Applications of Statistical and Machine Learning Models in Mining Engineering

Siyi Li

Department of Mining and Materials Engineering

McGill University, Montreal, Quebec

December 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Engineering

Abstract

Statistical and machine learning models are useful tools that can be used to extract valuable information from raw data and make accurate predictions and can be applied in the optimization of mining related systems through various means. This thesis aims to further contribute to the applications of such techniques in mining engineering by providing 4 different cases where statistical and machine learning models could facilitate design and decision making. Principal component analysis (PCA) was used to reduce the dimensions of the problem and simplify the design of stockpiles in bed-blending operations, generalized linear models (GLM) were introduced to model non-linear relationships among variables in quality control and safety related problems, factor analysis methods including structural equation models (SEM) were presented to be used in conjunction with cognitive work analysis to better analyze the underlying structures or latent variables in operational health and safety in mining operations, and finally clustering, which is a family of unsupervised learning methods, was applied to a mine planning problem to integrate mining and mineral processing and maximize recovery.

Résumé

Les modèles statistiques et d'apprentissage automatique sont des outils utiles qui peuvent être utilisés pour extraire des informations précieuses à partir de données brutes et faire des prévisions précises et peuvent être appliqués dans l'optimisation des systèmes liés à l'exploitation minière par divers moyens. Cette thèse vise à contribuer davantage aux applications de ces techniques en génie minier en fournissant 4 cas différents où les modèles statistiques et d'apprentissage automatique pourraient faciliter la conception et la prise de décision. L'analyse en composantes principales (ACP) a été utilisée pour réduire les dimensions du problème et simplifier la conception des stocks dans les opérations de mélange de lit. Des modèles linéaires généralisés (GLM) ont été introduits pour modéliser les relations non linéaires entre les variables dans le contrôle de la qualité et les problèmes liés à la sécurité, des méthodes d'analyse factorielle, y compris des modèles d'équations structurelles (SEM), ont été présentées pour être utilisées conjointement avec une analyse du travail cognitif afin de mieux analyser les structures sous-jacentes ou les variables latentes de la santé et de la sécurité opérationnelles dans les opérations minières, et enfin le clustering, qui est une famille de non supervisés méthodes d'apprentissage, a été appliquée à un problème de planification minière pour intégrer l'exploitation minière et le traitement des minéraux et maximiser la récupération.

Acknowledgements

I would like to express my sincerest gratitude towards Professor Mustafa Kumral for his invaluable guidance and patience. This work also would not have been possible without the financial support from my mother, Li Li-Qun, who made a lot of sacrifices for my education. Finally, I would like to acknowledge the comradery expressed by the Mineral Economics, mine reliability and asset management lab members, as well as my friends in statistics and computer science.

Contribution of Authors

The author of this thesis is the primary author as well as the first author of all journal articles and conference papers published as part of this thesis. Professor Kumral is the supervisor of the author's Master of Engineering degree and co-authored the following papers that originated from this research:

- Li, S., de Werk, M., St-Pierre, L. and Kumral, M. (2019) Dimensioning a stockpile operation using principal component analysis, International Journal of Minerals, Metallurgy and Materials.
    - The candidate student implemented the statistical analysis and regression models. The candidate student modified the PCA model. He also implemented the regression analysis. Furthermore, he primarily involved the paper writing and revisions. The candidate student also produced the figures.
- Li, S, and Kumral, M. (2019), Linear and generalized linear models in quality control and equipment reliability analysis, World Congress on Resilience Reliability and Asset Management, Singapore.
    - The candidate student conducted all statistical analysis and wrote most parts of the paper.
- Li, S., Sari, Y.A. and Kumral, M. (2019) New approaches to cognitive work analysis through latent variable modeling in mining operations, International Journal of Mining Science and Technology, 29(4): 549-556.
    - The candidate student established theoretical base and conducted the research, the candidate student was also the main contributor in writing the paper.
- Li, S., Sari, Y.A. and Kumral, M. (Minor revision) Optimization of Mining-Mineral Processing Integration Using Unsupervised Machine Learning Algorithms, Natural Resources Research.
    - The candidate student primarily conducted the research and wrote most parts of the paper.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike's information criterion
CFA	Confirmatory factor analysis
CLARA	Clustering large applications
CWA	Cognitive work analysis
EFA	Exploratory factor analysis
EPE	Integrated squared prediction error
FSMC	Firm safety management capability
GLM	Generalized linear model
GLS	Generalized least squares
LM	Linear model
LoM	Life of mine
ML	Maximum likelihood
NPV	Net present value
PAM	Partitioning around medoids
PC	Principal component
PCA	Principal component analysis
ROC curve	Receiver operating characteristic curve
SEM	Structural equation model
SSReg	Sum-of-squares regression
SSRes	Sum-of-squares residual
SST	Sum-of-squares total
SVD	Singular value decomposition
SVM	Support vector machine
TWSS	Total within cluster sum-of-squares
UCS	Uniaxial compressive strength
ULS	Unweighted least squares
VRR	Variance reduction ratio

Chapter 1 Introduction

1.1 Preliminaries

Over the past decades, a significant amount of research has been dedicated towards the optimization of mining systems including production scheduling, block sequencing, reliability analysis of assets and equipment, sustainability etc., which was partly driven by the rapid depletion of high grade deposits, the increasingly high operational costs as well as tightening environmental regulations. Hence it is imperative that mining systems become more efficient, safer and more environmentally friendly. Indeed, decision making and planning based on statistical and machine learning models could aid to the further optimization of mining systems, and the implementation of these methods could have significant impacts on the overall performance of the mining industry, as even small improvements in production and efficiency can lead to greater profitability due to the size and scale of the industry. There exists tremendous potential of applications of statistical and machine learning models in mining engineering. Statistical models could facilitate decision making under uncertainty by quantifying risks, and by identifying influential variables via rigorous inference, which can be particularly helpful in reliability analysis of systems in the presence of censored and truncated data. On the other hand, machine learning methods have become more prevalent in recent years due to the affordability of high-performance computers, and their applications can be very versatile. For instance, supervised and deep learning methods, such as support vector machines and neural networks could be used to make accurately predictions based on large numbers of input data. Unsupervised learning methods on the other hand, could be utilized to find patterns in data or conduct dimension reduction without pre-existing labels. While machine learning models such as naïve Bayes, recurrent neural networks and convolutional neural networks have been widely utilized in fields including natural language processing and computer vision, so far there have been relatively limited applications of machine learning in mining engineering related problems. This work aims to introduce innovative ways of applying statistical and machine learning models to optimize mining and related systems.

1.2 Research Objectives

The primary objective of this work is to optimize mining related problems via the introduction of statistical and machine learning methods. In particular:

- A principal component analysis (PCA) based solution was presented to aid to the design of stockpiles that could achieve high variance reduction ratio of polymetallic inputs in bed-blending operations,
- Generalized linear models (GLM) were introduced in order to conduct regression analysis on response variables with non-normal distributions. Two case studies were presented regarding applications of GLMs in mining related problems
- A factor analysis based statistical model was used to quantitively analyze safety and accident related data and extract underlying latent variables to facilitate in-depth understanding and improvement of operational health and safety of workers. Moreover, a proposition was made to combine the statistical model with cognitive work analysis (CWA) to enhance work safety in the mining industry.
- Clustering, which is a family of unsupervised machine learning techniques, was used to partition block data into different groups each for one unique mineral processing destination. This method has been shown to perform better than traditional cut-off-based methods when taking into considerations loss of recovery due to fluctuations in input grades.

1.3 Originality and Success

Mineral processing plants generally have narrow tolerances for the grades of their input raw materials, so stockpiles are often maintained to reduce material variance and ensure consistency. However, designing stockpiles has often proven difficult when the input material consists of multiple sub-materials that have different levels of variances in their grades. In this thesis, this issue was addressed by applying principal component analysis (PCA) to reduce the dimensions of the input data. The study was conducted in three steps. First, PCA was applied to the input data to transform them into a lower-dimension space while retaining 80% of the original variance. Next, a simulated a stockpile operation was simulated with various geometric stockpile configurations using a stockpile simulator in MATLAB. The variance reduction ratio was used as the primary criterion for evaluating the efficiency of the stockpiles. Finally, multiple regression was used to identify the relationships between stockpile efficiency and various design parameters and analyzed the regression results based on the original input variables and principal components. The results showed that PCA is indeed useful in solving a stockpile design problem that involves multiple

correlated input-material grades. Statistical methods including regression analysis has been widely utilized in the modelling of quality characteristics, systems reliability and safety engineering data. However, due to the many restrictive assumptions of the traditional regression models, including normality and homoscedasticity of the error term, such models could be rendered inappropriate when dealing with non-normal, binary or count data. In chapter 3 and 4, this study presents an overview of the intuition and mathematical foundations of the extension from linear models to generalized linear models (GLM), as well as factor analysis techniques including Structural Equation Models (SEM). Case studies incorporating Gamma regression, binomial/multinomial regression are used to demonstrate the potential applications of GLMs in quality, reliability and safety engineering, as well as the possibility of combining cognitive work analysis and structural equation modelling to investigate underlying structures and reasons behind work safety and operational health. Traditional ore-waste discrimination schemes often cause the loss in recovery because applying a cut-off grade has no control the average grade of ore, resulting in grade fluctuations of input grades in mineral processing. Chapter 5 introduces target grades instead of cut-off grades for different processing streams and models the losses due to deviation from targets via the Taguchi loss function. Three unsupervised learning algorithms, k-means clustering, CLARA and k-means based approximate spectral clustering (KASP), were presented to group mine planning blocks into clusters of similar grades with different processing destinations. Also, a technique considering uncertainties associated with block grades was proposed to generate new sequences that reduce variation in processing capacities across mine life. The case study in this chapter involved the treatment of a realistically large mining dataset. The results showed that clustering methods outperform cut-off 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 life of mine.

1.4 Thesis Organization

This thesis consists of 6 chapters:

Chapter 1 lists the topics covered in this thesis, its main objectives as well as original contributions.

Chapter 2 provides the literature review on relevant statistical and machine learning methods

Chapter 3 introduces and explains in detail how PCA can be used to facilitate the design of stockpiles in bed-blending operations.

Chapter 4 provides the intuition and mathematical foundations of linear and generalized linear models and presents case studies regarding utilization of GLMs in modelling nonlinear relationships in quality control and safety related problems.

Chapter 5 discusses the utilization of latent variable modeling related to occupational health and safety in the mining industry using SEMs.

Chapter 6 presents the methodology of three clustering algorithms and explains how clustering-based solutions manage to minimize deviation from target grades in mineral processing.

Chapter 7 concludes by summarizing the works done in this thesis and discusses potential improvements and future works.

Chapter 2 Literature Review

2.1 Introduction

The primary objective of this literature review is to provide an overview of how statistical models and machine learning methods have been used to model and ameliorate the design of engineering systems, as well as the basic backgrounds of the relevant algorithms utilized.

2.2 Key Methods

2.2.1 Regression models

Regression has remained one of the most important tools in statistics for the past 30 years [1]. Given a vector of inputs (or predictors/features) $\boldsymbol{X} = [\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_p}]$ with $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{X_i} \in \mathbb{R}^{n\times 1}\ \forall i \in \{1,2,\dots,p\}$, as well as a vector of outputs (or responses/independent variables) $\boldsymbol{y} = [y_1, y_2, \dots, y_n]^T$, the aim is to predict the output via the model

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.2.1}$$

Where $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{p\times 1}$ is the estimated parameter coefficients.

The standard way to obtain an estimate for the regression coefficients $\boldsymbol{\beta}$ is via the least squares method, by minimizing

$$||y - \hat{y}||_2^2 = \sum_i (y_i - \hat{y}_i)^2 \tag{2.2.2}$$

It could be shown that when normality assumption is added to the model i.e. $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$, the least squares solution becomes the maximum likelihood solution. As

$$\log\left[\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y_i-\mu_i)^2}{2\sigma^2}\right\}\right)\right] = Constant - \frac{1}{2\sigma^2}\sum_i (y_i-\mu_i)^2 \tag{2.2.3}$$

It then becomes clear that the objective function for maximum likelihood and least squares are essentially equivalent. The estimator has form $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$, with $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$. Generalized linear models (GLMs) extend the standard linear regression models to allow for non-Gaussian distributions of the response and also possibly nonlinear relationships between its mean and variance. One additional component in a GLM is the link function which maps the expectation of the response to the linear predictor [2]. GLMs are widely used in the modelling of categorical data, for instance logistic regression and multinomial regression, as well as count data with Poisson log-linear model. More details of GLMs are given in Chapter 4.

2.2.2 Regularization methods

Suppose the true relationship between the response $\boldsymbol{y}$ and the covariates of interest $\boldsymbol{X}$ can be denoted as $\boldsymbol{y} = f(\boldsymbol{X}) + \boldsymbol{\varepsilon}$, and some regression techniques were used such that at a sample point x, the model estimate is $\hat{f}(x)$, then the integrated squared prediction error (EPE) can be written as

$$\begin{aligned} EPE(x) &= \mathbb{E}[(y - \hat{f}(x))^2] \\ &= (\mathbb{E}[\hat{f}(x) - f(x)])^2 + \mathbb{E}[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2] + \sigma^2 \\ &= Bias[\hat{f}(x)]^2 + Var[\hat{f}(x)] + Error \end{aligned} \tag{2.2.4}$$

Also known as the bias-variance tradeoff, where $Bias = \mathbb{E}\left[\hat{f}(x) - f(x)\right]$ and $Variance = \mathbb{E}[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2]$. This signifies that when the regression model is too simple compared to the true model, then the model would be underfitting and the model prediction would deviation too much from the mean. Whereas when the true relationship is simpler than the regression model, for instance when an abundance of higher order polynomial terms is added, then the model would be overfitting and a small perturbation in the input would lead to much larger variations in the output. Regularization is one way that can be adopted to avoid overfitting at the cost of some additional bias in the model. The most commonly used regularization methods in regression is the ridge regression and lasso regression, which apply L-2 and L-1 norm penalization respectively, as shown in Equation (2.2.5).

