Implementation of Statistical Analysis to Determine Optimum Ultimate Pit

Tawum Juvert Mbah* , **Haiwang Ye, Jianhua Zhang, Mei Long**

Department of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan, China
Email: *juvertmbah@whut.edu.cn

Abstract

The ultimate pit may affect other aspects in the life of a mine such as economical, technical, environmental, and social aspects. What makes it even more complex is that most often there are many pits which are economically minable. This calls for a heuristic approach to determine which of these pits is the ultimate pit. This study presents a means of selecting an ultimate pit during design operations of the Hebei Limestone mine. During optimization processes of the mine, many pit shells were created using Whittle Software. Normally, Whittle Software optimizes these processes and assigns a revenue factor of 1 for the ultimate pit. Unfortunately, the pit shells created did not satisfy the criteria with a revenue factor of 1 based on the parameters. As a result of this, statistical analysis was implemented to further understand the relationship, variability, and correlation of the pit shells created (data). Correlation Analysis, K-means++ Analysis, Principal Component Analysis, and Generalized Linear models were used in the analysis of the pit shells created. The results portray a salient relationship of the optimization variables. In addition, the proposed method was tested on Whittle Sample projects which satisfy the selection of ultimate pit selection with a revenue factor of 1. The results show that the proposed model produced almost the same results as the Whittle model with a revenue factor of 1 and was also able to determine the ultimate pit in cases which did not satisfy the Whittle selection criteria.

Keywords

Ultimate Pit, Statistical Analysis, Pit Shells, Variables, Mine Planning, Clusters

1. Introduction

The selection of the ultimate pit is a very important step in mine planning. Several researches have been made in determining the ultimate pit [1] [2]. The

techniques in ultimate pit selection may vary depending on the mineral resource model, location, estimation techniques, economic parameters, technical parameters, and data analysis. The ultimate pit may vary if a parameter is changed. For example, if the economic parameter is changed, the ultimate pit will be the pit that satisfies the defined parameters. As such, before computing for an ultimate pit, there is a need for a proper evaluation of these parameters by experts and a final decision by the administrative board [3]. Before the determination of the ultimate pit, the very first step involves the characterization of the reserve followed by preliminary assessment, and feasibility studies [4]. These processes include exploration, geological modelling, mineral resource estimation, ore reserve estimation, and statistical analysis. Exploration is the process of finding an economically viable ore to mine. The exploration process consists of geological mapping of the area, geochemical exploration, and geophysical exploration and drilling of the mine area to collect samples [5]. The samples collected are used for sampling and analysis of the drill hole data. This data is then used to create a geological database. Geostatistical analysis can be done from the database to understand the geological structures and spatial relationships that constitute the mining area. The variogram analysis provides results to a variogram map which explains the anisotropy of the data and continuity within a region. The report from geostatistical analysis can then be used in geostatistical estimations. Geostatistical estimations can be done by inverse distance estimation, ordinary kriging, indicator kriging, and nearest neighbor [6] [7]. The block estimates can then be used to classify and quantify the ore into a mineral resource [4]. The resulting mineral reserve can then be converted into a block model by means of mathematical functions under standard mining conditions (metallurgical, economic, technical, processing and other factors) [8]. The block model can then be used to create and calculate attributes which in turn will be used during estimation or filling of the block model, cut-off grade, and column processing [9]. The block model can be exported to Whittle Software after geostatistical estimation using Surpac Gemcom software. Optimization parameters are configured and pit shells are created. Further statistical analysis can be done on the pit shells created during optimization to determine the ultimate pit of the mine.

The technique used in statistical analysis depends on certain factors such as data type, variables (parameters), type, and the model of statistical analysis. The most common types of analysis are univariate, bivariate and multivariate analysis [10]) [11]. Univariate analysis consists of statistical summaries (mean, standard deviation, etc.) where the data is visualized as histograms and probability plots. Bivariate analysis consists of correlation analysis, scatter plots and regression analysis. Multivariate analysis consists of multiple regression analysis, e.g. bulk density-metals and multiple variable plots e.g. triangular diagrams. Correlation Analysis, Principal Component Analysis (PCA), K-means++ clustering, and Gaussian Linear Models (GLM) were used to display and analyze the results of pit shells. Correlation analysis was done to determine the type of relationship between the variables and which variables show positive or negative correlations.

Principal component analysis was used to investigate the correlation and the line of best fit. The PCA creates new axes for better display and interpretation of the results. K-means++ clustering classify or group the data based on the difference and similarities of their features. K-means++ is an unsupervised algorithm used in statistical analysis to classify or group data based on their similarities or features. The "k" in the name stands for fixed number of clusters in the data [12]. Cluster-analysis has been implemented in different fields such as geology, medicine, and economics [13]. The GLM model was used to understand the association between the variables as well as predict the ultimate pit based on the defined variables.

The mine is located in Hebei province in the district of Yanshan. The topography consists of low mountains and hills. Nikon Total Station was used to measure the coordinate points. The formation of the strata is Cambrian (Gushan bands, long Sankai, Fengshan) at the upper strata, Ordovician and Quaternary. The mining area is dominated by undeveloped folds and faults. The orebody is different in composition due to a stratigraphic and lithological difference associated. Its mineral composition, texture, and structure are different as well. The chemical composition is substantially the same in a given area and varies with the different types of ore. The main chemical composition is Calcium Oxide (CaO), and small amounts of Magnesium Oxide (MgO), Silicon dioxide (SiO_2), Aluminum oxide (Al_2O_3), Iron Oxide (Fe_2O_3), Potassium Oxide (K_2O), Sodium Oxide (Na_2O), phosphorus pentoxide (P_2O_5), Sulphur Trioxide (SO_3), and Chlorine (Cl^-).

This study presents the determination of ultimate pit by statistical analysis. The exploratory data was collected from the Hebei Limestone mine. The database was created from the exploratory data and the block model from the database. The block model was then exported to Whittle Software for optimization processes. During the optimization processes, pit shells were created. The pit shells created were all economically minable. With respect to this, statistical analysis was implemented in the determination of the ultimate pit.

2. Methodology

Exploratory analysis was done on the mine data and the results were used to create a database. A block model was further created from the database under standard mining requirements. The block model was exported to whittle for optimization. Optimization parameters were configured and pit shells were created. Further statistical analysis was implemented on the pit shells to select the ultimate pit of the mine.

2.1. Geological Database

The database was created using SURPAC software which uses a relational database that contains tables or items. The database was created using the raw drill hole data. The database contains the collar, survey, geology and the assay tables.

The first 30 values of each table have been uploaded on Github for viewing purposes. The collar table gives the location of the hole in space (collar table). A downhole survey was done and readings of the dip and azimuth plotted in the reflex software to show how the actual hole in the field deviates from the plan (survey table). A metallurgical test was done on the sample and the good values were recorded in the assay table (assay table). Geological analysis was done on the samples to record the weathering state, lithology, sample recovery, moisture content, etc. and results recorded in the geologic table (geology table). The geologic database is created by importing all the tables from exploratory analysis using Surpac software. The above process was done on standard mining conditions to meet the requirement of creating a geologic database [14]. From the database, the drill holes were sectioned and digitized using the graphical method to provide mirror images of the shape and size of the ore depending on the geological interpretation of the data (image).

