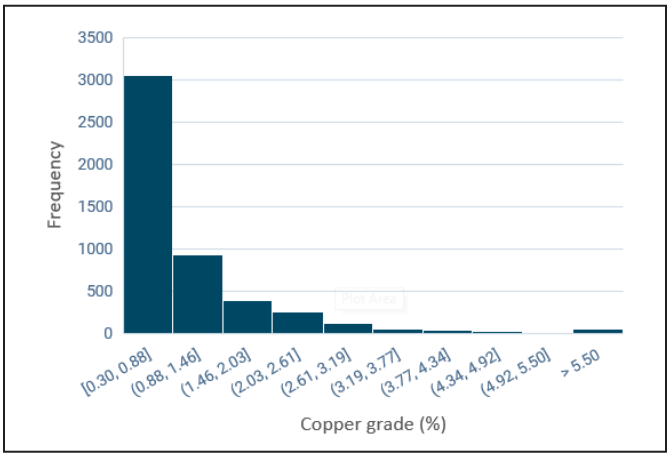

Figure 7 . Variation in the grade of main metal cooper in the raw dataset.

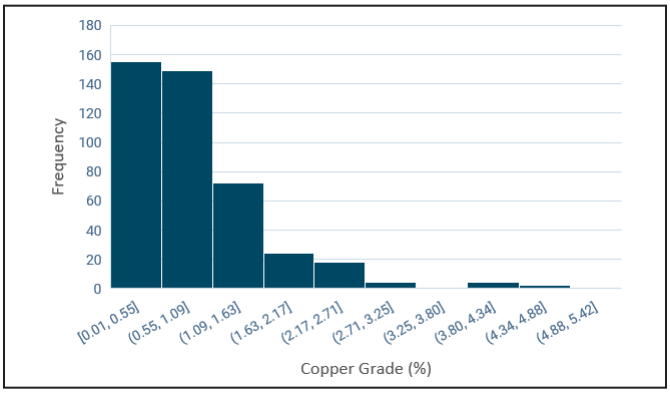

Figure 8. Variation in the grade of main metal copper in the samples selected.

The variation in lithology of the samples selected was compared to the raw data to check for similarity (Tables 2 and 3). When the variation in lithology of the raw data is similar to that of the samples selected, the samples selected can move on to the next stage. In this study, good alignment between the lithology variation in the raw data and the samples selected was achieved on the first pass. The selected samples and the rationale for choosing them are presented to metallurgists and geologists, and once all are convinced the selected samples are sent for test work.

Table 2. Lithology variation of raw data

Rock Type	Percentage
B3_Magnetite_carbonate Ore	22.61
B1_Biotite_calcsilicate Ore	18.41
B2_Calcsilicate_carbonate Ore	15.85
NOC	0.47
BAS	0.23
A4_Basaltic_carbonate-calcsil. Ore	4.20
A1_Graphitic Ore	4.43
CBBX	3.03
MAF	0.47
GDI	1.17
D1_Amphibole_talc_mag. Ore	3.50
A3_Tuffitic_carbonate-calcsil. Ore	11.19
O_Oxidised Ore	2.80
MSE	3.50
D2_Magnetite Ore	5.13
CLA	0.93
SCB	0.23
GAB	1.86

Table 3. Lithology variation of samples selected

Rock Type	Percentage
B3_Magnetite_carbonate Ore	22.10
B1_Biotite_calcsilicate Ore	16.91
B2_Calcsilicate_carbonate Ore	12.90
NOC	0.18
BAS	0.87
A4_Basaltic_carbonate-calcsil. Ore	1.54
A1_Graphitic Ore	3.83
CBBX	1.40
MAF	0.63
GDI	0.97
D1_Amphibole_talc_mag. Ore	5.11
A3_Tuffitic_carbonate-calcsil. Ore	8.90
O_Oxidised Ore	1.07
MSE	4.16
D2_Magnetite Ore	9.79
CLA	0.55
SCB	0.93
GAB	1.26

TIME AND COST SAVINGS

The methodology presented enables significant time savings which translates to cost savings. It takes time to write code in Python initially to standardize the procedure, but once it is completed the process is efficient and cuts down time from an average of 1200 seconds per drillhole for data classification and 12 hours for handpicking samples to around 60 seconds per drillhole for higher-quality data classification and 8 hours for handpicking samples from the classified dataset. Initial data processing might differ for each dataset, as multiple files may have to be combined and null rows may have to be dealt with, but after that a standard procedure is followed in which the cleaned database is fed into Python functions to obtain the clustered database from which samples are selected. The time savings exponentially increase as the number of drill holes increases. This allows for multiple iterations of sample selection quickly if needs change during the study.