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{ridge} &= \underset{\boldsymbol{\beta}}{argmin} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\beta}^T \boldsymbol{X}_i)^2 + \lambda ||\boldsymbol{\beta}||_2^2 \\ \widehat{\boldsymbol{\beta}}_{lasso} &= \underset{\boldsymbol{\beta}}{argmin} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\beta}^T \boldsymbol{X}_i)^2 + \lambda ||\boldsymbol{\beta}||_1 \end{aligned} \tag{2.2.5}$$

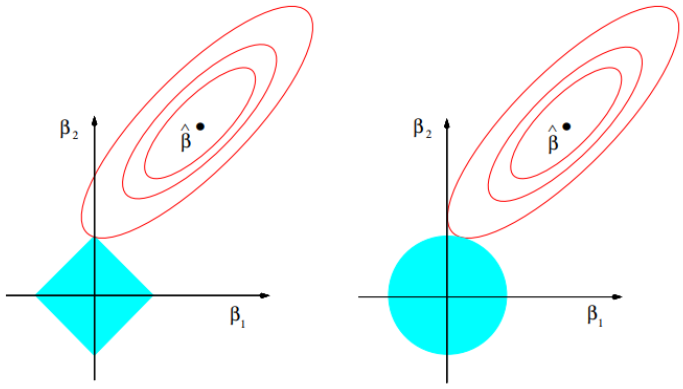


Figure 2.1. Illustrations of lasso (left) and ridge (right).
Source: Hastie et al. [1]

The effects of ridge and lasso regression are illustrated Figure 2.1, where the red ellipses signify contours of equal likelihood, and the area in cyan represent the contours for the constraint function. Solutions are found where the likelihood and constraint contour intersect. Combining ridge and lasso penalization is termed elastic net regression, and has been shown to be particularly useful in high-dimensional problems [3]. Regression analyses have been used to model a wide range of data in mining engineering. Sauvageau and Kumral used various robust regression methods including least absolute regression, M-estimation, MM-estimation etc. to model mineral processing data in the presence of outliers [4], Wang et al. used logistic and Poisson GLMs to identify the primary factors that contribute to unsafe behavior of coal miners [5], Rezania et al. proposed using evolutionary polynomial regression to assess complex civil engineering systems [6]. Detailed explanation regarding multiple regression, generalized linear models, factor analysis and relevant literature reviews are given in Chapters 4 and 5.

2.2.3 Ensemble Learning

Ensemble learning approaches combine multiple base supervised learning algorithms (regressors and classifiers) in order to obtain superior predictive performance. Bagging (bootstrap aggregation), boosting and stacking are the three most common methods in ensemble learning. Bagging refers to the training of a series of models on multiple bootstrapped samples (randomly sampled uniformly with replacement) of the original data set, then making a majority vote for classification or taking average for regression. Empirically, bagging has been known to be able to reduce variance at the cost of increasing bias. Random forest, which is a bagging technique, has been widely used in modelling mining engineering related problems. Mishra et al. used random forest decision based approaches for blast design, Tingxin et al. used similar methods to accurately predict slope stability in open-pit coal mines [7, 8]. Another ensemble learning method, boosting, incrementally constructs a series of different weak models with high bias and low variance, and re-weights the mis-classified data points in each iteration. Stacking is a third ensemble method that trains a meta-model to combine the predictions of an arbitrary set of base learning models.

2.2.4 Support Vector Machines (SVMs) and Neural Networks

SVMs and neural networks are among some of the most powerful supervised machine learning algorithms, and have seen greatly increased number of applications in various fields of engineering in recent years. The theory behind some of the more complicated machine learning algorithms such as SVM and various types of neural networks will not be detailed in this thesis, as they cannot

be easily summarized in a very small number of paragraphs. In recent years, there have been plenty of research on the applications of SVM in mining engineering related problems. For instance, Li et al. utilized SVM to train a regression model that could predict outcomes in mechanized mining faces using a set of conditions that incorporate geological factors, technical factors and management factors [9], Chatterjee used an ensemble of SVMs to estimate reliability of mining equipment, and performed hyperparameter tuning using genetic algorithms [10], Chen et al. proposed a new form of probabilistic back-analysis that combines Bayesian priors for geo-mechanical parameters and least-squares SVM for prediction of displacements [11]. Although this thesis does not include case studies related to SVM, there is tremendous potential for it to be applied to various mining related problems that concerns prediction.

Neural networks, on the other hand, could refer to a series of methods including multi-layer perceptron, convolutional neural networks (CNNs), recurrent neural networks (RNNs), generative adversarial neural networks (GANs) etc., and can be supervised or unsupervised. Neural networks have very vast applications in natural language processing, image analysis, time series analysis and many other different areas. Neural networks have also been proven to be useful in prediction related tasks in mining engineering, Lv et al. trained an improved back-propagated neural network for the prediction of surface subsidence coefficient in backfilled coal mining areas [12], Rakhshani et al. used an artificial neural network to detect and predict faults in boilers of power plants, Gonzalo et al. utilized nonlinear autoregressive exogenous neural networks to model the availabilities of heavy duty mining equipment [13]. Although not a major topic of this thesis, neural networks have tremendous potentials to be applied particularly in image analysis related tasks in mining engineering, as well as through the modelling of subsurface geological data with generative adversarial neural networks.

2.2.5 Unsupervised Learning

The primary task of a supervised learning algorithm is to construct a model from a training data set with associated labels for each entry, and to maximize the out-of-sample prediction accuracy. In unsupervised machine learning however, the objective is to identify patterns in data sets without knowing the labels. There are three tasks that are commonly associated with unsupervised learning, namely dimensionality reduction, cluster analysis and anomaly detection. Principal component analysis (PCA) is one very popular method for dimensionality reduction and is often based on the eigen/spectral decomposition of the variance-covariance matrix or correlation matrix of the data

set, or the singular value decomposition of the data set. The mechanism of PCA is detailed in Chapter 3. Understandably dimensionality reduction techniques are extremely popular, particularly in high-dimensional problems, and are often used in conjunction with supervised learning methods. Shao et al. used PCA together with support vector regression to predict pressure in natural gas desulfurization process [14], Xu et al. used PCA and a back-propagated neural network to accurately model coal and gas outburst [15]. Clustering is another unsupervised learning method that is heavily involved in this thesis. Some of the most common clustering algorithms, including the k-means algorithm, partitioning around medoids (PAM), clustering large applications (CLARA) and spectral clustering are introduced in detail in Chapter 6, with relevant literature reviews.

Chapter 3

Dimensioning a Stockpile Operation Using Principal Component Analysis (PCA)

3.1 Bed-blending Operations

Bed-blending operations are applied across a variety of industries, including the mining industry, which uses stockpiles to homogenize and reduce the variability of the raw materials before delivery to mineral processing plants. The reason being that unfavourable residual variations always persist even in materials from the same source, due to the discontinuous, cyclic, random, and autocorrelated nature of ore [16]. Optimization of processing efficiency relies heavily on homogenizing input materials [17]. A bed-blending system has two phases. In the first phase, a stacker traverses the ground with constant velocity along the stockpile, during which process materials are laid down on the same level as the stacker. As the stacker gradually reaches the end of the stockpile, it decelerates until it entirely stops, before starting again traveling back in the opposite direction. In the reclaiming phase, a reclaimer (either a bucket-wheel or a harrow-type scraper, etc.) cuts slices of the stockpile that is perpendicular to the direction of stacking [18]. In the past, many years researchers have constantly been working towards optimizing the design of blending operations and various theories and methods have been put forward [19-21]. Gy 1992 introduced the concepts of using the variance reduction ratio (VRR) to evaluate the effectiveness of blending [22], Dowd [23] presented the use of geostatistical approaches to improve stockpile design by predicting the output characteristics of given stockpile parameters, Kumral [24] incorporated multiple regression and genetic algorithms into optimization of stockpile design. In particular, there have been increasingly many applications of statistical methods and mathematical models in the optimization of metallurgical and minerals engineering operations [25-28]. The designing of a bed-blending operation could be relatively straightforward when there exists but one mineral grade that is of concern to the processing plant; however, this is rarely the case as raw material grades have a multivariate nature, for instance, certain types of iron ores might have more than six different chemical compositions that need to be homogenized [29, 30]. Therefore, challenges arise in situations where there are different levels of variations across the material grades that together make up the stockpile input. Consequently, this research proposes the

utilization of PCA, which is a dimension reduction technique that has been widely applied in many fields including image and signal processing, statistical mechanics and multivariate quality control etc. By introducing principal component analysis (PCA) to this problem, it is possible to reduce the number of varying materials to a much smaller value, while preserving most of the information contained from the original data, and thereby facilitate designing of the stockpile with minimum loss of information. This research is conducted in three primary steps:

i. The principal component analysis is performed on the input data, projecting it to lower dimension space while retaining most of the information. The input data used in this chapter is a serially correlated dataset with realistic statistical properties that could well occur in a real-world problem.

ii. A computer algorithm is built to simulate the process of bed-blending, where stacking and reclaiming are mimicked by laying down discrete blocks of unit weight and volume to form a cuboid and then slicing across its lengths. The stockpile simulator computes the input and output variances of all the material grades, including those of the principal components'.

iii. Multiple regression is used to find the relationships between the response and the predictors, with the response being the variance reduction ratio, and the predictors being stockpile design parameters. This step is repeated for all the input materials including the principal components.

3.2 Methodology of PCA and Stockpile Simulator

3.2.1 Principle Component Analysis

Principal component analysis (PCA) is a multivariate statistical/machine learning technique that seeks to reduce p-dimensional correlated variables to a set of ordered and uncorrelated k-dimensional linear projections. Mathematically it is related to finding the spectral/eigen decomposition of the positive-semidefinite variance-covariance matrix or the singular value decomposition (SVD) of the rectangular data matrix.

3.2.2 Spectral Decomposition of the Variance-Covariance Matrix

Let $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ be a data matrix where n represents the number of observations and p the number of variables with $n > p$. $\boldsymbol{X}_{\mathbf{c}}$ is the mean-centred data matrix with $\boldsymbol{X}_{\boldsymbol{c}} = \boldsymbol{X} - \mathbf{1}_{\mathrm{n}} \boldsymbol{\mu}^{\mathrm{T}}$, with $\mathbf{1}_n$ being a $n \times 1$ column vector of 1's and $\boldsymbol{\mu}^T$ being the $1 \times p$ row vector denoting the variable means. Let $\boldsymbol{x}_{\mathrm{c}}$ be the row vector representing the variables $\boldsymbol{x}_c = [x_1\ x_2\ \ldots\ x_p]$. The variance-covariance matrix can

then be found by:

$$\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{X_c}^T\boldsymbol{X_c} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix} \tag{3.2.1}$$

The variance-covariance matrix is symmetric and positive semidefinite with the diagonal entries being the variances of each variable and the rest being covariances between variables. The variance-covariance matrix can then be used to solve for its eigenvectors (v) and their corresponding eigenvalues (λ) by finding its spectral decomposition, i.e.

$$\boldsymbol{\Sigma} = \boldsymbol{VEV}^T \tag{3.2.2}$$

By definition of the spectral decomposition, the columns of the orthogonal matrix $\boldsymbol{V}$ are the eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{E}$ is a diagonal matrix with its diagonal entries being the corresponding eigenvalues in descending order. This could be seen by multiplying $\boldsymbol{V}$ to the right on equation (3.2.2), which results to $\boldsymbol{\Sigma V} = \boldsymbol{VEV}^T\boldsymbol{V} = \boldsymbol{VE}$, and by looking at the columns of $\boldsymbol{V}$ and diagonal entries of $\boldsymbol{E}$, $\boldsymbol{\Sigma}\mathrm{v_i} = \lambda_i\mathrm{v_i}$. The p eigenvectors resulting from the spectral decomposition are orthogonal and thereby linearly independent and forms a p dimensional space. Let the reduced k linear projections be $\boldsymbol{\xi}$, with $\boldsymbol{\xi} = [\xi_1\ \xi_2 \cdots \xi_k]$.

$$\boldsymbol{\xi}_k = \boldsymbol{v}_k^T\boldsymbol{x}_c = \boldsymbol{v}_{k1}\boldsymbol{x}_1 + \boldsymbol{v}_{k2}\boldsymbol{x}_2 + \cdots + \boldsymbol{v}_{kp}\boldsymbol{x}_p \tag{3.2.3}$$

The variance-covariance matrix of the transformed data ξ is a diagonal matrix of eigenvalues.

$$var(\boldsymbol{\xi}) = var(\boldsymbol{V}^T\boldsymbol{x}_c) = \boldsymbol{V}^T var(\boldsymbol{x}_c)\boldsymbol{V} = \boldsymbol{V}^T\boldsymbol{VEV}^T\boldsymbol{V} = \boldsymbol{E} \tag{3.2.4}$$

Therefore, it is evident that the k-components of ξ are uncorrelated, and their variances are the eigenvalues. The proportion of variance explained by the reduced k-dimensional principal components is given by $Var_{exp} = \frac{\sum_i^k \lambda_i}{\sum_i^n \lambda_i}$.

3.2.3 Singular Value Decomposition

A unique singular value decomposition exists for any real matrix $\boldsymbol{X} \in \mathbb{R}^{\mathrm{n}\times\mathrm{p}}$.

$$\boldsymbol{X} = \boldsymbol{UDV}^T \tag{3.2.5}$$

Where $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{V} \in \mathbb{R}^{p\times p}$ are orthogonal matrices. The columns of $\boldsymbol{U}$ are called the left singular vectors and those of $\boldsymbol{V}$ right singular vectors. $\boldsymbol{D} \in \mathbb{R}^{n\times p}$ has positive singular values only for its diagonal entries, and the number of diagonal entries is equal to rank($\boldsymbol{X}$). It is generally assumed in this thesis that n > p holds for the data matrix $\boldsymbol{X}$. Finding the SVD of $\boldsymbol{X}$ is associated with spectral decompositions of the matrices $\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{XX}^T$.