From the sections, triangulation was done and a digital terrain model (DTM) of the orebody created as shown in Figure 1(a). The drill holes were partitioned by compositing for easy calculation of grade Figure 1(b). For a dataset to produce good results, the outliers need to be removed or down weighted [15]. In this study, the outliers were removed by means of geostatistical analysis using a histogram plot. Figure 2 shows a histogram which was modified to mark the outliers using a blue circle and a blue rectangle. The outliers were removed by applying a cutoff of 22.5 using the "iff" formula (outliers removed).

2.2. Block Model Creation

The block model is a form of a spatially-referenced database that provides a means for modelling the 3D Body points and intervals of the drill hole sample data. It provides a method for estimating volume, tonnage, and average grade of the 3D body from parse drill hole data. The blocks are arranged based on define coordinates which are minimum and maximum of Northing (Y), Easting (X), and Elevation (Z). The centroid of each block defines its geometric dimensions in each axis, that is, its Y, X, and Z coordinates. Blocks can be of varying sizes, the sizes of the blocks in this block model is $10 \times 10 \times 5$ and a total number 1,123,402 blocks. An attribute is a model that contains properties or information of the model space. It can vary from characters, numbers, decimals, integers or even character code. A total of 35 attributes were created for the block model (attributes). The block model was constrained to control the selection of blocks from which the information was retrieved to make interpolations. Estimation of the block model was done by Ordinary Kriging using results from the variogram study. Figure 3 shows the constrained block model colored by limestone grade attribute. Other attributes such as rock, zone, and economic attributes (mining costs, and processing costs) were added to the block model to meet Whittle Software standards. All negative grades and air blocks error were checked. A block model report was used to validate the block model. The block model was then exported to Whittle (Block Model Validation Report).
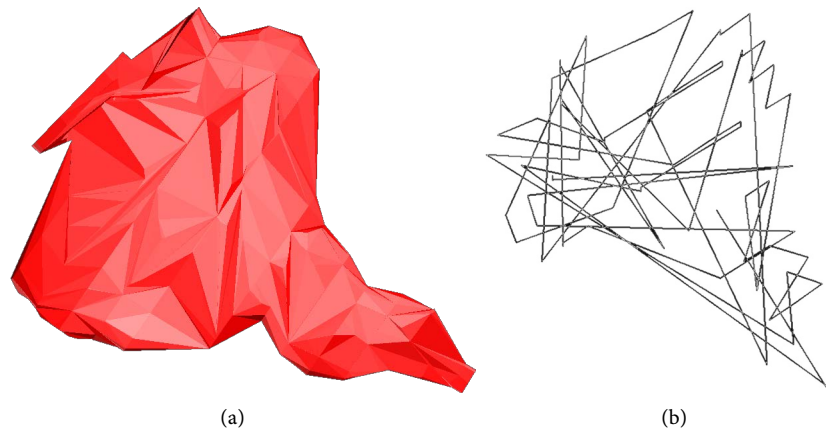
(a)   (b)

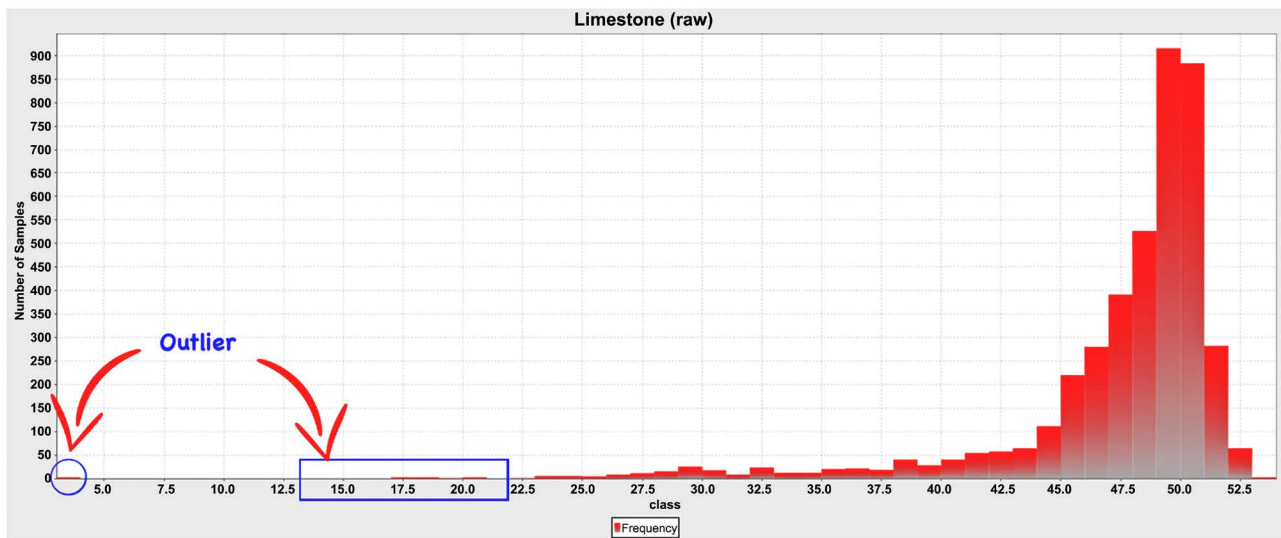**Figure 1.** (a) Triangulated ore save as .dtm; (b) Composite of the orebody.

Figure 2. Outliers marked by a blue circle and a blue rectangle.
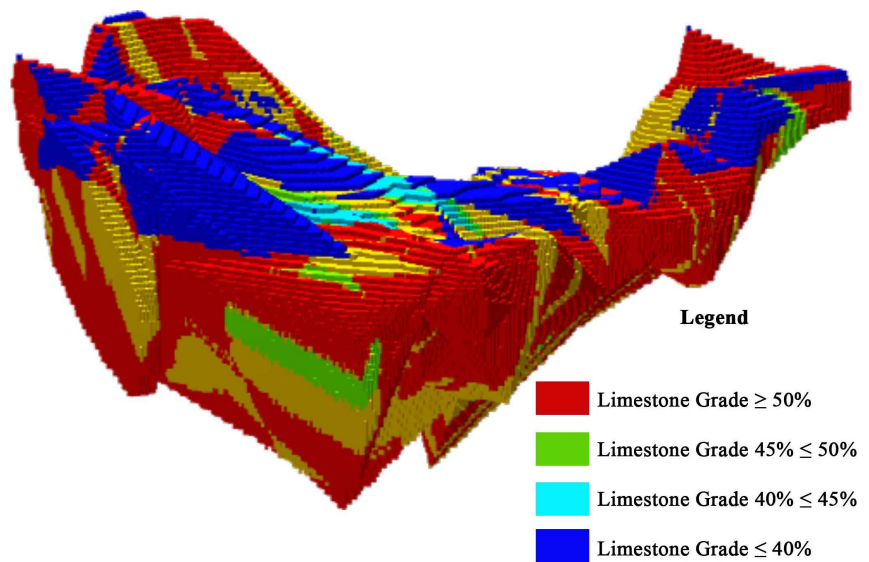

Figure 3. Block model constrained by limestone grade attribute.

2.3. Optimization in Whittle

The block model was imported into Whittle Software for optimization purposes. Table 1 shows the optimization parameters used during the optimization process. During the Optimization process, several revenue factors were investigated for best results. A revenue factor (RF) is a variable which when multiplied with other parameters such as metal price will produce an outline for each value which can be used to create pit shells. The RF are either in range or single values and can be done in two methods which are the fixed intervals and geometric intervals [16]. In Whittle, the optimal pit shell is the pit shell with a revenue factor of 1. In the scope of this work, the pit shells created did not satisfy this rule as there was no exact RF of 1. The Hebei Limestone optimization results can be found on this link. The Marvin blend optimization results can be found on this link (An internet connection maybe needed to view the tables). As a result of this, further investigations were applied to determine the optimal pit.