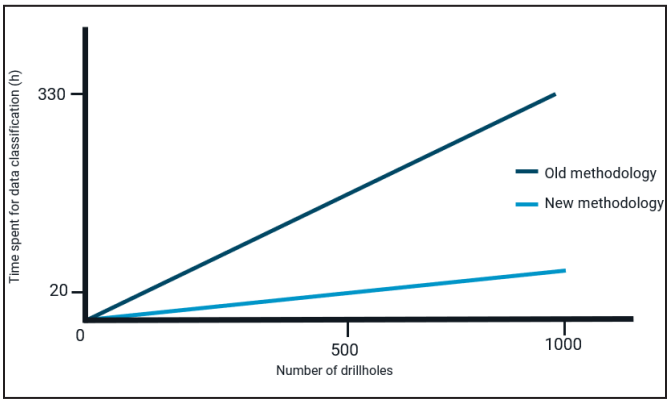

Figure 9 . Comparing total time spent for data classification using the old methodology and the new methodology

CONCLUSIONS

Selecting samples which are representative of the ore body for metallurgical test work is an important part of a geometallurgical study. A streamlined and standardized sample selection methodology utilizing modern-day tools such as Python is presented in this paper. The methodology was used in a pre-feasibility study of a copper mine to select 40 samples with a minimum mass of 20 kg for flotation test work. It resulted in grouping sections of drillhole lengths with similar characteristics and elegantly selecting samples with sufficient mass from those groups. Automating the process resulted in significant time savings from an average of 1200 seconds per drillhole to around 60 seconds per drillhole for data classification and from 12 hours to 8 hours for handpicking samples from the classified dataset. There is potential to use supervised machine learning techniques in the clustering exercise of the methodology to capture outliers more accurately, but cost and skill limitations may have to be overcome as training the models requires time and expertise. As with any sample selection program, eventually, the test work results dictate if the samples selected resulted in the geometallurgy program's desired outcome, but this methodology allows multiple iterations quickly.

REFERENCES

[1] Potakey, N. E., & Ortiz, J. M. Defining geological units using geochemical data and unsupervised machine learning. Predictive Geometallurgy and Geostatistics Lab, 47.

[2] Michaux, S., & O'Connor, L. (2020). How to set up and develop a geometallurgical program. Geological Survey of Finland.

[3] Dominy, S. C., O'Connor, L., Glass, H. J., Purevgerel, S., & Xie, Y. (2018). Towards representative metallurgical sampling and gold recovery testwork programmes. Minerals, 8(5), 193.

[4] Jansson, N. F., Allen, R. L., Skogsmo, G., & Tavakoli, S. (2022). Principal component analysis and K-means clustering as tools during exploration for Zn skarn deposits and industrial carbonates, Sala area, Sweden. Journal of Geochemical Exploration, 233, 106909.

[5] Lishchuk, V., Koch, P. H., Ghorbani, Y., & Butcher, A. R. (2020). Towards integrated geometallurgical approach: Critical review of current practices and future trends. Minerals Engineering, 145, 106072.

24-093

Validation of Modeled Rockmass Permeability Against Field Measurements in a Longwall Mine

Zoheir Khademian
CDC NIOSH, Pittsburgh Mining Research Division, Pittsburgh, PA

Steven J. Schatzel
CDC NIOSH, Pittsburgh Mining Research Division, Pittsburgh, PA

Marcia M. Harris
CDC NIOSH, Pittsburgh Mining Research Division, Pittsburgh, PA

Kayode M. Ajayi
CDC NIOSH, Pittsburgh Mining Research Division, Pittsburgh, PA

ABSTRACT

Predicting rockmass permeability is critical in evaluating various engineering designs, including estimating gas inflow to a longwall mine in the case of a hypothetical breach in the gas well. This study conducted field permeability measurements to validate a geomechanical model capable of predicting rockmass permeability during longwall mining. A series of slug permeability tests were conducted in an active mine in Pennsylvania. A model of the mine was constructed in 3DEC numerical modeling software, and permeabilities were calculated. The modeling results agreed well with the pre- and post-mining permeability measurements, showing the applicability of this tool to evaluate gas well stability near mine workings.

INTRODUCTION

The intrinsic permeability of a rock or soil is a measure of the rock or soil's ability to transmit fluid as the fluid moves through it (Schwartz and Zhang, 2003). Evaluation of rockmass permeability is essential in various rock engineering designs such as constructing dams (Thawatchai, Bunpoat, & Warakorn, 2021), tunnels (K. Zhang et al., 2021), geothermal reservoirs (Tomac & Sauter, 2018), oil and gas reservoirs (Gehne & Benson, 2019), and waste containment structures (Sasaki & Rutqvist, 2021). Another application is the evaluation of permeabilities enhanced by mining where shale gas reservoirs underlie active or future longwall mining operations (PADEP, 2018). In the Northern Appalachian Basin in the United States, some of the production wells in shale gas reservoirs intersect with minable coal seams and thus require specific design considerations to allow both operations to coexist. In most cases, the gas wells are positioned in the mine abutment pillars for protection against mining-induced ground deformations. However, one of the safety concerns is excessive ground movement that might damage gas well production casing, leading to the leakage of explosive gas into the mine working. Under such scenario, mining-induced rockmass permeabilities and relevant changes induced by mining becomes of main importance.

In previous work (Khademian, et al. 2021)(Khademian, et al. 2021), a geomechanical modeling methodology was developed based on the Discrete Fracture Network (DFN) technique to calculate the permeability evolution during mining of a shallow, 145-m-cover mine in the Pittsburgh coal seam. In-situ measurements were used to constrain and calibrate the range of DFN parameters in the model. The methodology was then validated against field permeability measurements in a deeper mine, at a 341-m-cover site in the Pittsburgh seam (Khademian et al., 2022). In this paper, another case is studied in a 352-m deep mine, and field permeability measurements are conducted before