The right singular vectors $\boldsymbol{V}$ are the eigenvectors of the matrix $\boldsymbol{X}^T\boldsymbol{X}$ as shown in equation (3.2.6).

$$\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{V}\boldsymbol{D}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T = \boldsymbol{V}(\boldsymbol{D}^T\boldsymbol{D})\boldsymbol{V}^T$$
$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{D}^T\boldsymbol{D})_{p\times p} \quad (3.2.6)$$

The left singular vectors $\boldsymbol{U}$ are the eigenvectors of the matrix $\boldsymbol{X}\boldsymbol{X}^T$ as shown in equation (3.2.7).

$$\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{D}^T\boldsymbol{U}^T = \boldsymbol{U}(\boldsymbol{D}\boldsymbol{D}^T)\boldsymbol{U}^T$$
$$\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{D}^T\boldsymbol{D})_{n\times n} \quad (3.2.7)$$

Equation (8) displays the SVD of the data matrix $\boldsymbol{X}$ when it has full column rank with n > p, where $\boldsymbol{u}_i$, i ∈ {1,2,…,n} is the i[th] column of the matrix of left singular vectors $\boldsymbol{U}$ and $\boldsymbol{v}_j^T$, j ∈ {1,2,…,p} is the j[th] row of the matrix of right singular vectors $\boldsymbol{V}$. The upper partition of matrix $\boldsymbol{D}$ is a p by p diagonal matrix of singular values in descending order and the lower patriation is a (n-p) by p matrix of zeros.

$$\boldsymbol{X}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}=[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \cdots\ \boldsymbol{u}_n]\begin{bmatrix}\sigma_1 & 0 & 0 & 0 & \cdots\\ 0 & \sigma_2 & 0 & 0 & \cdots\\ 0 & 0 & \sigma_3 & 0 & \cdots\\ \vdots & \vdots & \vdots & \ddots & \cdots\\ 0 & \vdots & \vdots & \vdots & \sigma_p\\ 0 & \vdots & \vdots & \vdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & \cdots & \cdots & \cdots & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{v}_1^T\\ \boldsymbol{v}_2^T\\ \vdots\\ \boldsymbol{v}_p^T\end{bmatrix}=\sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^T+\sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^T\cdots+\sigma_p\boldsymbol{u}_p\boldsymbol{v}_p^T=\sum_{\mathrm{i}}^{\mathrm{p}}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T \quad (3.2.8)$$

The result is p rank-1 matrices with linearly dependent rows and columns, which represents the principal components in descending importance.

Computationally, SVD and spectral decomposition are similar. SVD can be computed via the QR-SVD algorithm whereas the spectral decomposition can be found by the symmetric QR-algorithm. Both algorithms are based on orthogonal similarity transformations which preserve eigenvalues. The algorithms are iterative as for eigenvalue problems with p greater or equal to five, no general formula exists for the roots of the characteristic polynomial. In this case, the symmetric QR algorithm converges faster. However, SVD is more numerically stable as explicitly forming the variance-covariance matrix unnecessarily enlarges the condition number of the problem.

3.2.4 Stockpile Simulator

This chapter of the research looks into the effects of chevron and windrow stacking methods across a variety of different stockpile configurations. Simply put, chevron stacking method is done by stacking materials horizontally in one direction followed by stacking another layer of material on top in the opposite direction. Windrow stacking method puts down the materials in parallel rows with triangular cross-sections and then stacks more rows on top between the gaps using the multiple peaks. Chevron stacking tends to lead to particle segregation, whereas the windrow method does not bring about such concerns as it reduces fluctuations in particle size distribution by traversing the stacker much more frequently [24]. Due to the complexity of blending operations

in real life, it is very difficult, if not impossible, to build a model that perfectly replicates their effects. Hence relatively simple linear block models are built, in MATLAB, to simulate the effects of chevron and windrow blending methods. Similar to the simulator developed by Marques 2013, the stockpile simulator in this research is essentially a homogenization simulator for linear cuboid stockpile [31]. The input to the simulator is a series of predefined mining sequences, and the output is the blocks re-arranged by the algorithm. The simulated stockpiles are defined by three parameters, namely the stockpile height (h), length (l) and width (w). The stockpile capacity can be found as $\text{Capacity} = \text{h} \times \text{l} \times \text{w}$. Chevron stockpiles are simulated by laying down blocks along the direction of the stockpile length until the predefined stockpile length (l) is reached, then laying more blocks on the next height level in the opposite direction. Stockpile widths are the simulated chevron stockpiles are set to 1. In the case of the windrow stockpiles, blocks are laid down in the direction of the stockpile length, when the row is filled (stockpile length is reached), another row is added in the stockpile width direction, but blocks in the row are laid down in a direction opposite to the previous row. This process is repeated until the stockpile width is reached, after which more rows of blocks are put down on top but with rows and blocks in a row laid down in opposite directions. In other words, the direction of laying down blocks reverse with each increment of stockpile width, and direction of putting down rows of blocks reverse with each increment of stockpile height. This process is repeated until the stockpile height is reached. The reclaiming process is simulated by taking the average grades of all blocks in the same reclaiming slice, i.e., all blocks with the same stockpile length (l) value. In other terms, each reclaiming slice has $\text{h} \times \text{w}$ number of blocks, and the stockpile has a total of h layers with each layer having $\text{l} \times \text{w}$ blocks.

Effect of the blending process is evaluated primarily using the Variance Reduction Ratio (VRR). VRR is given by [22]

$$\text{VRR} = \frac{\sigma_{out}^2}{\sigma_{in}^2} \tag{3.2.9}$$

Where σ_{out}^2 and σ_{in}^2 are, respectively, the output and input variances. It is of paramount importance that the VRR is calculated based on the same weight or volume, and in the case of this chapter, the number of blocks of material. Since the output of the simulation finds takes the average grade of all blocks within the same reclaiming slice, mean of the same number of blocks is calculated while finding the input variance.

The multiple numbers of different stockpile configurations are tested for the two stacking methods, and VRR value is calculated for each variable in each configuration scenario. Illustration of chevron and windrow stockpiles are shown in Figure 3.1 and Figure 3.2 respectively.

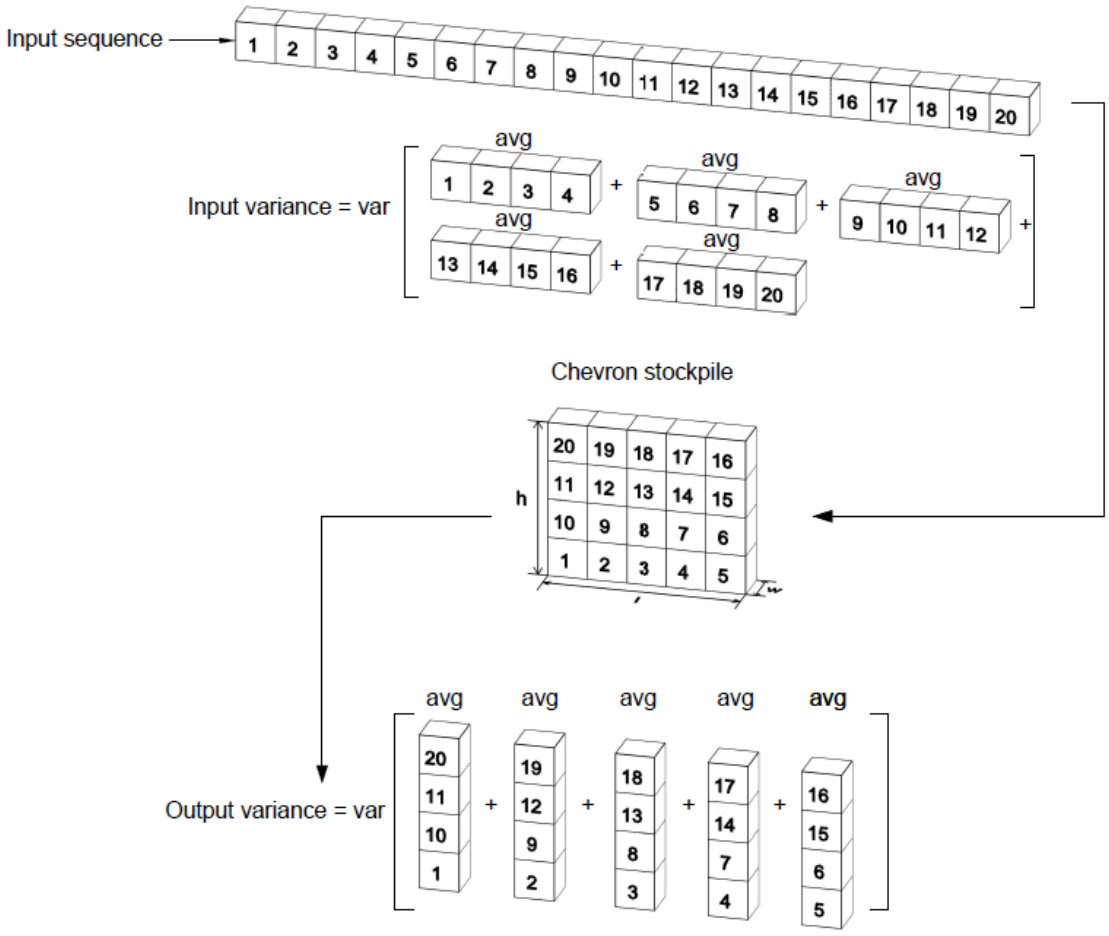


Figure 3.1. Chevron stockpile illustration

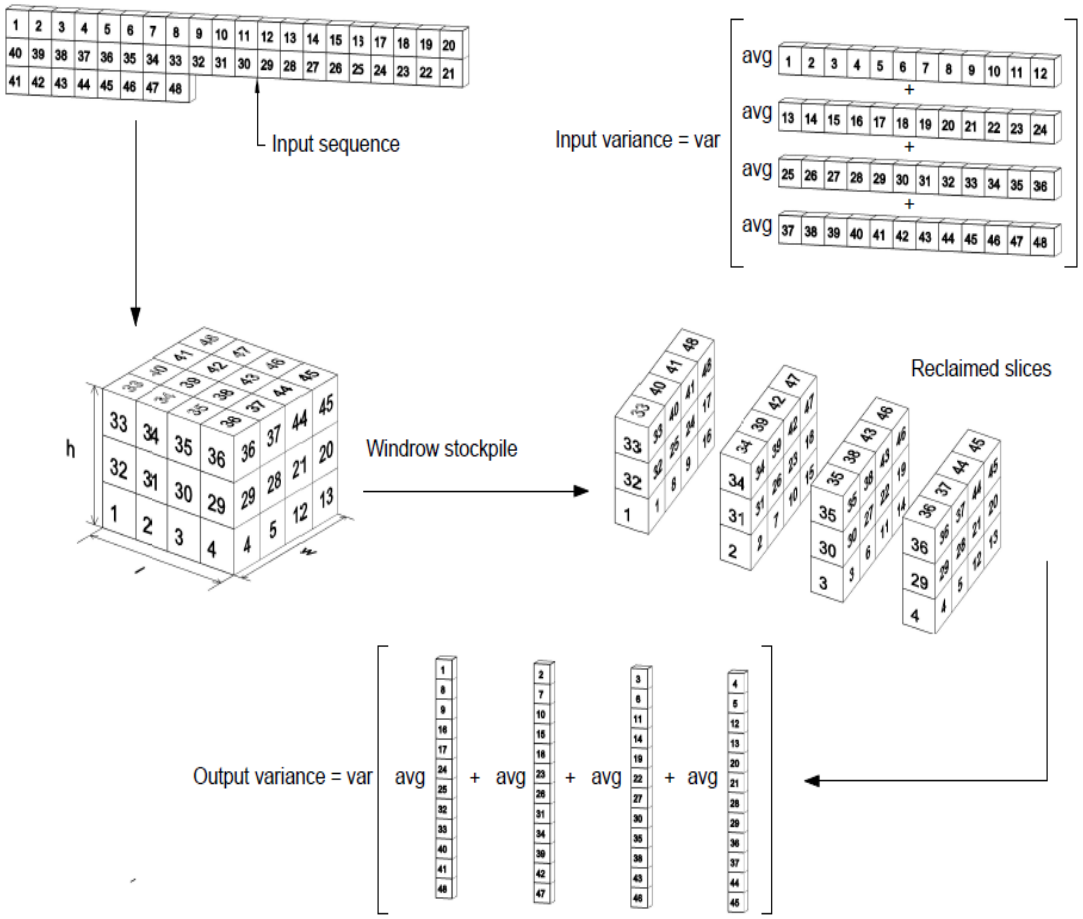


Figure 3.2. Windrow stockpile illustration

3.3 Bed Blending Case Study

3.3.1 Input Data and PCA Results

The case study has a data input of 15,000 blocks containing grade information for iron, silica, alumina, and lime. The input data was briefly analysed and run through the PCA algorithm in R.

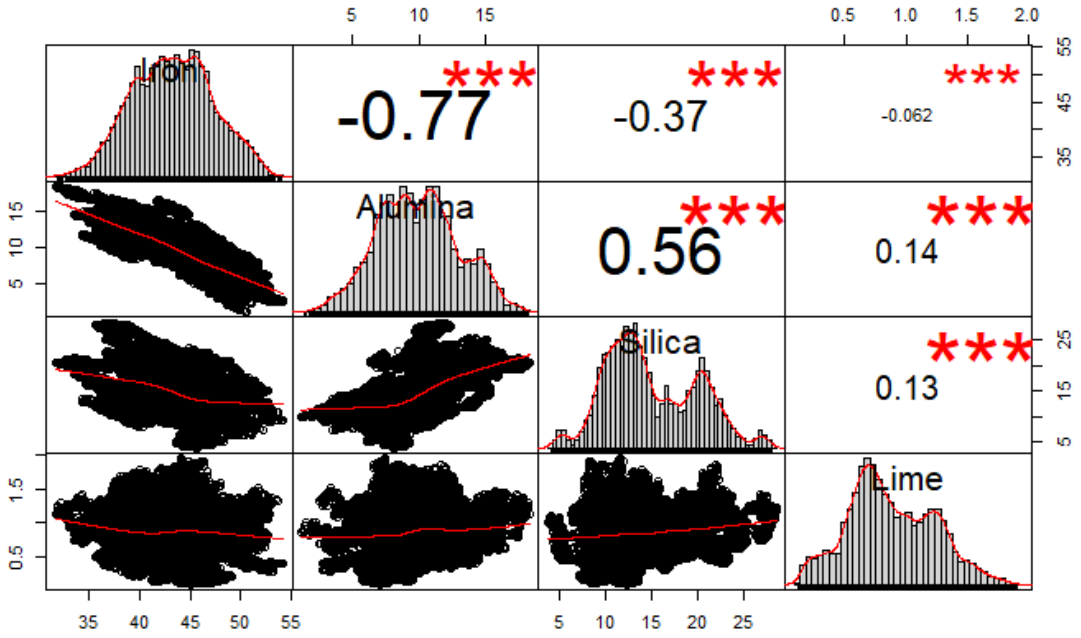


Figure 3.3. Scatterplot matrix of case study input data

Figure 3.3 shows the scatterplot matrix for the input data, with the lower left panels being scatterplots, diagonal panels being histograms of each variable and upper right panels being correlations between variables. It could be observed that the input variables, i.e., mineral grades have very complex relationships with each other and are highly correlated, except for lime.

PCA is conducted on the dataset using the prcomp() function in R, the data matrix is centred and scaled and PCA is done via the Singular Value Decomposition (SVD) method. Table 3.1 shows the proportion of variance explained by each of the principal components. Since PC1 along only accounts for 54.3% of the original variation in the dataset, the first two principal components are used so that approximately 80% of the variation is preserved.