3. Statistical Analysis of Pit Shells

Four different statistical analytical models were implemented in analyzing the pit shells created during optimization processes. The pit shells created were recorded in a comma separated value file (.csv). The columns of the .csv file are considered variables in this study. The pit shells created was recorded after using a fixed interval revenue factor of 0.3 - 2 (pit shells created). It was the best result after using several revenue factors. From the results, it can be seen that there's no pit with an exact revenue factor of 1. Normally in Whittle, the optimum pit is that which satisfies a revenue factor of 1. Unfortunately, this condition was not met based on our optimization criteria. As a result of this, statistical analysis methods were implemented to determine the ultimate pit.

Table 1. Optimization parameters.

Parameters	Hebei limestone Mine	Marvin Gold/Copper
	Value (¥)/(tonne)	Value ($) (AU/Oz) (CU/ton)
Selling Price	150¥	12.00/0.907
Selling Cost	5¥	0.20/0.326
Reference Mining cost	11¥	0.9
Mining recovery fraction	0.95	1
Mining Dilution Fraction	1.05	1
Processing Cost	30 - 45	4
Processing Recovery	0.95	0.6/0.8
Slope Angle	45	40/45
Discount rate	10	10
Revenue Factors	0.3 to 2 Using 99 FF	0.5 to 2, using 54 FF

3.1. Correlation Analysis (CA)

Correlation analysis was done to better understand the relationship between the variables of the optimization process. Correlation is the strength between two or more variables [17]. The data was normalized to −1 and 1. −1 means that the correlation is negative and 1 means the correlation is positive. The most common correlation methods are Pearson, Spearman, and Kendall. In this study, the Spearman correlation analysis was used to investigate the pit shells created during optimization data. The Spearman correlation assumes that the relationship between the variables is monotonic (an increase in the value of one variable will result to an increase/decrease of the other variables), they are measured ordinally, internal or ratio scale and that the observations are independent [18]. It also measures the direction and strength associated to the variables. The data is normalized to +1 and −1. The rank value can be +1 to −1. A +1 indicates a perfect correlation, 0 means there is no correlation and −1 means it has a negative correlation [19]. Equation (1) shows the Spearman's correlation coefficient where $R(y)$ is the rank of y_i, $R(x)$ is the rank of x_i, $\overline{R(x)}$ is the mean of x, $\overline{R(y)}$ is the mean of y, and "n" is the number of pairs (variable pairs). The main purpose of the correlation analysis was to determine which pair of variables shows high or positive correlation. The results are outlined in Section 4.1.

$$\rho = \frac{\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)\left(R(y_i) - \overline{R(y)}\right)}{\sqrt{\sum_{i=1}^{n}\left(R(x_i) - \overline{R(x)}\right)^2 \cdot \sum_{i=1}^{n}\left(R(y_i) - \overline{R(y)}\right)^2}}$$
$$= 1 - \frac{6\sum_{i=1}^{n}\left(R(x_i) - \overline{R(y_i)}\right)^2}{n\left(n^2 - 1\right)}$$

(1)

3.2. Principal Component Analysis (PCA)

PCA is a statistical analysis algorithm used in dimensionality reduction of data with many variables while minimizing the loss of information for better interpretation [20]. Firstly, the algorithm creates new variables called 'Principal Components' (PC) from the initial variables by mixtures or linear combinations [21]. More weight or information is given to the first PC than the second and so forth. There are five major steps in principal component analysis which are standardization, covariance matrix computation, identification of PC by Eigenvalues and Eigenvectors, feature vector, and reorientation of the data to PC axes. The data was normalized to the same scale by standardization in order to prevent biased results Equation (2). The covariance matrix was used to understand the relationship between the variables. The eigenvectors represent the direction of the axes with the most information (variance) of the covariance matrix while the eigenvalues are the weight or coefficient of the eigenvectors. The feature vector is a matrix that has the values of the eigenvectors with the highest information. In the last step, the algorithm uses the eigenvectors of the covariance matrix and orientates it to the principal component axes by multiplication of

feature vector and transpose of the original standardized dataset as shown in equation (3). The first PC known as principal component one (PC1) is the line which maximizes the variance and the second (PC2) is calculated perpendicularly from the first and so on. The data was divided into five clusters each with a different color and four PC axes which represent the revenue factors, stripping ratio, ore tones, and limestone grade. The newly created values were plotted on the four PC based on their weight. The results are outlined in Section 4.2.

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} \tag{2}$$

$$\text{Final Dataset} = \text{Feature Vector}^{\text{T}} * \text{Standardized Original Dataset}^{\text{T}} \tag{3}$$

3.3. K-Means++ Clustering

Jupyter Notebook Environment was used in the implementation of K-means++ algorithm. The mean, standard deviation, and count was calculated and the results displayed to have an idea of the dataset. The pit number, revenue factors, limestone grade, and stripping ratio were used to implement the algorithm based on the correlation results in Section 4.1. The elbow method was used to determine the optimum number of clusters. The elbow method uses the sum of squared distances between each point of cluster and its centroid [22]. Clusters were plotted on a graph and the optimum number of cluster (k) is the point where the change in Within Cluster Sum of Squares (WCSS) becomes constant or level off as shown in Equation (4). The optimum number of cluster (k) is then fitted in the dataset and the results plotted (code). In addition, the data was plotted on a value-controlled line chart based on the optimization variables [23]. The value-controlled line chart represents the normalized values of the principal component on the y-axes and the variables on the x-axis. The variables span across the x-axis and the weight of each variable varies along the y-axis. The standard deviation of each point can be calculated based on the relationship (importance, distance, and cost percentage). The optimal values that satisfy the best possible outcomes are fitted along the $y = 0$ axes.

$$\text{wcss} = \sum_{i=1}^{m} \left(x_i - c_i \right)^2 \tag{4}$$

where wcss is the Within Cluster Sum of Squares; x_i is the data point belonging to the cluster; and c_i is the mean value of points assigned to the cluster. Each observation x_i is assigned to a given cluster such that the sum of squares (ss) distance of the observation to their assigned cluster centers c_i is minimized. The wcss measures the compactness (goodness) of the clustering and the smaller the distance between the points the better the clustering.

3.4. Generalized Linear Models (GLM)

This section implements a multivariate statistical model called Generalized Linear Models (GLM) popularized by McCullagh and Nelder [24]. GLM refers to a large group of regression models used in estimation by exponential distributions

[25]. GLM models include Logistic Regression, ANCOVA, Linear Regression, ANOVA, Poisson Regression, Multinomial response *etc.* GLM models have three major components which are Random Component, Systematic Component, and a Link Function [26]. The random components refer to the predictor or dependent variable (pits along the *y*-axis) defined by the density function (*f*) Equation (5). The systematic components refer to the target or independent or explanatory variables (revenue factors, stripping ratio, limestone grade and ore tones along the *x*-axis) defined by the observation matrix (*n*) Equation (6). The link function (*g*) models the predictions (*u*) to the systematic component's Equation (7). The GLM model is implemented by applying the gaussian distribution, identity link function to the dataset by maximizing the gaussian family likelihood Equation (8) to reduce the least squared error Equation (9) [27]. The results were plotted on a graph as described in Section 4.4.

$$f\left(y;\theta,\varnothing\right) \tag{5}$$

$$\eta = X\beta \tag{6}$$

$$E\left(y = \mu\right) = g^{-1}\left(\eta\right) \tag{7}$$

$$\max_{\beta,\beta_0} -\frac{1}{2}\sum_{i=1}^{N}\left(x_i^{\mathrm{T}}\beta + \beta_0 - y_i\right)^2 - \lambda\left(\alpha\|\beta\|_1 + \frac{1}{2}(1-\alpha)\|\beta\|_2^2\right) \tag{8}$$

$$D = \sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 \tag{9}$$

4. Statistical Analysis of the Results

Statistical analysis was done on the statistical models and the results plotted using scatterplot, line plot, biplot and boxplot for interpretation. The four statistical analysis methods show positive results when compared with that of Whittle. The statistical analysis results are outlined below.

4.1. Correlation Analysis Results

From the correlation analysis, the correlation value is greater than 0, which indicates that there is monotonic relationship. The highest correlation was realized between the revenue factor (RF) and the pit number which had a correlation of 1 (perfect monotonic relationship) as shown in Table 2. Also, there was a high

Table 2. Correlation analysis results.

| Column 1	Column 2	Correlation
Rev Factor	Pit	1
Rev Factor	Stripping Ratio	0.990213211
Pit Number	Stripping Ratio	0.990213211
Limestone Grade	Rev Factor	0.947286057
Limestone Grade	Pit Number	0.947286057
Limestone Grade	Stripping Ratio	0.945555937

correlation between the revenue factor and the stripping ratio, pit number, and stripping ratio. This means that, in order to determine the optimum pit, the revenue factor and the stripping ratio can show better results as compared to other variables.

4.2. PCA Results

The PCA results were plotted on a biplot graph. According to the PCA theory, the highest variance starts from the origin (*x, y*) and decreases as it moves away from the origin. Also, the highest variance is plotted on the PC axes based on the weight or amount of information they carry [28]. From **Figure 4(a)**, we can see that the closest plot to the origin is pit 31. This means that pit 31 has the highest variance in terms of revenue factor, limestone grade and stripping ratio, hence the optimum pit. A verification of this analytical hypothesis was investigated using the optimization results from the Whittle Gemcom sample project (Marvin blend). The results show that the maximum pit (pit 42) was the closest pit to the origin with a revenue factor of 1 as stated in the Whittle handbook. The results are displayed on **Figure 4(b)** shows the PCA results of the Marvin blend mine.

4.3. K-Means++ Results

The K-means++ displayed the results in the form of clusters. The optimum number of clusters (*k*) was chosen to be 3 as shown in **Figure 5** where WCSS almost levels off according to the elbow theory [29]. The optimum cluster (3) was then used to compute the data and the results plotted. From **Figure 6(a)**, the results show that when the number of clusters was set to 1, the centroid was displayed at the interception of pit 31 on the *x*-axis and stripping ratio of 0.33 on
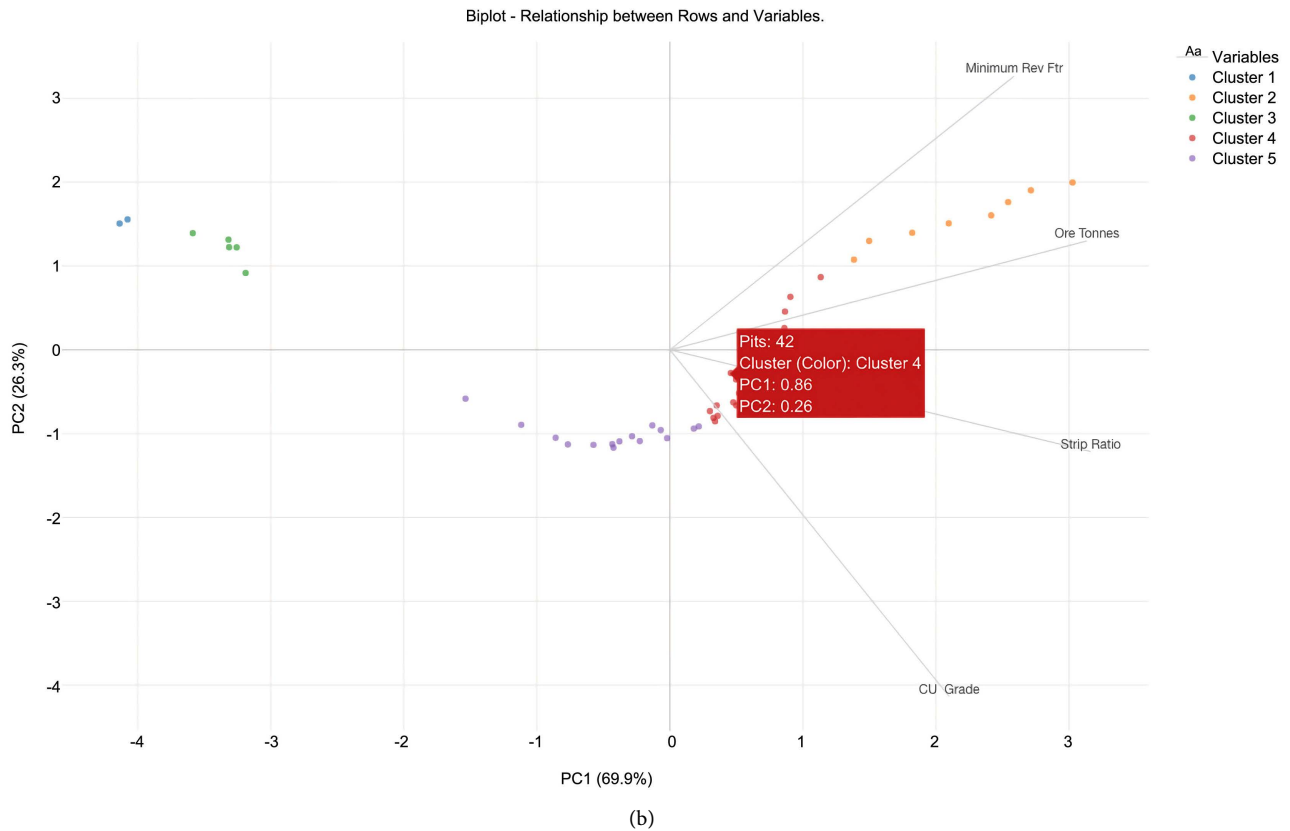

(a)

Biplot - Relationship between Rows and Variables.