Table 3.1. Importance of principal components

	Principal component 1	Principal component 2	Principal component 3	Principal component 4
Standard deviation	1.474	0.990	0.807	0.443
Proportion of Variance	0.543	0.245	0.163	0.049
Cumulative Proportion	0.543	0.788	0.951	1.000

Table 3.2 displays the principal component loadings which are essentially sorted eigenvectors

based on their corresponding eigenvalue in descending order. The principal components are linear combinations of the original variables and the loadings represent the relative coefficients. In other words, variables that have large loadings contribute more to a certain principal component. In the case of this dataset, iron and alumina are the primary contributors to PC1 while lime contributes overwhelmingly to PC2.

Table 3.2. Principal component loadings

	Principal component 1	Principal component 2	Principal component 3	Principal component 4
Iron	0.576	0.200	0.506	0.610
Alumina	-0.631	-0.090	-0.156	0.755
Silica	-0.496	0.036	0.835	-0.237
Lime	-0.158	0.975	-0.149	-0.046

Fundamentally Table 3.2 means the two principal components used to reconstruct the data have forms as follows:

$$\begin{aligned} PC1 &= 0.576 \times Iron - 0.631 \times Alumina - 0.496 \times Silica - 0.158 \times Lime \\ PC2 &= 0.2 \times Iron - 0.09 \times Alumina + 0.036 \times Silica + 0.975 \times Lime \end{aligned} \tag{3.3.1}$$

Figure 3.4 is a biplot of the principal component scores i.e. the transformed/reduced data. The x and y-axis represent standardized PC1 and PC2 scores respectively. The four vectors are the transformed variables, which are essentially original variables rebuilt with the chosen principal components. Quality of representation of each vector by the chosen two PCs is displayed in different colours based on their respective squared cosine values. For any given variable, the sum of the squared cosines from all PCs should be equal to one. Since the reduced data only consists of two PCs, the better a variable is represented by these two PCs, the closer is it to the circumference of the circle [32]. For this dataset, the first two principal components represent fairly well lime, alumina, and iron, but some of the information from silica is lost from the transformation. The correlations between variables are largely preserved as can be told from the angles between vectors.

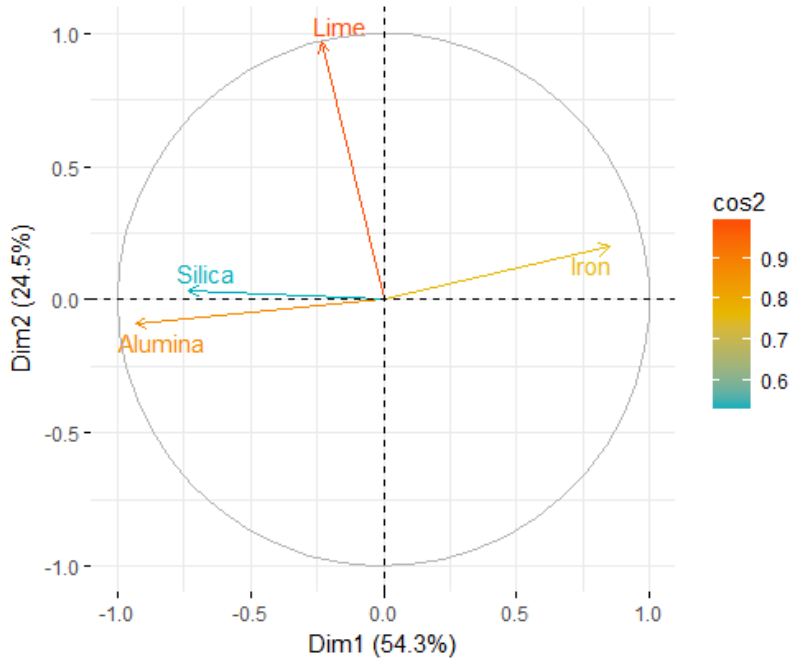


Figure 3.4. Principal component biplot

3.3.2 Output Data Analysis

45 stockpile scenarios in total were generated, 15 of them being chevron and the rest windrow, half of the windrow scenarios were created by switching the values of stockpile height and width. For all scenarios, the stockpile capacity is kept at a constant of 15,000 blocks. The input data and the principal components are run through the stockpile simulator and the resulting VRR values are shown in Table 3.3 and Table 3.4.

Table 3.3. Chevron stockpile output

Height	Length	Width	VRR_{Iron}	VRR_{Silic}	VRR_{Alu}	VRR_{Lim}	VRR_{PC1}	VRR_{PC2}
5	3000	1	0.0575	0.0574	0.1134	0.1517	0.0719	0.1403
6	2500	1	0.1480	0.1565	0.0425	0.2959	0.1126	0.3209
8	1875	1	0.1960	0.0813	0.0970	0.0555	0.1486	0.0580
10	1500	1	0.0192	0.0231	0.0098	0.0874	0.0152	0.0858
12	1250	1	0.0636	0.1281	0.0189	0.0239	0.0759	0.0263
15	1000	1	0.0200	0.0205	0.0091	0.0080	0.0122	0.0079
20	750	1	0.0060	0.0090	0.0037	0.0085	0.0049	0.0098
24	625	1	0.0169	0.0142	0.0088	0.0183	0.0088	0.0216
25	600	1	0.0101	0.0074	0.0127	0.0230	0.0072	0.0228
30	500	1	0.0047	0.0073	0.0015	0.0037	0.0049	0.0020
40	375	1	0.0018	0.0010	0.0014	0.0041	0.0009	0.0048
50	300	1	0.0074	0.0054	0.0015	0.0045	0.0048	0.0073
60	250	1	0.0008	0.0008	0.0003	0.0005	0.0008	0.0004
75	200	1	0.0011	0.0013	0.0006	0.0006	0.0012	0.0005
100	150	1	0.0020	0.0018	0.0010	0.0003	0.0020	0.0006

Table 3.4. Windrow stockpile output

Height	Length	Width	VRR_{Iron}	VRR_{Silic}	VRR_{Alu}	VRR_{Lim}	VRR_{PC1}	VRR_{PC2}
1	3000	5	0.0575	0.0574	0.1134	0.1517	0.0719	0.1403
5	3000	1	0.0575	0.0574	0.1134	0.1517	0.0719	0.1403
2	2500	3	0.1480	0.1565	0.0425	0.2959	0.1126	0.3209
3	2500	2	0.1480	0.1565	0.0425	0.2959	0.1126	0.3209
2	1875	4	0.1960	0.0813	0.0970	0.0555	0.1486	0.0580
4	1875	2	0.1960	0.0813	0.0970	0.0555	0.1486	0.0580
2	1500	5	0.0192	0.0231	0.0098	0.0874	0.0152	0.0858
5	1500	2	0.0192	0.0231	0.0098	0.0874	0.0152	0.0858
2	1250	6	0.0636	0.1281	0.0189	0.0239	0.0759	0.0263
6	1250	2	0.0636	0.1281	0.0189	0.0239	0.0759	0.0263
3	1000	5	0.0200	0.0205	0.0091	0.0080	0.0122	0.0079
5	1000	3	0.0200	0.0205	0.0091	0.0080	0.0122	0.0079
4	750	5	0.0060	0.0090	0.0037	0.0085	0.0049	0.0098
5	750	4	0.0060	0.0090	0.0037	0.0085	0.0049	0.0098
3	625	8	0.0169	0.0142	0.0088	0.0183	0.0088	0.0216
8	625	3	0.0169	0.0142	0.0088	0.0183	0.0088	0.0216
5	600	5	0.0101	0.0074	0.0127	0.0230	0.0072	0.0228
5	600	5	0.0101	0.0074	0.0127	0.0230	0.0072	0.0228
3	500	10	0.0047	0.0073	0.0015	0.0037	0.0049	0.0020
10	500	3	0.0047	0.0073	0.0015	0.0037	0.0049	0.0020
4	375	10	0.0018	0.0010	0.0014	0.0041	0.0009	0.0048
10	375	4	0.0018	0.0010	0.0014	0.0041	0.0009	0.0048
2	300	25	0.0074	0.0054	0.0015	0.0045	0.0048	0.0073
25	300	2	0.0074	0.0054	0.0015	0.0045	0.0048	0.0073
3	250	20	0.0008	0.0008	0.0003	0.0005	0.0008	0.0004
20	250	3	0.0008	0.0008	0.0003	0.0005	0.0008	0.0004
3	200	25	0.0011	0.0013	0.0006	0.0006	0.0012	0.0005
25	200	3	0.0011	0.0013	0.0006	0.0006	0.0012	0.0005
5	150	20	0.0020	0.0018	0.0010	0.0003	0.0020	0.0006
20	150	5	0.0020	0.0018	0.0010	0.0003	0.0020	0.0006

For the windrow scenarios, it could be seen just by switching the values for width and height does not change the VRR values at all. Moreover, for both windrow and chevron stacking, VRR is generally minimized by reducing the stockpile length, which is equivalent to increasing the number of blocks in each reclaiming slice.

3.3.3 Autocorrelation and the Effectiveness of Blending

The effectiveness of blending operations regarding this particular dataset is very high with generally very low VRR values as the data is strongly autocorrelated, as shown in Figure 3.5. The number of lags was chosen to be one-tenth the size of the dataset which is 1500. It could be seen that there exists significant autocorrelation for all 4 variables, far exceeding the 95% quantile for noise.

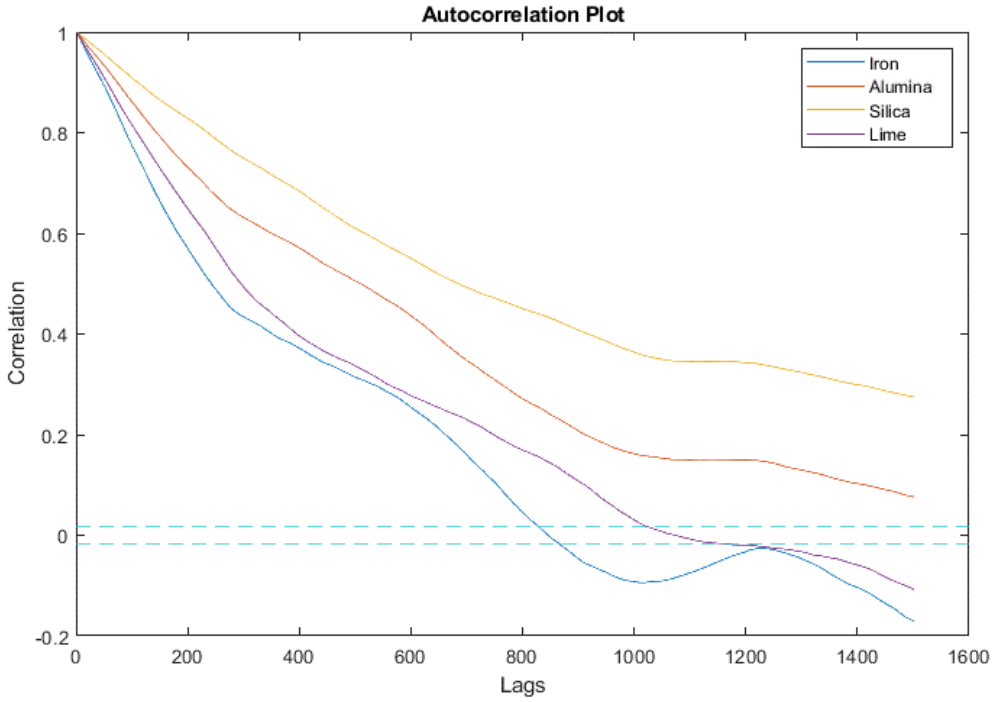


Figure 3.5. Autocorrelation plot for original dataset

An alternative simulated dataset was generated through Monte-Carlo simulation in R, removing autocorrelation while preserving the correlation between variables. The scatterplot matrix depicting the simulated data is shown in Figure 3.6. Figure 3.7 shows the autocorrelation of the variables in the simulated data.

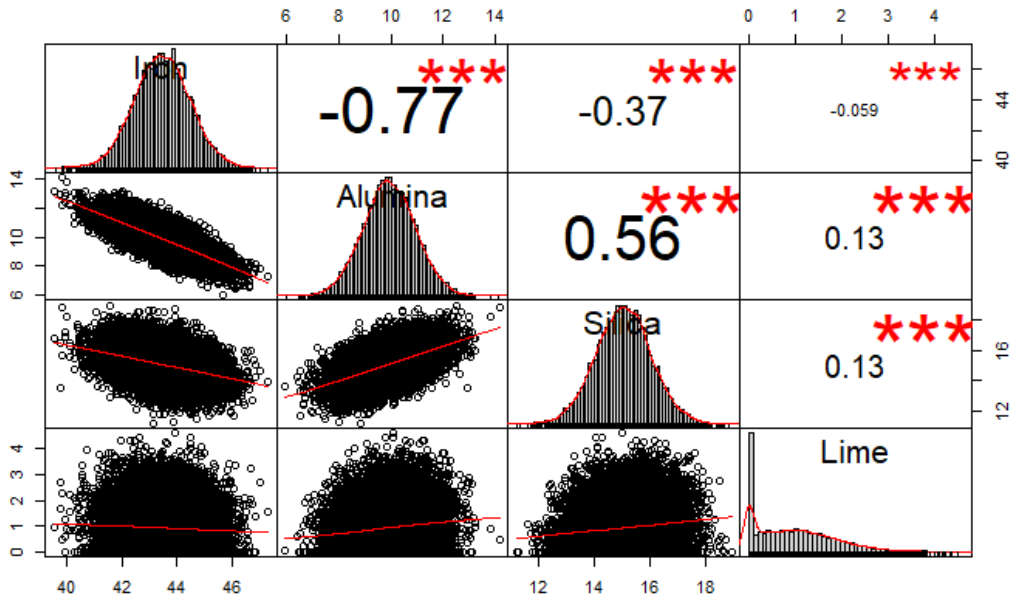


Figure 3.6. Scatterplot matrix of simulated data

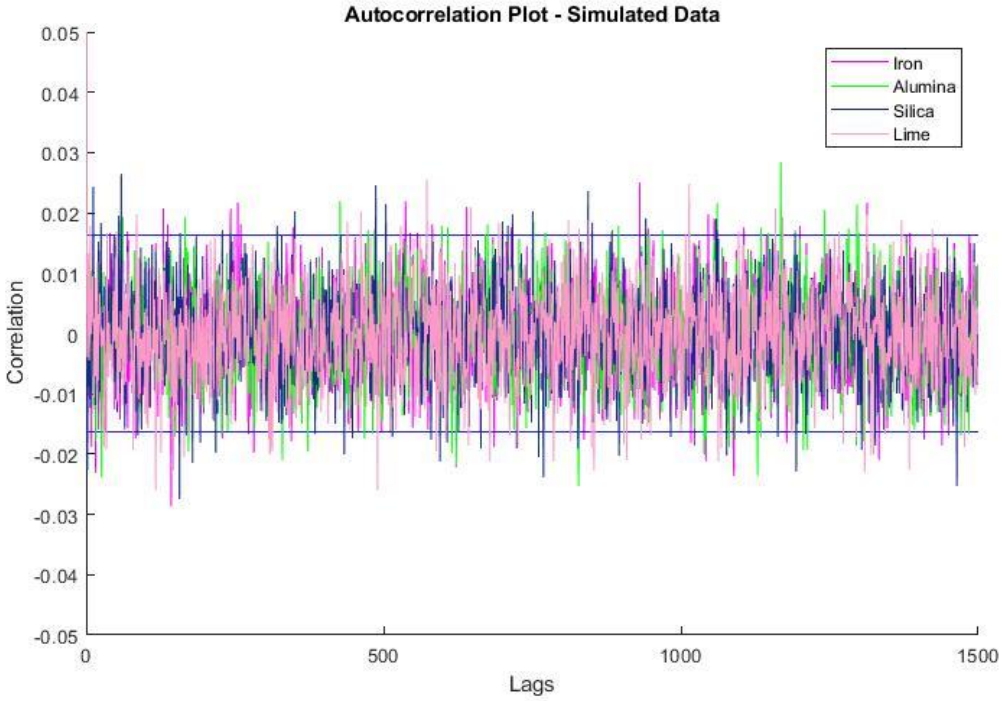


Figure 3.7. Autocorrelation plot of simulated data

Variables of the simulated data exhibit no autocorrelation as the plot follows no obvious pattern

and lies almost entirely within the 95% noise region. The simulated data were run through the same stockpile simulator with identical stockpile configurations. The results are shown in Table 3.5 and Table 3.6.