(b)
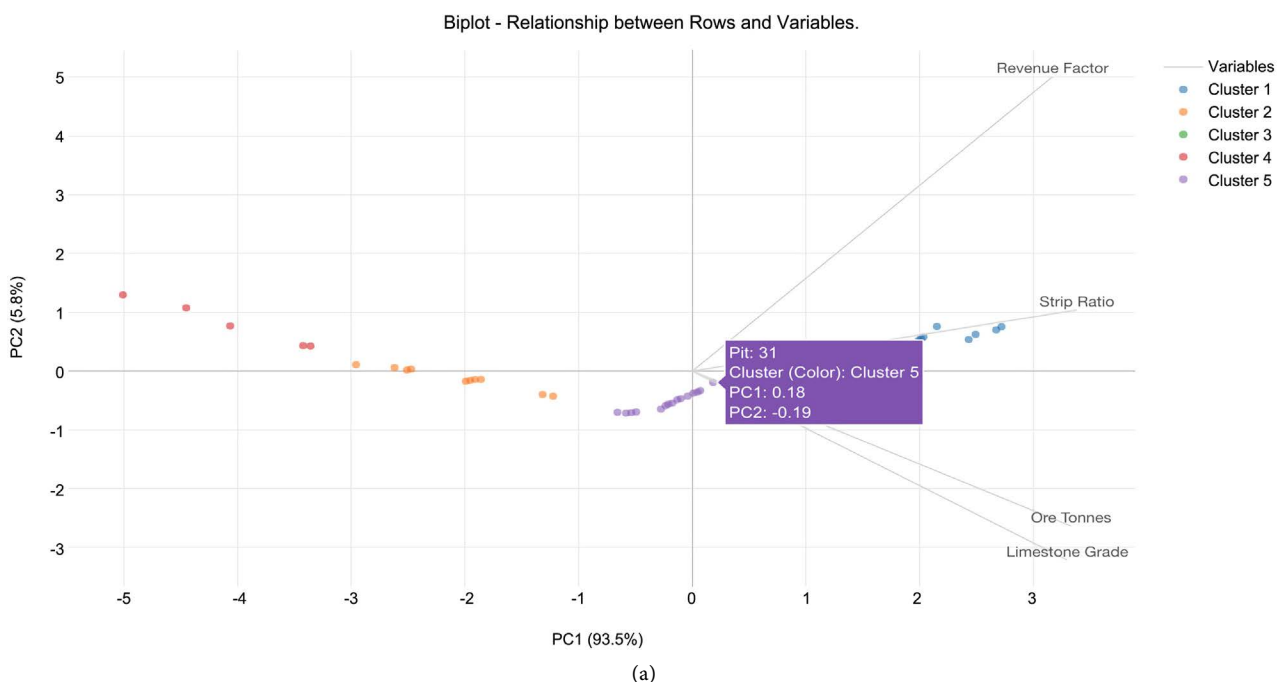
Figure 4. (a) Hebei Limestone mine PCA results; (b) Marvin blend PCA results.
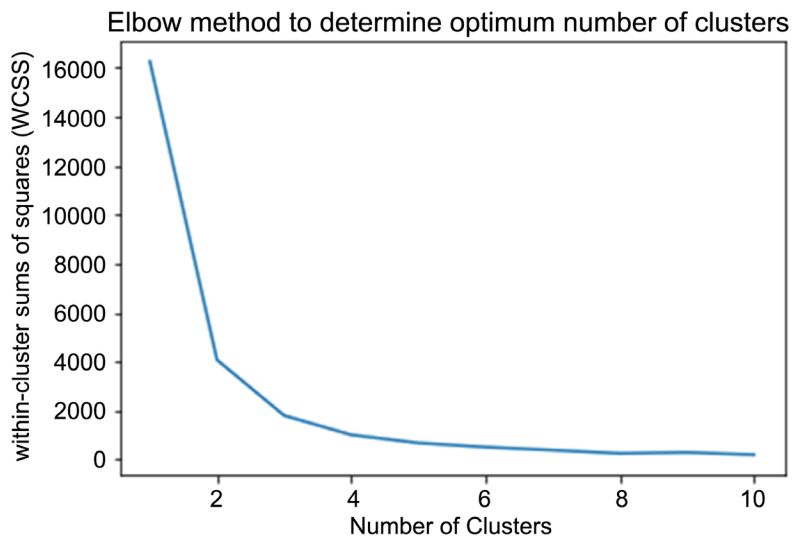

Figure 5 Using the Elbow method to determine the optimum number of clusters.

the y-axis. This shows that pit 31 is the pit with the highest variance of occurrence. When the optimum number of clusters was set to 3, three different centroids were displayed **Figure 6(b)**. The mean μ log of the three centroids defines the variability and deviation from the normal distribution of the component value. The calculated mean was 31 which represents the ultimate pit. In addition,
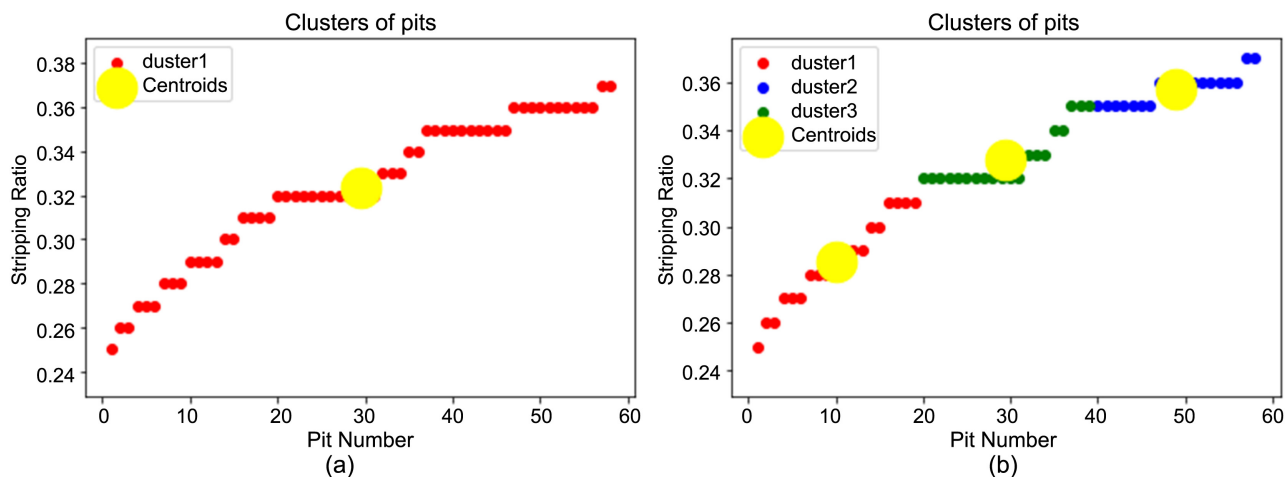
Figure 6. (a) Using 1 cluster; (b) using 3 clusters.

the results were plotted on a line chart which shows the variability of the different optimization variables with respect to pit number (**Figure 7**). When a mouse hovers over a particular point, the readings of that particular point are displayed. **Figure 7(a)** and **Figure 7(b)** has been modified to display the readings of the interception of each variable to avoid too many images. The results of Hebei Limestone mine shows that the ultimate pit 31 with respect to the revenue factor, stripping ratio and Limestone grade **Figure 7(a)**. While for Marvin blend mine was pit 42 **Figure 7(b)**.

4.4. GLM Results

Interpreting results from the GLM model is the key in understanding the underlying model as well as the relationship between the models.

From the results in **Table 3**, we can see that the p-value for some variables is less than 0.05 while others are greater than 0.05 and a summary of the results shows that the p-value is less than 0.05. With regards to this, the ANOVA test was further implemented to test how well the model fits the data and how the model reacts with respect to the variables. A summary of the test shows that $p < 0.05$ which means that the model has a statistically significant association between the predictor variables and the target variables. Also, the coefficients are significant which means that, a change in a variable is associated with a change in the mean response value. The revenue factors and stripping ratios coefficients were relatively high. For each change in the stripping ratios, the mean pit value output increases by 86.8 units and for each change in the revenue factor the pit value increases by 25.1 units contrary to the limestone grade which shows a negative value of −519.2 units. This may be due to the fact that the Limestone grade from the optimization results is almost constant or changes insignificantly.