Table 3.5. Chevron stockpile output – simulated data

Height	Length	Width	VRR_{Iron}	$VRR_{Alumina}$	VRR_{Silica}	VRR_{Lime}	VRR_{PC1}	VRR_{PC2}
5	3000	1	1.015	1.028	1.093	0.965	1.052	0.970
6	2500	1	0.995	1.013	1.052	0.950	1.020	0.961
8	1875	1	1.028	1.014	1.052	0.913	1.011	0.931
10	1500	1	0.982	0.979	1.121	0.909	1.020	0.914
12	1250	1	0.996	1.034	1.086	0.924	1.032	0.923
15	1000	1	1.067	0.986	1.140	0.907	1.058	0.913
20	750	1	1.026	0.972	1.103	0.851	1.017	0.877
24	625	1	1.071	1.120	1.169	0.915	1.110	0.953
25	600	1	1.265	1.115	1.148	0.849	1.186	0.847
30	500	1	1.076	0.956	1.130	0.813	1.013	0.836
40	375	1	1.176	1.003	1.144	0.865	1.088	0.919
50	300	1	1.225	1.123	0.962	0.818	1.103	0.841
60	250	1	1.216	1.004	0.994	0.851	1.016	0.932
75	200	1	1.679	1.188	1.092	0.864	1.317	0.871
100	150	1	1.561	1.443	1.003	0.741	1.375	0.809

Table 3.6. Windrow stockpile output – simulated data

Height	Length	Width	VRR_{Iron}	VRR_{Alu}	VRR_{Silic}	VRR_{Lim}	VRR_{PC1}	VRR_{PC2}
1	3000	5	1.015	1.028	1.093	0.965	1.052	0.970
5	3000	1	1.015	1.028	1.093	0.965	1.052	0.970
2	2500	3	0.995	1.013	1.052	0.950	1.020	0.961
3	2500	2	0.995	1.013	1.052	0.950	1.020	0.961
2	1875	4	1.028	1.014	1.052	0.913	1.011	0.931
4	1875	2	1.028	1.014	1.052	0.913	1.011	0.931
2	1500	5	0.982	0.979	1.121	0.909	1.020	0.914
5	1500	2	0.982	0.979	1.121	0.909	1.020	0.914
2	1250	6	0.996	1.034	1.086	0.924	1.032	0.923
6	1250	2	0.996	1.034	1.086	0.924	1.032	0.923
3	1000	5	1.067	0.986	1.140	0.907	1.058	0.913
5	1000	3	1.067	0.986	1.140	0.907	1.058	0.913
4	750	5	1.026	0.972	1.103	0.851	1.017	0.877
5	750	4	1.026	0.972	1.103	0.851	1.017	0.877
3	625	8	1.071	1.120	1.169	0.915	1.110	0.953
8	625	3	1.071	1.120	1.169	0.915	1.110	0.953
5	600	5	1.265	1.115	1.148	0.849	1.186	0.847
5	600	5	1.265	1.115	1.148	0.849	1.186	0.847
3	500	10	1.076	0.956	1.130	0.813	1.013	0.836
10	500	3	1.076	0.956	1.130	0.813	1.013	0.836
4	375	10	1.176	1.003	1.144	0.865	1.088	0.919
10	375	4	1.176	1.003	1.144	0.865	1.088	0.919
2	300	25	1.225	1.123	0.962	0.818	1.103	0.841
25	300	2	1.225	1.123	0.962	0.818	1.103	0.841
3	250	20	1.216	1.004	0.994	0.851	1.016	0.932
20	250	3	1.216	1.004	0.994	0.851	1.016	0.932
3	200	25	1.679	1.188	1.092	0.864	1.317	0.871
25	200	3	1.679	1.188	1.092	0.864	1.317	0.871
5	150	20	1.561	1.443	1.003	0.741	1.375	0.809
20	150	5	1.561	1.443	1.003	0.741	1.375	0.809

VRR values for the simulated scenarios all approximate 1, which means that for a dataset without autocorrelation, the effects of blending operations are insignificant.

3.3.4 Regression Analysis

We used multiple regression to identify the relationships between the VRRs of the input materials and the design parameters of the stockpile and used stepwise regression to choose regressors that best describe the models. The possible predictor variables are the stockpile length, width, and height, is windrow (a binary factor variable that equals 0 if the chevron stockpile is used, 1 otherwise), as well as all their first-order interactions and the second-order terms of stockpile

length, width and height. For each response variable, a model was forwardly selected from an initial model with intercepts only, backwardly eliminated another model from an initial model with all possible predictors, and selected a third and final model stepwise that initially consists of only the four main effects. The variable selection criterion is based on Akaike's information criterion (AIC), which measures the closeness between the sample fit and true model fit, where the relative closeness is defined as the Kullback–Leibler divergence from the true model [18]. The AIC can be calculated as follows: AIC = –2(Maximum loglikelihood – Number of parameters), and models with lower AIC values are generally preferred. We performed this process using R software with the stepAIC() function. Table 3.7 shows an illustration of the process of finding the best model for VRR_{Iron}.

Table 3.8 shows the final results for all the response variables.

Table 3.7. Stepwise VRR_{Iron} model selection

VRR_iron_step1$anova						
Initial Model: vrr_iron ~ 1						
Final Model: vrr_iron ~ length + I(length^2)						
Step	name	Df	Deviance	Resid. Df	Resid. Dev	AIC
1	—	—	—	44	0.145221	−256.13
2	+ length	1	0.07402	43	0.071201	−286.20
3	+ I(length^2)	1	0.008424	42	0.062777	−289.87

VRR_iron_step2$anova						
Initial Model: vrr_iron ~ height + length + width + iswindrow						
Final Model: vrr_iron ~ height + length + width + I(length^2) + height:width						
Step	name	Df	Deviance	Resid. Df	Resid. Dev	AIC
1	—	—	—	40	0.071188	−280.21
2	+ I(length^2)	1	0.012905	39	0.058283	−287.21
3	+ height:width	1	0.006292	38	0.051991	−290.35
4	– iswindrow	1	1.03×10^{-5}	39	0.052002	−292.34

VRR_iron_step3$anova						
Initial Model: vrr_iron ~ (height + length + width + iswindrow)^2 + I(height^2) + I(length^2) + I(width^2)						
Final Model: vrr_iron ~ height + length + width + iswindrow + I(length^2) + height:length + height:iswindrow + length:width + length:iswindrow						
Step	name	Df	Deviance	Resid. Df	Resid. Dev	AIC
1	—	—	—	32	0.039573	−290.63
2	– width:iswindrow	0	0	32	0.039573	−290.63
3	– I(width^2)	1	0.000717	33	0.040291	−291.82
4	– height:width	1	0.001334	34	0.041625	−292.36
5	– I(height^2)	1	0.001859	35	0.043484	−292.39

Table 3.8. Regression results

Models	Adjusted R-squared	AIC
$VRR_{Iron}=1.294\times10^{-3}\times height + 2.384\times10^{-4}\times length + 5.875\times10^{-3}\times width - 3.319\times10^{-1}\times IsWindrow - 4.835\times10^{-8}\times length^2 - 1.913\times10^{-5}\times(height\times length) + 4.681\times10^{-3}\times(height\times IsWindrow) + 6.723\times10^{-5}\times(length\times IsWindrow) - 1.913\times10^{-5}\times(length\times width) + 1.412\times10^{-1}$	0.623	−292.4
$VRR_{Alumina} = 1.651\times10^{-4}\times height + 9.612\times10^{-5}\times length - 1.811\times10^{-8}\times length^2 - 3.269\times10^{-2}$	0.57	−357.0
$VRR_{Silica} = 1.171\times10^{-8}\times length^2 - 1.571\times10^{-3}$	0.75	−361.4
$VRR_{Lime} = -6.454\times10^{-4}\times height + 1.187\times10^{-4}\times length + 4.283\times10^{-3}\times width - 3.319\times10^{-1}\times IsWindrow - 2.693\times10^{-5}\times(height\times length) + 3.638\times10^{-3}\times(height\times IsWindrow) + 9.391\times10^{-5}\times(length\times IsWindrow) - 2.693\times10^{-5}\times(length\times width) + 3.338\times10^{-1}$	0.73	−287.1
$VRR_{PC1}=1.166\times10^{-4}\times height + 1.585\times10^{-4}\times length + 4.187\times10^{-4}\times width - 3.229\times10^{-8}\times length^2 + 8.414\times10^{-4}\times(height\times width) - 1.015\times10^{-1}$	0.67	−321.5
$VRR_{PC2} = -6.338\times10^{-4}\times height + 1.240\times10^{-4}\times length + 4.601\times10^{-3}\times width - 3.840\times10^{-1}\times IsWindrow - 3.122\times10^{-5}\times(height\times length) + 3.967\times10^{-3}\times(height\times IsWindrow) + 1.090\times10^{-4}\times(length\times IsWindrow) - 3.122\times10^{-5}\times(length\times width) + 3.986\times10^{-1}$	0.69	−277.1

The resulting model for the principal components is similar to but differs from those for the rest of the variables. Optimizing the principal components rather than the original variables will lead to different stockpile design parameters. However, as PCA retains as much information as possible during the transformation, the design that minimizes the VRR of the PCs is clearly the mathematically optimal design that aims to minimize the variances for all input variables. This effectively addresses the issue of having to assign a weight or importance to each variable.

We note that the interaction term for the stockpile height and width is unique for the VRR_{PC1} model but otherwise, minimizing VRR_{PC1} and VRR_{PC2} is equivalent to minimizing $\sum w_i VRR_i$,, where w_i refers to the weight of each variable and should be set to equal to the sum of the factor loadings of the principal components.

3.4 Chapter Summary

Principal component analysis (PCA) can be used in conjunction with multiple regression to design and optimize stockpiles when there are multiple types of materials whose output grades must be controlled. The performance and benefit of applying PCA may potentially increase with the number of material-grade variables studied. Input data that are autocorrelated have a significant impact on the performance of the stockpiles, with reduced variance reduction ratios (VRRs) for increased levels of autocorrelation. The multiple regression results of Table 5.2 have relatively low adjusted R-squared values, which may be due to some of the variance being uniquely determined by the degree of autocorrelation in the block input. Nevertheless, it was found that the VRR is generally reduced with an increasing number of reclamation slices (length) and that the performances of the windrow and chevron methods do not differ significantly. However, additional scenarios and data input are needed to better determine the effects of the design parameters on the VRR.

Chapter 4

Linear and Generalized Models in Quality Control, Reliability and Safety in Mining Engineering

4.1 Modelling of Non-linear Relationships

4.1.1 Applications of Regression Analysis in Mining Engineering

The optimization of mining systems including reliability of mining equipment, safety of assets and personnel has become an increasingly important topic. The costliness of some large mining equipment means that downtimes or failures tend to be associated with high costs. Other issues such as improving occupational safety by studying accidents in workplace and optimizing product quality characteristics are also imperative in augmenting the overall profitability of a mining operation [33, 34]. In these regards, statistical techniques including regression analysis are helpful both in predicting future values and in helping researchers understand the underlying causal relationships among variables. Indeed, in recent years numerous advanced regression and machine learning techniques are put forward to better analyze data in engineering quality control, personnel safety and equipment reliability analysis [35-37]. However, nonlinear relationships among variables is a very commonly encountered phenomenon when modelling these kinds of engineering systems. While nonparametric machine learning techniques tend to have fairly accurate prediction results, parametric models such as the generalized linear models are a lot more interpretable and therefore could also be used to study the interactions between variables [38]. This study aims to show that generalized linear models are suitable tools in modelling engineering problems in many cases, including many potential problems in the mining industry, such as safety related issues, optimization of quality characteristics of systems or products and equipment reliability.

4.1.2 The Challenger Example

Nonlinear relationships among variables is a very common phenomenon in many disciplines of engineering. A classic example is the O-ring data which came from experiments conducted on the O-rings that eventually led to the Challenger space shuttle disaster. As shown in Figure 4.1, generalized linear could successfully capture the discreteness within the data, when the assumptions of linear model do no hold.

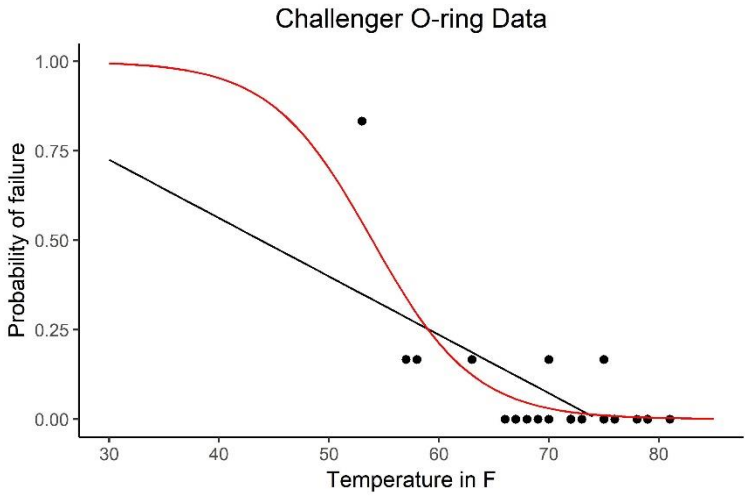


Figure 4.1. Illustration of linear model and generalized linear model for the O-ring data (Linear model in black and logistic GLM in red)

This study aims to show that generalized linear models are suitable tools in modelling engineering problems in many cases, including many potential problems in the mining industry, such as safety related issues and equipment reliability.

4.2 Methodology and Mathematical Intuitions

4.2.1 The Ordinary Linear Model

The most commonly used linear model, i.e. the ordinary linear model, has the following form:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{4.2.1}$$

Where $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^T$ is the $n \times 1$ vector of independent observations, with $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{y}) = (\mu_1, \dots, \mu_n)^T$, $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is the design matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of predictors and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the error term which represents both the measurement errors and random fluctuations. Generally, it is assumed that the model has homoscedastic error with mean 0, i.e. $\mathbb{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ with $var(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$. We are interested in obtaining the best fitting $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ with respect to the ordinary linear model. Intuitively this could be achieved by minimizing the residual, i.e. $\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{argmin} \ \|\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\|_2^2 = \underset{\boldsymbol{\beta}}{argmin} \ \|\hat{\boldsymbol{\epsilon}}\|_2^2$. Consider the QR-decomposition of the design matrix, given by $\boldsymbol{X} = \boldsymbol{Q}\begin{bmatrix}\boldsymbol{R} \\ \boldsymbol{0}\end{bmatrix}$, where $\boldsymbol{Q} = [\underset{p}{\boldsymbol{Q}_1} \ \underset{n-p}{\boldsymbol{Q}_2}] \in \mathbb{R}^{n \times n}$ is orthogonal and $\boldsymbol{R} \in \mathbb{R}^{p \times p}$ is upper triangular. It follows that:

$$\begin{aligned}\|\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\|_2^2 &= \left\|\boldsymbol{Q}^T(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})\right\|_2^2 = \left\|\boldsymbol{Q}^T\boldsymbol{y}-\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{bmatrix}\widehat{\boldsymbol{\beta}}\right\|_2^2 \\ &= \left\|\begin{matrix}\boldsymbol{Q}_1^T\boldsymbol{y}-\boldsymbol{R}\widehat{\boldsymbol{\beta}}\\ \boldsymbol{Q}_2^T\boldsymbol{y}-\boldsymbol{0}\end{matrix}\right\|_2^2 = \left\|\boldsymbol{Q}_1^T\boldsymbol{y}-\boldsymbol{R}\widehat{\boldsymbol{\beta}}\right\|_2^2 + \left\|\boldsymbol{Q}_2^T\boldsymbol{y}\right\|_2^2\end{aligned} \tag{4.2.2}$$

It is obvious that the terms in Equation (4.2.2) is minimized when $\left\|\boldsymbol{Q}_1^T\boldsymbol{y}-\boldsymbol{R}\widehat{\boldsymbol{\beta}}\right\|_2^2 = 0$ and consequently $\boldsymbol{Q}_1^T\boldsymbol{y} = \boldsymbol{R}\widehat{\boldsymbol{\beta}}$. It follows that the optimal residual is within the range of $\boldsymbol{Q}_2$, and therefore the orthogonal complement of the design matrix, as shown in Equation (4.2.3).

$$\begin{aligned}\hat{\boldsymbol{\epsilon}} &= \boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y}-\boldsymbol{Q}_1\boldsymbol{R}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{y}-\boldsymbol{Q}_1\boldsymbol{Q}_1^T\boldsymbol{y} = (\boldsymbol{I}-\boldsymbol{Q}_1\boldsymbol{Q}_1^T)\boldsymbol{y} \\ &= \boldsymbol{Q}_2\boldsymbol{Q}_2^T\boldsymbol{y} \in \mathcal{R}^{\perp}(\boldsymbol{X})\end{aligned} \tag{4.2.3}$$

It follows that an explicit solution for $\widehat{\boldsymbol{\beta}}$ could be found in Equation (4.2.4). This way of solving for $\widehat{\boldsymbol{\beta}}$ is called the normal equation method.

$$\begin{aligned}\boldsymbol{X}^T\hat{\boldsymbol{\epsilon}} &= \boldsymbol{X}^T\boldsymbol{y}-\boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{0} \\ \boldsymbol{X}^T\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}^T\boldsymbol{y} \\ \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}^{\dagger}\boldsymbol{y} \\ \widehat{\boldsymbol{\mu}} &= \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{y} = \boldsymbol{H}\boldsymbol{Y}\end{aligned} \tag{4.2.4}$$

Where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is the hat matrix, and $\boldsymbol{X}^{\dagger}$ is the Moore-Penrose generalized inverse of $\boldsymbol{X}$. The predicted value has form: $\widehat{\boldsymbol{\mu}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{y} = Proj_{\mathcal{R}(\boldsymbol{X})}\cdot\boldsymbol{y}$, i.e. The least squares normal equations make predictions by projecting of the response onto the column space of the design matrix. A geometric illustration is displayed in Figure 4.2.

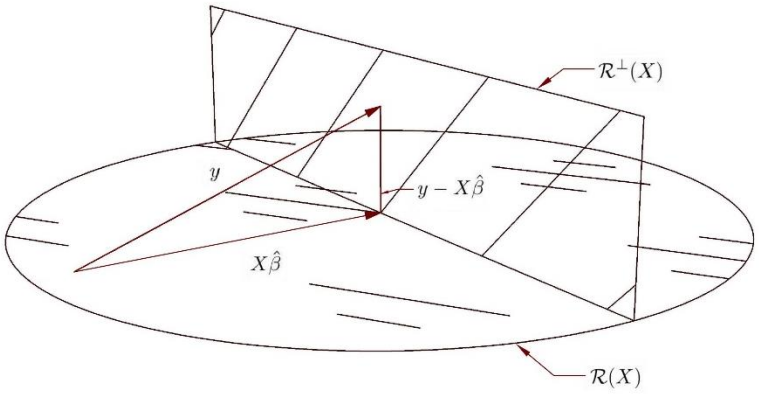


Figure 4.2. Illustration of data projection by least squares normal equations

By taking the squared two norms of the orthogonally decomposed vectors, using Pythagorean

theorem and some algebraic manipulation, one can arrive at the sum-of-squares decomposition as shown in Equation (4.2.5).

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \hat{\boldsymbol{\epsilon}} = \widehat{\boldsymbol{\mu}} + \hat{\boldsymbol{\epsilon}} \\ \boldsymbol{y} - \overline{\boldsymbol{y}} \cdot \mathbb{1}_n &= \widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \cdot \mathbb{1}_n + \hat{\boldsymbol{\epsilon}} \\ ||\boldsymbol{y} - \overline{\boldsymbol{y}} \cdot \mathbb{1}_n||_2^2 &= ||\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}} \cdot \mathbb{1}_n||_2^2 + ||\hat{\boldsymbol{\epsilon}}||_2^2 \\ \sum_{i=1}^{n} (\boldsymbol{y_i} - \overline{\boldsymbol{y}})^2 &= \sum_{i=1}^{n} (\widehat{\boldsymbol{y}} - \overline{\boldsymbol{y}})^2 + \sum_{i=1}^{n} (\boldsymbol{y} - \widehat{\boldsymbol{y}})^2 \\ SST &= SSReg + SSRes \end{aligned} \tag{4.2.5}$$

Where the Total Sum of Squares (SST) is the variation in the data explained by the intercept only model, Regression Sum of Squares (SSReg) is the variation in the data explained by the full model and Residual Sum of Squares (SSR) is the variation in the data the is left unexplained. The predictive power of a linear model could be summarized by the well-known R-squared metric, with $R^2 = \frac{SSReg}{SST}$, signifying the percentage of variation in the data that is explained by the regression coefficients.

4.2.2 Generalized Linear Models

4.2.2.1 Components of a GLM

A generalized linear model has three components, a random component of response $\boldsymbol{Y}$, with independent observations, that follows the exponential-dispersion family:

$$f_Y(\boldsymbol{y}; \theta, \phi) = \exp\{\frac{\boldsymbol{y} \cdot \theta - b(\theta)}{a(\phi)} + c(\boldsymbol{y}, \phi)\} \tag{4.2.6}$$

Where θ is termed the canonical parameter and ϕ the dispersion parameter. A systematic component made up of the $n \times p$ design matrix $\boldsymbol{X}$ and p-dimensional regression coefficients $\boldsymbol{\beta}$, with $\boldsymbol{\eta} = \boldsymbol{X\beta}$, and a monotone, differentiable link function $g(\cdot)$ that maps the mean of the response to the linear predictor. In summary, a GLM has form as follows:

$$g[\mathbb{E}(\boldsymbol{Y})] = \boldsymbol{X\beta} \tag{4.2.7}$$

The link function that maps the mean to the canonical parameter is called the canonical link and has many unique mathematical qualities.

4.2.2.2 Maximum likelihood estimation

Under the assumption that the distribution of $\boldsymbol{Y}$ is within the exponential-dispersion family, its mean and variance could be conveniently expressed with the corresponding canonical and

dispersion parameters

$$\begin{aligned} \mathbb{E}(\boldsymbol{Y}) &= b'(\theta) \\ Var(\boldsymbol{Y}) &= b''(\theta)a(\phi) \end{aligned} \tag{4.2.8}$$

The log-likelihood, as the sum of log-likelihoods of individual observations, is:

$$l(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi_i) \tag{4.2.9}$$

In order to maximize the log-likelihood, one must first take the partial derivate with respect to β. By the chain rule:

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n} \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} \\ &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \frac{1}{g'(\mu_i)} \cdot \boldsymbol{X}_{ij} \\ &= \sum_{i=1}^{n} \frac{y_i - \mu_i}{var(Y_i)}\frac{1}{g'(\mu_i)}\boldsymbol{X}_{ij} \end{aligned} \tag{4.2.10}$$

Therefore, the score equations, also known as likelihood equations, are:

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{var(Y_i)}\frac{1}{g'(\mu_i)}\boldsymbol{X}_{ij} = 0, \qquad j = 1,2, \dots, p \tag{4.2.11}$$

The score equations are functions of β through the fitted means. There exist no closed form solution hence iterative methods must be used to solve these equations. A very commonly used iterative method to solve non-linear equations is the Newton-Raphson method.

The Newton-Raphson method functions as follows:

- Generate some initial solutions $\beta^{(0)}$
- At the neighbourhood of $\beta^{(i)}$, with i being the iteration number, use a second-order multivariate Taylor series expansion to approximate the likelihood function. As shown in Equation (4.2.12), where $\boldsymbol{u}$ is the vector of score equations (first order derivatives) and $\boldsymbol{H}$ is the Hessian matrix of second order derivatives, with $\boldsymbol{H}_{jk} = \frac{\partial^2 l}{\partial \beta_j \partial \beta_k}$.

$$l(\boldsymbol{\beta}) = l(\boldsymbol{\beta}^{(i)}) + \frac{\boldsymbol{u}^{(i)^T}}{1!}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}) + \frac{1}{2!}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)})^T \boldsymbol{H}^{(i)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(i)}) \tag{4.2.12}$$

The term is then maximized with respect to $\boldsymbol{\beta}$ by taking derivative and setting to zero, arriving at:

$$\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)} - [\boldsymbol{H}^{(i)}]^{-1}\boldsymbol{u}^{(i)} \tag{4.2.13}$$

This process is repeated until per-iteration adjustment is smaller than a prespecified tolerance factor.

Another way for estimation is to use the expected value of the hessian matrix. The method is called Fisher-scoring. The expected hessian is named the Fisher information matrix. Its properties are shown in Equation (4.2.14).

$$\begin{aligned} \boldsymbol{\mathcal{I}}_{jk}(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \mathbb{E}\left(-\frac{\partial^2 l_i}{\partial\beta_j \partial\beta_k}\right) = \sum_{i=1}^{n} \mathbb{E}\left(\frac{\partial l_i}{\partial\beta_j} \cdot \frac{\partial l_i}{\partial\beta_k}\right) \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\left\{\frac{Y_i - \mu_i}{var(Y_i)}\right\}^2 \cdot \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 X_{ij}X_{ik}\right] \\ &= \sum_{i=1}^{n} X_{ij}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 \mathbb{E}\left[\left\{\frac{Y_i - \mu_i}{var(Y_i)}\right\}^2\right] X_{ik} \\ \boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) &= \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} \end{aligned} \tag{4.2.14}$$

Fisher scoring is the default method used to estimate parameters in GLMs because the hessian matrix isn't always negative definite, while the Fisher information matrix is always symmetric positive definite. In fisher scoring, the adjustment per iteration uses a negative version of the Fisher information matrix. i.e. $\boldsymbol{\beta}^{(i+1)} = \boldsymbol{\beta}^{(i)} + [\boldsymbol{\mathcal{I}}^{(i)}]^{-1}\boldsymbol{u}^{(i)}$ The inverse Fisher information matrix also coincide with the asymptotic variance-covariance matrix of the linear predictors.

4.3 GLM Case Studies

4.3.1 Gamma Regression in a Quality-Improving Experiment

4.3.1.1 Problem statement and interpretation via linear models

Suppose that a company is investigating the effect of the concentration of a certain chemical and two different mechanical treatments on the UCS of its product, in order to optimize design. A data set with 90 observations was collected, with UCS as the response, CON (Chemical concentration) as a continuous predictor and TRE (Mechanical treatment) as a factor predictor of two levels. A

scatter plot of the data is shown in Figure 4.3.