A plot of the predicted versus the actual values follows a straight line which shows a strong relationship and that the values are normally distributed (there's no unidentified variable, outliers or nonnormality) as shown in **Figure 8**.
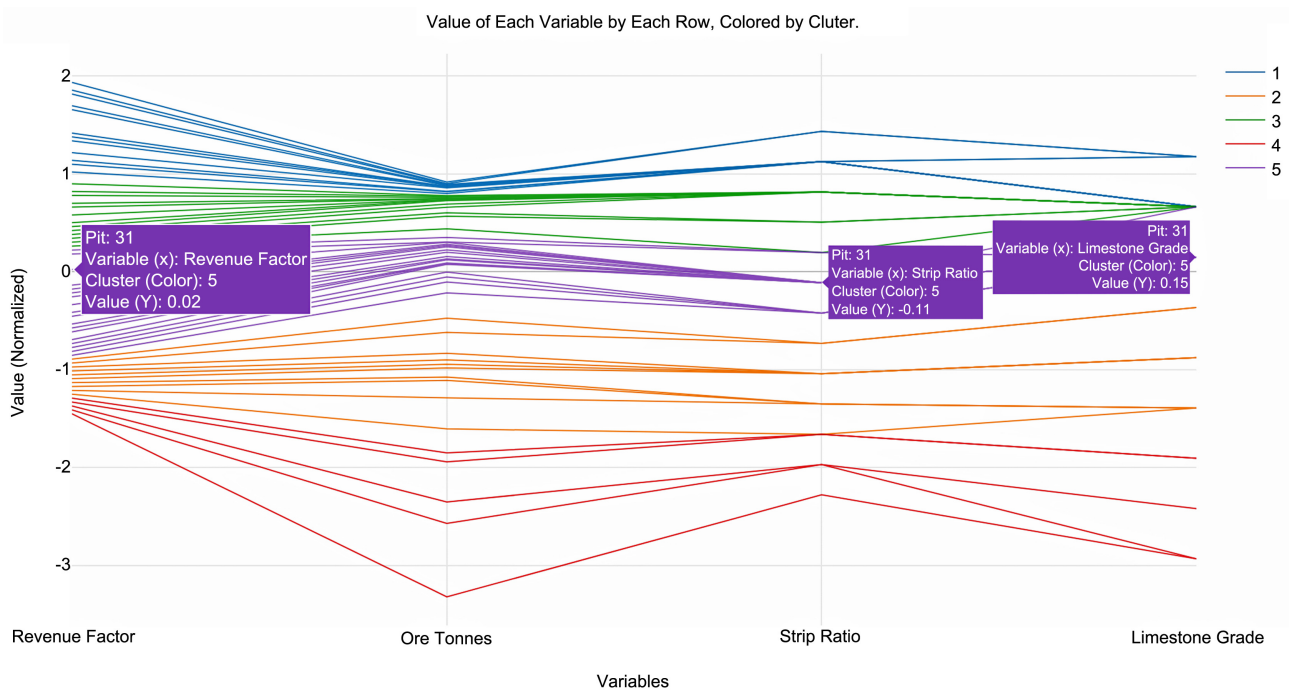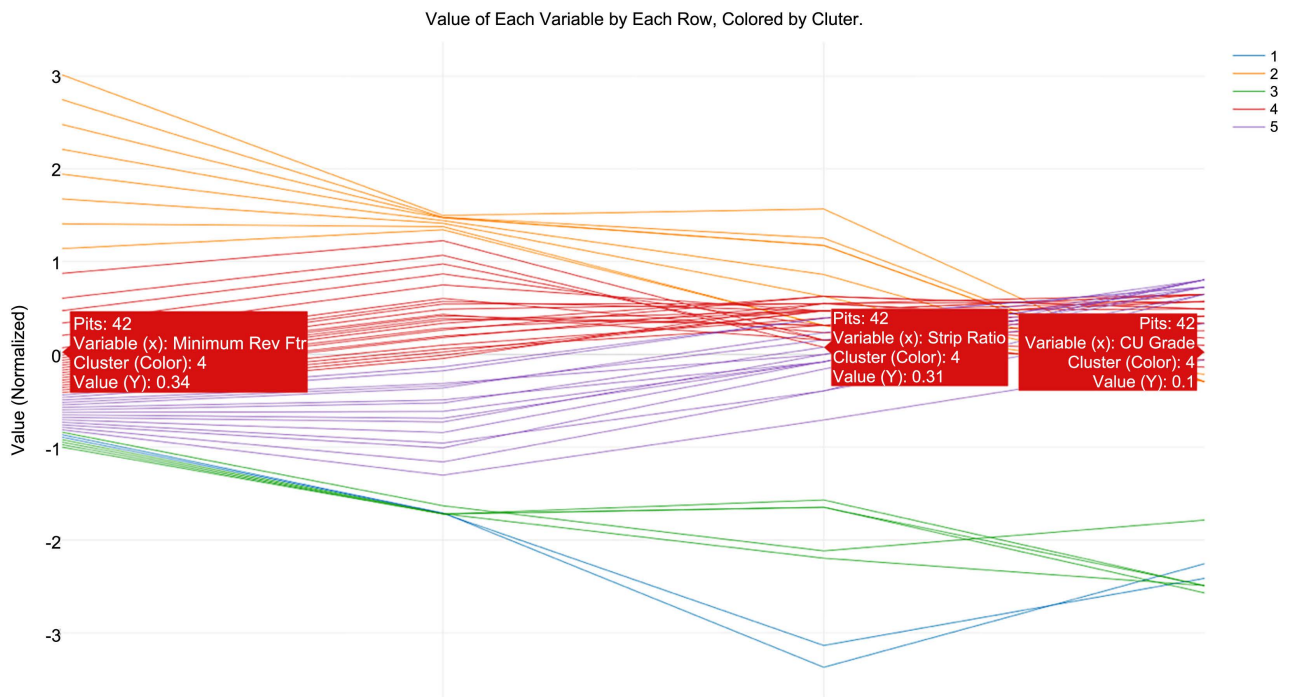
Figure 7. (a) variation of variables with respect to pit number for Hebei Limestone Mine; (b) variation of variables with respect to pit number for Hebei Limestone Mine.

The information criteria (IC) measures the log-likelihood function that produced the data. The asymptomatic assumptions between Akaike's Information

Table 3. GLM results.