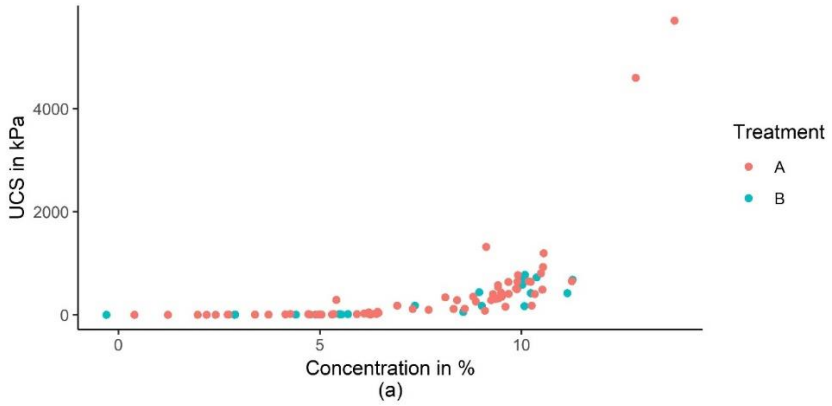


Figure 4.3. Scatter plot UCS vs Chemical concentration

A main-effect only model showed that the mechanical treatment, TRE, in fact isn't a significant predictor of UCS, with a large p-value of 0.45.

Table 4.1. Regression summary (LM identity)

Model: UCS ~ CON + TRE				
AIC = 1430.895		R-squared = 0.3033		
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-642.051	190.0251	-3.37877	0.001091
CON	141.9753	23.30625	6.091728	2.96E-08
TRE	-121.461	161.2538	-0.75323	0.453346

A subsequent model with only CON as predictor had smaller AIC of 1429, with R^2 of 0.30 The regression line and residual plot of this linear regression model are shown in Figure 4.4. It could be seen that the relationship between the response and the predictor is clearly not linear, and the residuals do not center around zero, nor is the variance constant across the mean.

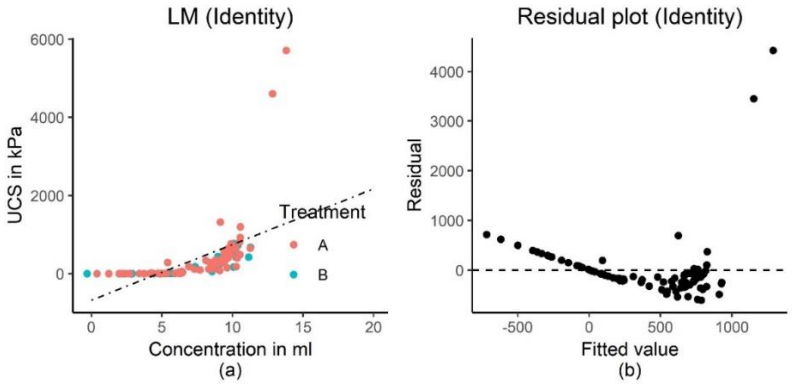


Figure 4.4. Regression results of LM (identity) (a) Regression line (b) Residual plot

Yet another model could be fitted by log-transforming both the response and the continuous predictor. By the log-transformed model had a R^2 of 0.92. As shown in Figure 4.5, the residuals plot showed significant improvement, with the mean generally centered around 0, but the constant variance assumption is not impeccable but holds approximately.

Table 4.2. Regression summary (LM log-transform)

Model: log(UCS) ~ CON				
AIC = 199.86		R-squared = 0.91		
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.67617	0.198954	-8.42489	6.25E-13
CON	0.788545	0.025094	31.42377	1.30E-49

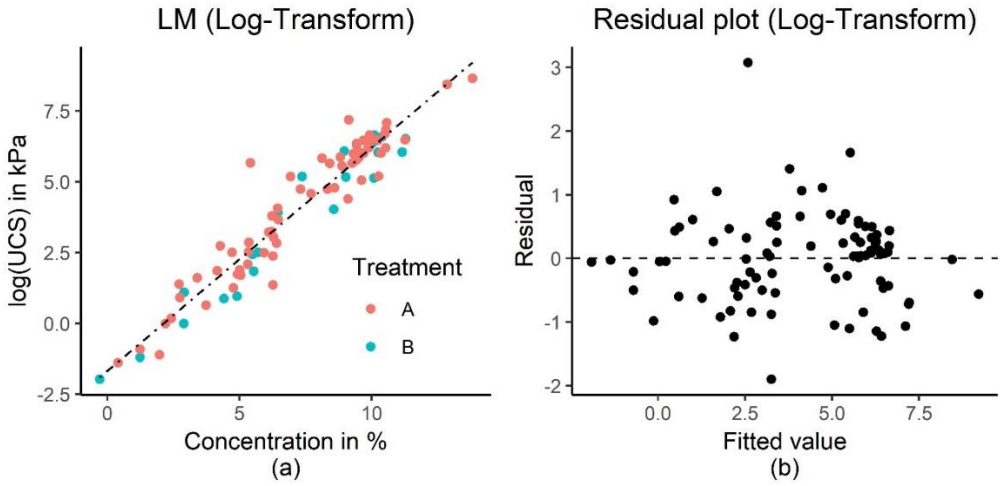


Figure 4.5. Regression results of LM (log-transform) (a) Regression line (b) Residual plot

4.3.1.2 A GLM approach

Since the data is skewed continuous and positive, it is reasonable to consider using a Gamma GLM. The gamma pdf has form as follows:

$$f_Y(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} exp\{-\beta y\} \qquad (4.3.1)$$

A reparameterization of $a = \frac{\alpha}{\beta}$ and $b = \alpha$ allows it to be rewritten as a member of the exponential dispersion family:

$$
\begin{aligned}
f_Y(y;a,b) &= \frac{b^b}{a^b\Gamma(b)} y^{b-1} exp\{-\frac{b\cdot y}{a}\} \\
&= exp\{b\ log(b) - b\ log(a) - log[\Gamma(b)] + (b-1)log(y) - \frac{b\cdot y}{a}\} \\
&= exp\{\frac{(-\frac{1}{a})\cdot y - log(a)}{\frac{1}{b}} + b\ log(b) - log[\Gamma(b)] + (b-1)log(y)\}
\end{aligned}
\tag{4.3.2}
$$

The mean and variance of the re-parameterized gamma pdf are:

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{Y}) &= b'(\theta) = a \\
var(\boldsymbol{Y}) &= b''(\theta)a(\phi) = \frac{a^2}{b}
\end{aligned}
\tag{4.3.3}
$$

It could be observed that the variance is a quadratic function of the mean, as opposed to the normal pdf who's mean and variance are independent. Therefore, the Gamma GLM could be used to model continuous, positive data whose variance increases with mean, which is applicable for this dataset.

Table 4.3. Regression summary (GLM with inverse and log link functions)

Model: UCS ~ CON Family = Gamma(link = 'inverse')				
AIC = 1108.8				
Null deviance: 324 on 89 degrees of freedom				
Residual deviance: 211.13 on 88 degrees of freedom				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.012474	0.001487	8.38676	7.49E-13
CON	-0.0009	0.000108	-8.29207	1.17E-12
Model: UCS ~ CON Family = Gamma(link = 'log')				
AIC = 973.99				
Null deviance: 324 on 89 degrees of freedom				
Residual deviance: 60.17 on 88 degrees of freedom				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.05598	0.431234	-2.44875	0.016318
CON	0.749549	0.054391	13.78074	1.08E-23

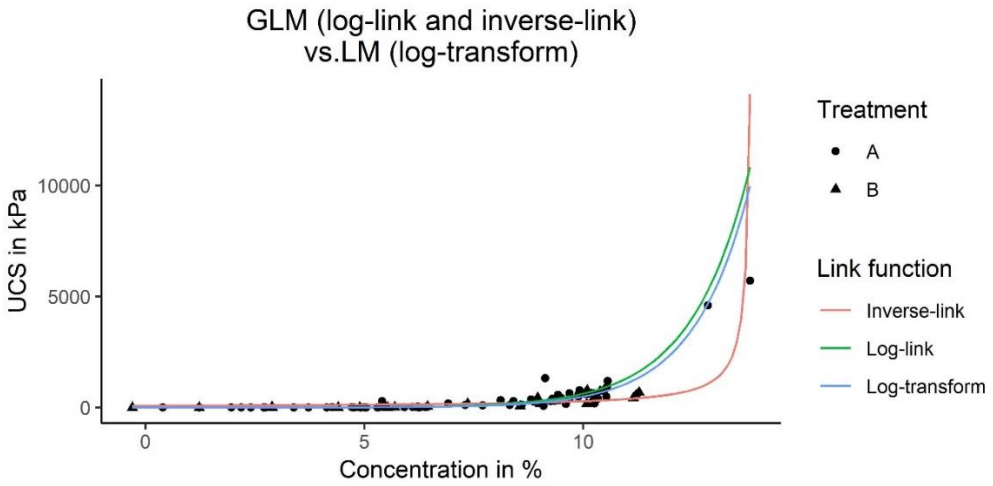


Figure 4.6. Regression results of Gamma GLM compared to LM (log-transform)

Table 4.4. Interpretation of regression models

Regression Model	Interpretation
Gamma GLM with inverse link	$\frac{1}{\mathbb{E}(Capacity)} = 0.012 - 0.0009 \cdot CON$
Gamma GLM with log link	$log\{\mathbb{E}(Capacity)\} = -1.06 + 0.75 \cdot CON$
LM with log-transformed response	$\mathbb{E}\{log(Capacity)\} = -1.68 + 0.79 \cdot CON$

It should be noted that $log\{\mathbb{E}(UCS)\} \neq \mathbb{E}\{log(UCS)\}$ since the expectation is an integral, i.e. $log\left[\int_{y\in Y} y \cdot f_Y(y)dy\right] \neq \int_{y\in Y} log(y) \cdot f_Y(y)dy$. In cases like this, gamma regression is the only way though which one may get direct estimate on the response. Hence the use of gamma regression in treating similar datasets in mining related problems should be promoted.

4.3.2 Logistic Regression in an Occupational Safety Study

4.3.2.1 Problem statement and the analysis of binary responses

The second case study of this chapter is based on safety related data from a certain manufacturing facility. The study involves performing a voluntary safety intervention on a portion of the 151 employees. After the safety intervention employees were monitored by their respective supervisors for 2 months, performance was evaluated by a binary variable, taking value 1 if an employee performed safety-breaching activities within the period, and 0 otherwise. The covariates used in this study are: INT: a 2-level factor representing if an employee undertook safety intervention;

SUP: a 2-level factor representing if an employee is a worker or a supervisor, and EXP, which represents the amount of working experience of the employee measured in months. The aim of the study is to investigate the relationship between an employee's working experience and their safety behaviours, as well as if the designed safety intervention worked as intended. Since the responses are binary, binomial regression could be used to formulate the model.

The probability mass function of Binomial distribution with probability of success π and size m is: $f_Y(y;\pi,m) = \binom{m}{y}\pi^y(1-\pi)^{m-y}$, which could be reparametrized to an exponential-dispersion family form given that y in this case represents not the number of success, but the proportion of success. The reparametrized form is:

$$\begin{aligned} f_Y(y;\pi,m) &= \binom{m}{my}\pi^{my}(1-\pi)^{m-my} \\ &= \exp\{\frac{y\ log\frac{\pi}{1-\pi} + \log(1-\pi)}{\frac{1}{m}} + log\binom{m}{my}\} \end{aligned} \tag{4.3.4}$$

As could be seen from Equation (4.3.4), the canonical parameter of the binomial distribution is $log\frac{\pi}{1-\pi}$. When the canonical link is used in binomial GLMs, the linear predictors conveniently represent the log-odds ratio.

4.3.2.2 Logistic regression analysis

A histogram depicting the working experience distribution of the studied employees are shown in Figure 4.7. As part of the exploratory analysis, the proportion of safety breach are plotted against different groups of employees and quantile values of experience, as shown Figure 4.7.

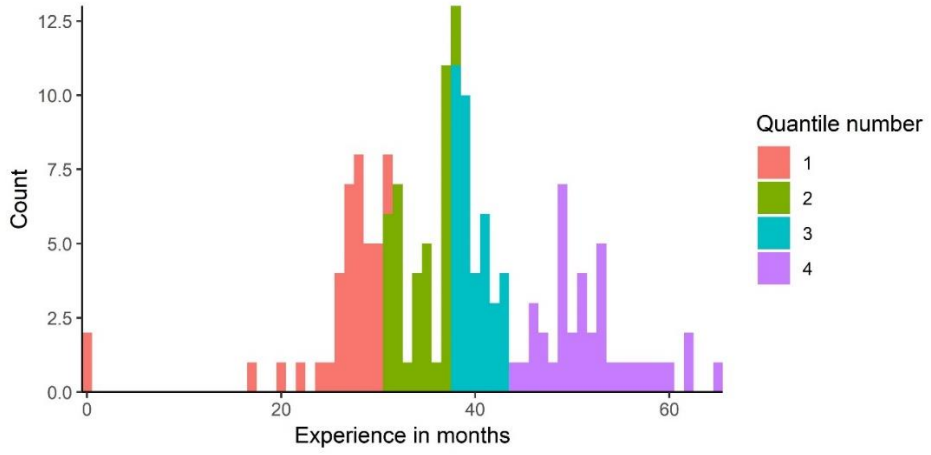


Figure 4.7. Histogram of the safety intervention study data

It could be seen in the plot of observed means that there does indeed appear to be an effect of the safety intervention on the safety behavior of the employees, with the group of employees who participated in the safety intervention generally having a lowered proportion of committing safety breaches. As well as that, the group of employees that are supervisors also tend to breach safety regulations less often.

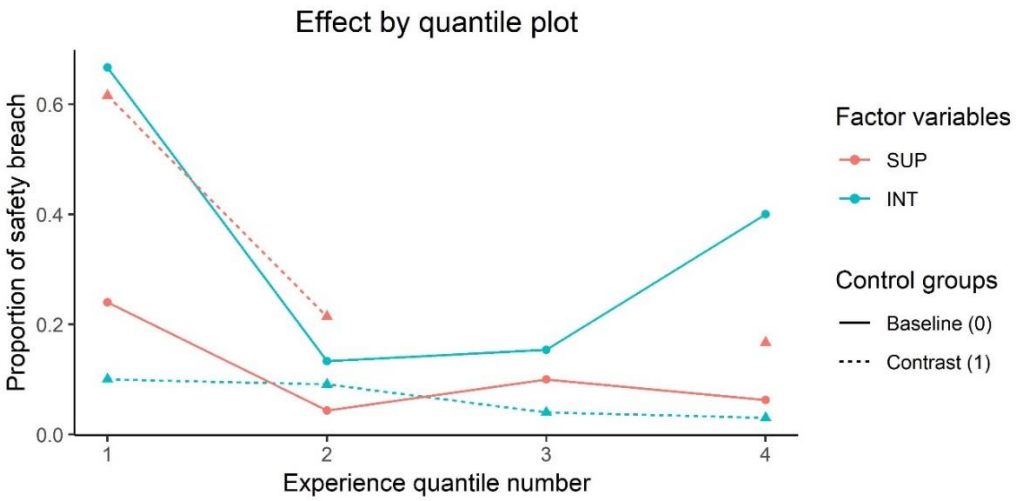


Figure 4.8. Observed mean for each group

A backward model selection was carried out, starting from the most complicated model with three-way interactions, within each iteration the difference in residual deviance between models were used to selection between nested models and eventually the final model was selected to be the model y ~ EXP + SUP + INT + EXP:SUP, at a 10% significant level.