Term	Coefficient	Std Error	t Ratio	P Value	Conf High	Conf Low
(Intercept)	21977.85	14543.38	1.511193	0.136679	50482.88	−6527.18
Min Rev Ftr	25.06232	0.959057	26.13226	5.97E−32	26.94207	23.18257
Ore Tonnes	1.47E−06	4.32E−07	3.398316	0.001294	2.32E−06	6.21E−07
Strip Ratio	86.80532	27.58524	3.146803	0.002708	140.8724	32.73825
Limestone Grade	−519.218	329.8802	−1.57396	0.121448	127.3469	−1165.78
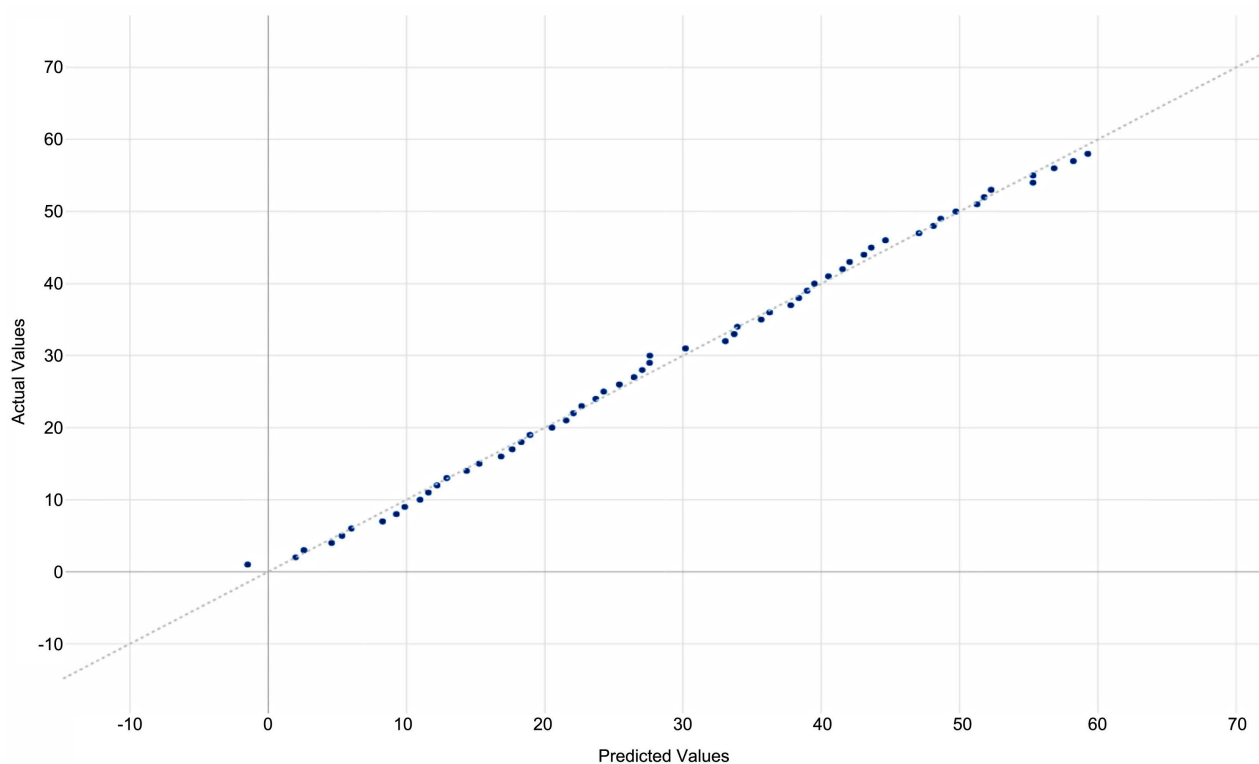
Figure 8. Predicted versus actual values follow a straight line.

Criteria (AIC) and Bayesian Information Criterion (BIC) is heuristic and sometimes unrealistic based on a given situation [30]. The AIC is smaller than BIC which is normal for a well fitted model. Also, the AIC/BIC for the GLM is lower compared to the AIC/BIC for other models which is one of the reasons GLM was used in this study [31].

Since the GLM model proves to fit the data well, a plot of the predictor variable (pit) versus the target variables (revenue factors, stripping ratios, limestone grade, ore tones) shows that the ultimate pit is at the intercept (The intercept (a) is the predicted value of the dependent variable) between the variables **Figure 9**. In **Figure 9**, the pit number at the intersection between the variables for Hebei Limestone mine is 31, hence the optimum pit is 31. Also, the variation between the revenue factor and the pit numbers are almost in the same direction. This shows a high degree of correlation as compared to the other variables.
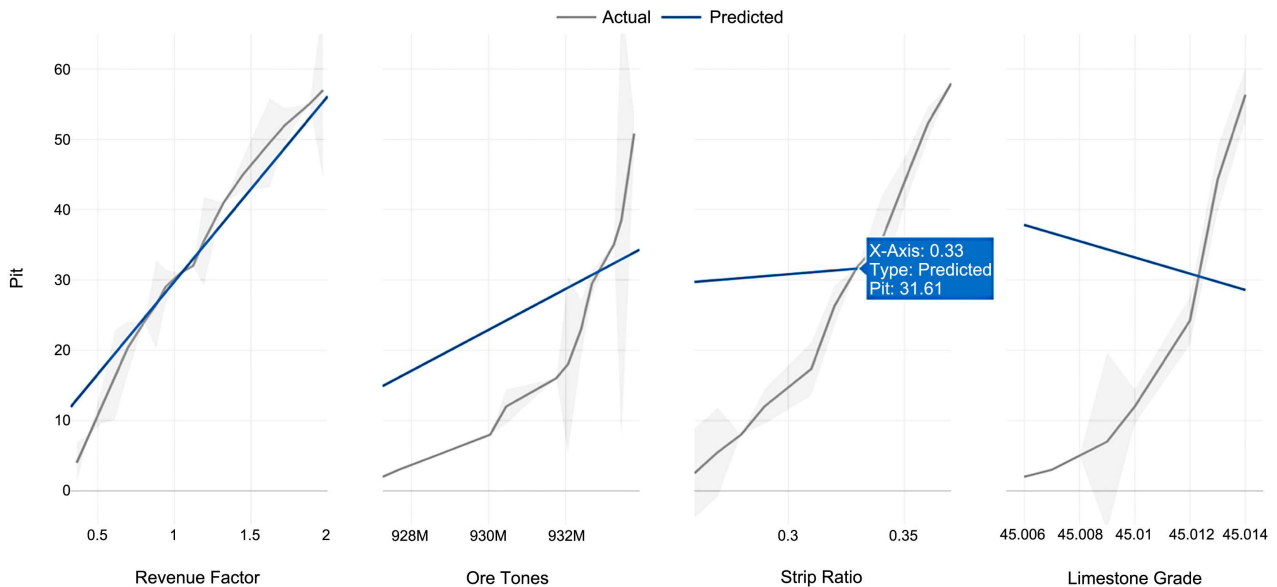
Figure 9. Hebei limestone variation of target variables versus predictor variable.

5. Discussions

This study presents a summary of the statistical analysis of the pit shells created during optimization process in Whittle Software. The goal was to choose and simulate as many statistical models possible to see if statistical analysis can be adopted in selecting the ultimate pit. Despite the fact that some technical details about the models were left out, more emphasis was laid on the type of model and the results they produced with the data set presented in this study. The ultimate pit was exported to Surpac and used to design the mine as shown in **Figure 10**. The final ultimate pit NPV was 3,541,615,321, the total ore tone was 1,207,747,800 tones, Limestone grade of 45.013 and the life of the mine was 20 years.

6. Conclusions

Choosing an ultimate pit is a step that can affect all other processes in mine planning. Also, tuning optimization parameters may produce different results, especially when a certain number of parameters has been defined by the administration and there is doubt in choosing which pit number is the ultimate pit, statistical analysis may help to better interpret the results. It may also be of advantage in a situation where the pit shells created do not satisfy the Whittle selection criteria of optimum ultimate pit with a revenue factor of 1 based on defined parameters.

In this paper, the statistical analysis uses revenue factors, stripping ratios, and the grade of limestone to simulate the ultimate pit. The ultimate pit determined in this case is the optimum ultimate pit which gives the best predictions with respect to these parameters.

The ultimate pit can be used to evaluate the economy of the mine and in making administrative decisions in mine planning.
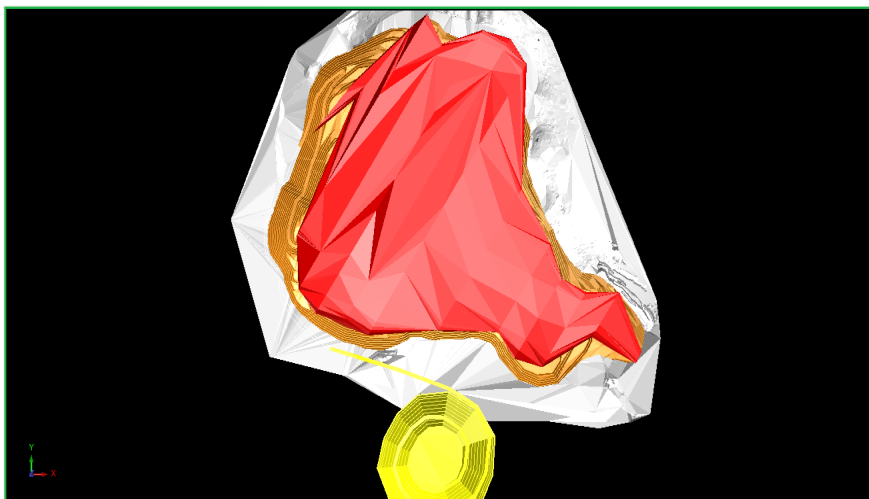
Figure 10. Final ultimate pit showing the ore zone, topography and waste dump.

Limitations of the Model and Future Work

It was found that the models described above had some few limitations in predicting the ultimate pit. The limitations were noted when computing a multi-ore mine at the same time (a mine with more than one ore type). The results from the multi-ore predictions were plus or minus compared to a model with just one ore type. Future work consists of modelling multi-ore mines with two or more ore types by tuning optimization parameters in the statistical models.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ebrahimi, A. (2019) Ultimate Pit Size Selection, Where Is the Optimum Point? SRK Consulting Inc., Canada.
https://docs.srk.co.za/sites/default/files/file/AEbrahimi_UltimatePitSizeSelection_2019_1.pdf

[2] Caccetta, L. and Giannini, L.M. (1987) An Application of Discrete Mathematics Design of an Open Pit Mine. *Discrete Applied Mathematics*, **21**, 1-19.
https://doi.org/10.1016/0166-218X(88)90030-3

[3] Turpin, R.A. (1980) A Technique for Assessing Ultimate Open-Pit Mine Limits in a Changing Economic Environment.

https://preserve.lehigh.edu/cgi/viewcontent.cgi?article=3291&context=etd

[4] Godoy, M. (2009) Mineral Reserve Estimation in the Real World: A Review of Best Practices and Pitfalls. Golder Associates, Los Andeles.

[5] Michaud, D. (2018) Mining Engineering and Mineral Exploration Economics. https://www.911metallurgist.com/blog/mining-engineering

[6] Gemcom (2013) Geostatistics. Gemcom Software International Inc. (Gemcom), Vancouver.

[7] Noumea, (2013) Resource Estimation and Surpac Noumea. Mining Associates. Spring Hill.

[8] Jalloh, A.B. and Sasaki, K. (2014) Geo-Statistical Simulation for Open Pit Design Optimization and Mine Economic Analysis in Decision-Making. In: Drebenstedt, C. and Singhal, R., Eds., *Mine Planning and Equipment Selection*, Springer, Cham, 93-101. https://doi.org/10.1007/978-3-319-02678-7_10

[9] Gemcom (2013) Block Modelling. Gemcom Software International Inc. (Gemcom), Vancouver.

[10] Khademi, A. (2016) Applied Univariate, Bivariate, and Multivariate Statistics. *Journal of Statistical Software*, **72**, 1-4. http://dx.doi.org/10.18637/jss.v072.b02

[11] Uraibi, S.H. and Midi, H. (2019) On Robust Bivariate and Multivariate Correlation Coefficient. *Economic Computation and Economic Cybernetics Studies and Research*, **53**, 221-239. https://doi.org/10.24818/18423264/53.2.19.13

[12] Garbade, M.J. (2018) Understanding K-Means Clustering in Machine Learning. https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1

[13] Efendiyev, G.M., Mammadov, P.Z., Piriverdiyev, I.A. and Mammadov, V.N. (2016) Clustering of Geological Objects Using FCM-Algorithm and Evaluation of the Rate of Lost Circulation. *Procedia Computer Science*, **102**, 159-162.

[14] Gemcom (2013) Geological Database. Vol. 6.5, Gemcom Software International Inc., Vancouver.

[15] Pernet, C.R., Wilcox, R. and Rousselet, G.A. (2013) Robust Correlation Analyses: False Positive and Power Validation Using a New Open Source Matlab Toolbox. *Frontiers in Psychology*, **3**, Article No. 606. https://doi.org/10.3389/fpsyg.2012.00606

[16] Thobaven, D. and Willoughby, J. (2012) Strategic Mine Planning in Whittle. Gemcom Software International Inc., Vancouver.

[17] Drezner, Z. (1995) Multirelation—A Correlation among More than Two Variables. *Computational Statistics & Data Analysis*, **19**, 283-292. https://doi.org/10.1016/0167-9473(93)E0046-7

[18] Data Science (2019) Pearson vs Spearman vs Kendall. https://datascience.stackexchange.com/questions/64260/pearson-vs-spearman-vs-kendall

[19] Statstutor (2004) Spearman's Correlation. http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf

[20] Jolliffe, I.T. and Cadima, J. (2016) Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions A*, **374**, Article: 20150202. https://doi.org/10.1098/rsta.2015.0202

[21] Jaadi, Z. (2020) A Step by Step Explanation of Principal Component Analysis. https://builtin.com/data-science/step-step-explanation-principal-component-analys

is

[22] Maklin, C. (2018) K-Means Clustering Python Example.
https://towardsdatascience.com/machine-learning-algorithms-part-9-k-means-exa
mple-in-python-f2ad05ed5203

[23] Shillito, M.L. and Marle, D.J.D. (1992) Value: Its Measurement, Design, and Management. Wiley & Sons, Inc., New York, 368 p.

[24] McCullagh, P. and Nelder, F.J. (1989) Generalized Linear Models. Chapman and Hall, Chicago.

[25] Haq, Z., Hanna, T., Michal, K., Kraljevic, T. and Stubbs, M. (2018) Generalized Linear Model (GLM).
https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glm.html

[26] Müller, M. (2004) Generalized Linear Models. Fraunhofer Institute for Industrial Mathematics, Kaiserslautern.

[27] Nykodym, T., Kraljevic, T., Wang, A. and Wong, W. (2017) Generalized Linear Modeling with H_2O. H2O.ai, Inc., CA.
http://docs.h2o.ai/h2o/latest-stable/h2o-docs/booklets/GLMBooklet.pdf

[28] Gniazdowski, Z. (2017) New Interpretation of Principal Components Analysis. *Zeszyty Naukowe WWSI*, **11**, 43-65.

[29] Grekousis, G. (2020) Spatial Analysis Theory and Practice: Describe-Explore-Explain through GIS. Cambridge University Press, New York.
https://doi.org/10.1017/9781108614528

[30] Burnham, K.P. and Anderson, D.R. (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, **33**, 261-304.
https://doi.org/10.1177/0049124104268644

[31] Dziak, J.J., Coffman, D.L., Lanza, S.T. and Li, R. (2012) Sensitivity and Specificity of Information Criteria. The Methodology Center, Pennsylvania.