Table 4.5. Analysis of deviance table

Model 1: y ~ EXP + SUP + INT + EXP:SUP + EXP:INT + SUP:INT				
Model 2: y ~ EXP * SUP * INT				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
144	101.3479			
143	100.1706	1	1.177322	0.277902
Model 1: y ~ EXP + SUP + INT + EXP:SUP + SUP:INT				
Model 2: y ~ EXP + SUP + INT + EXP:SUP + EXP:INT + SUP:INT				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
145	101.3532			

144	101.3479	1	0.005285	0.942049
Model 1: y ~ EXP + SUP + INT + EXP:SUP				
Model 2: y ~ EXP + SUP + INT + EXP:SUP + SUP:INT				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
146	102.1354			
145	101.3532	1	0.782225	0.376462
Model 1: y ~ EXP + SUP + INT				
Model 2: y ~ EXP + SUP + INT + EXP:SUP				
Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
147	104.8461			
146	102.1354	1	2.710695	0.099678

Again, due to the responses being binary, the deviance of the saturated model has certain qualities that prevents the utilization of residual deviance to evaluate the model goodness of fit. In this case, the only option was to plot the regression lines against the means of different groups, as shown in Figure 4.9. Unfortunately, however, the optimal model didn't seem to agree very much with the mean of each group, especially the group of supervisors who participated in the safety intervention (in blue) and the group of normal employees (in green) who did not participate. Interestingly, it turned out that there are a few observations in the study that were particularly influential. i.e. a few new employees with zero amount of experience who weren't supervisors. After dropping those observations, a new quantile histogram was plotted in Figure 4.10 and the analysis was rerun with the new optimal model being y ~ EXP + TEN, the model with only two of the main effects. The new diagnostic plot was shown in Figure 4.11, and the trend of the regression curves agreed a lot better with the means of each group.

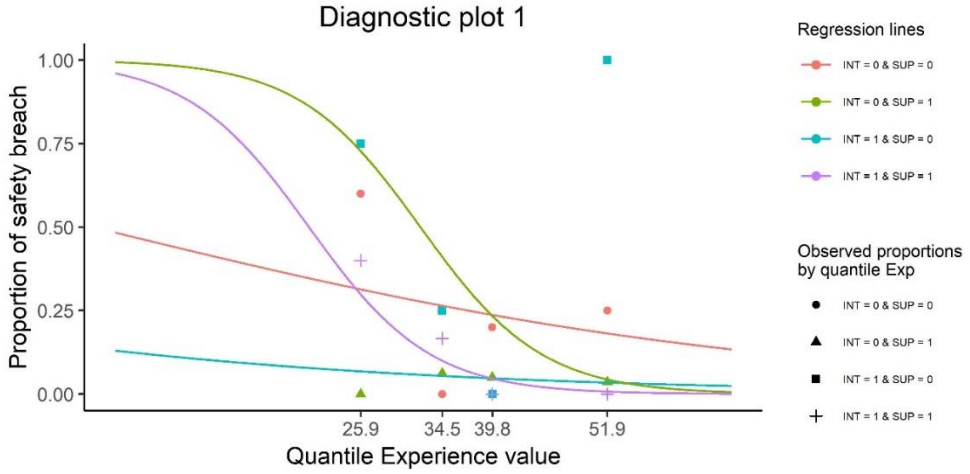


Figure 4.9. Diagnostic plot of the model with full observations

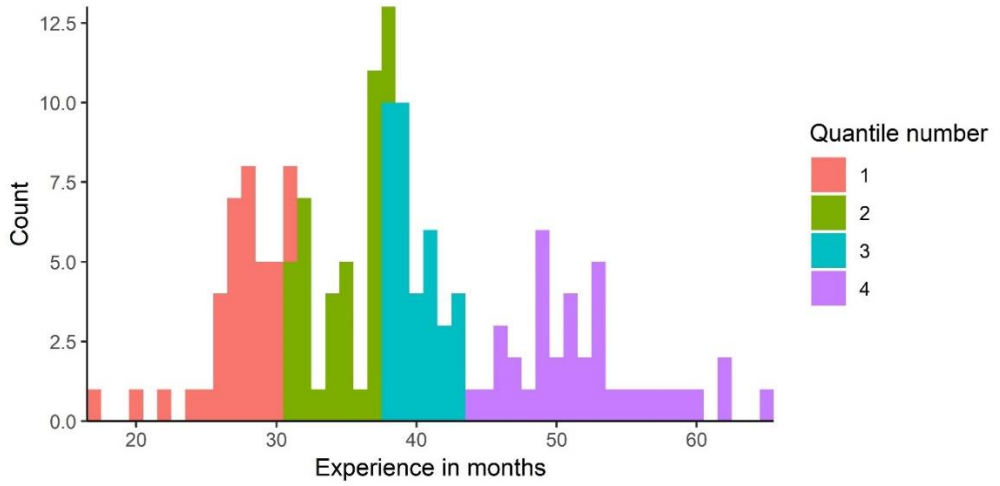


Figure 4.10. Quantile histogram with reduced observations

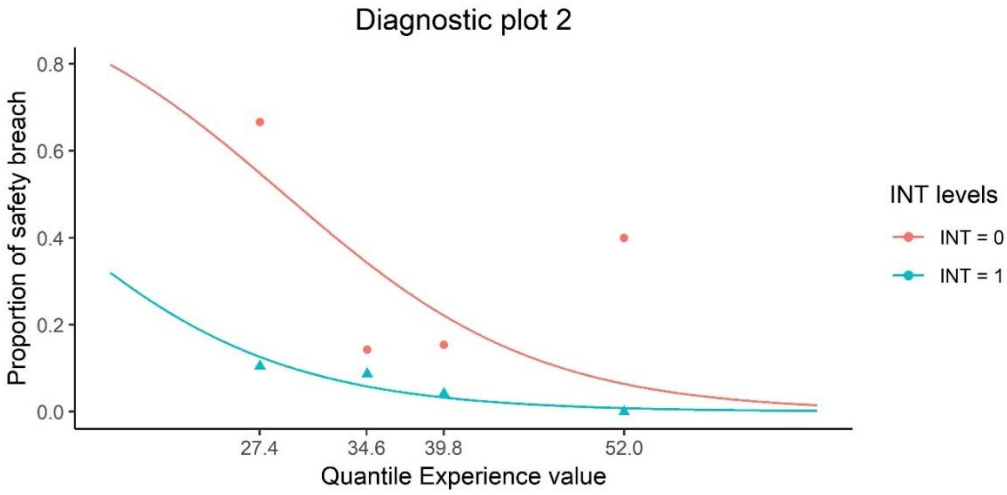


Figure 4.11. Diagnostic plot of the model with reduced observations

The ROC curve of the models is plotted in Figure 4.12. The final optimal model had the most area between itself and the diagonal line and therefore has the most prediction power.

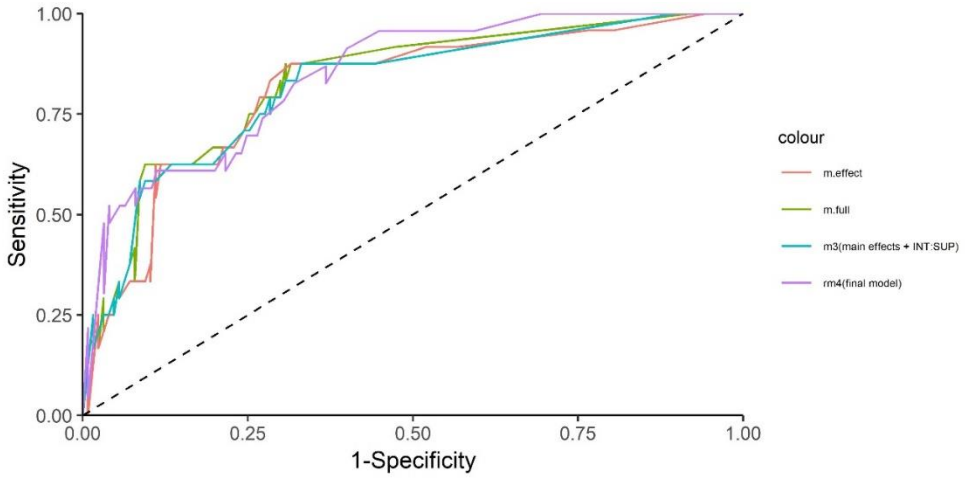


Figure 4.12. ROC plot for model comparison

Table 4.6. Logistic regression output

Call: glm(formula = y ~ LMI + TEN, family = binomial)		
Coefficients:		
(Intercept)	Exp	INT
3.403	-0.117	-2.131
Degrees of Freedom: 147 Total (i.e. Null); 145 Residual		
Null Deviance: 127.9		
Residual Deviance: 94.15	AIC: 100.2	

The regression coefficients are shown in Table 4.6, and the model could be interpreted mathematically as such:

$$log\left\{\frac{\mathbb{E}(Safety\ breach)}{1-\mathbb{E}(Safety\ breach)}\right\} = 3.403 - 0.117 \cdot Experience - 2.131 \cdot INT \qquad (4.3.5)$$

With the left-hand side of the equation representing the log-odds of committing a safety breach. After transforming, it could be interpreted verbally as:

- A month's increase in experience decrease the odds of breaching safety codes by a factor of 1.12
- Employees who took part in the safety intervention has lowered odds to commit a safety breach by a factor of 8.4.

4.4 Chapter Summary

This chapter uses two case to showcase the potential of GLMs being used in the optimization of product quality characteristics and in the analysis of safety engineering data. GLMs are very capable of modelling data that are discrete in nature, including binary, multinomial, and count data. Moreover, for continuous skewed datasets GLMs provide a way to estimate responses directly, which is unachievable via a linear model, even if the variance stabilizing transformation manages to achieve constant variance.

Chapter 5
Analysis of Latent Variables in Occupational Health and Safety in Mining Operations

5.1 Latent Variables in Mining Operations

5.1.1 Preliminaries

Occupational health & safety is a primary concern in the mining industry. Underground mining operations in particular, involve exposing workers to detrimental working environments including airborne respirable dust, excessive amount of potentially deafening noise, narrow openings with considerable heat and humidity as well as the possibility of rock falls and cave-ins. There has been a sizable amount of experience and research works on the technical and socio-technical aspects of mine safety. However, the complex mechanisms that underlie the causal relationships of safety behaviours and occupational injuries are still not fully understood. One way to quantitatively describe these relationships is through the analysis of the unobserved, hidden constructs, or latent variables.

This chapter aims to contribute to the application of quantitative methods such as latent variable analysis and modelling in topics related to mine safety as well as safety science in mining. The chapter is organized as follows. First, a latent variable is defined, followed by a review of the multivariate statistical modelling techniques including the exploratory factor analysis (EFA), the confirmatory factor analysis (CFA) and the structural equation model along with latent variables (SEM). A critical comparison of the three techniques is provided in reference to mine safety. Then, relevant literature in mine safety and safety science that utilizes the techniques mentioned above is discussed. A new approach to cognitive work analysis using (CWA) latent variables analysis is proposed. This approach combines the theoretical advancements in CWA with latent variables analysis to model and measures the effects. Finally, two latent variable models are presented that can be used in cognitive work analysis.

5.1.2 Modelling of Latent Variables

There have been many cases where variables are not directly present in the data, including unmeasured variables, unobserved variables, hidden constructs [39]. In a way, latent variables can be informally defined as variables that are not directly observed in a dataset, but whose existence

can be identified or inferred by the variables that did get directly observed. One possible formal definition of latent variables is the local/conditional independence definition [39]:

$$\mathbb{P}(\mathrm{Y}_1, \mathrm{Y}_2, \ldots, \mathrm{Y}_\mathrm{n}) = \prod_{i=1}^{n} \mathbb{P}\,(\mathrm{Y}_\mathrm{i}|\eta) \tag{5.1.1}$$

where Y_1, Y_2, …, Y_n are observed variables; and η is the vector of latent variables. The local independence definition states that the observed variables become independent if the latent variables that constitute the association between them are held constant. Another intuitive definition is the sample realization definition, which states that a latent variable is a variable for which some subset of a given sample is missing realization and therefore only observable through values of other observed variables.

Latent variables are typically represented as linear combinations of observed variables via factor analysis. Three of the most common techniques are summarized in this chapter, including the EFA, the CFA and the SEM with latent variables.

Exploratory factor analysis is a multivariate statistical technique that studies the underlying relationships between variables by conceptually grouping them, from an examination of appropriate statistics such as covariance or correlation [40]. The procedures of EFA are summarized in Figure 5.1 and detailed in this section.

The orthogonal factor model is the most basic form of factor models. For an observed random vector $\boldsymbol{X} \in \mathbb{R}^{\mathrm{p}\times 1}$ with mean denoted by $\boldsymbol{\mu} \in \mathbb{R}^{\mathrm{p}\times 1}$ and variance-covariance matrix denoted by $\boldsymbol{\Sigma}$, the factor model attempts to explain the total data variance by postulating that it essentially is made up of two parts, i.e., the common variance or communalities from m common factors, $\boldsymbol{F}^{\mathrm{T}} = [F_1, F_2, \cdots, F_m]$ and specific variances from p error terms or specific factors, $\boldsymbol{\varepsilon}^{\mathrm{T}} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]$. The factor model can be written as:

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\varepsilon} \tag{5.1.2}$$

$\boldsymbol{L} \in \mathbb{R}^{\mathrm{p}\times\mathrm{m}}$ is the matrix of factor loadings where the entry l_{ij} represents loading of the i^{th} variable on the j^{th} factor. $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are independent; both have zero expectation with $Cov(\boldsymbol{F}) = \boldsymbol{I}$ and $Cov(\boldsymbol{\varepsilon}) = \boldsymbol{\varphi}$, $\boldsymbol{\varphi}$ being a diagonal matrix. The formulated model then implies a variance-covariance matrix with the form: $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}} + \boldsymbol{\varphi}$. The implied covariance structure has properties as follows:

$$\begin{gathered} Var(\boldsymbol{X}_i) = \sigma_{ii}^2 = h_i^2 + \varphi_i = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \varphi_i \\ Cov(\boldsymbol{X}_i, \boldsymbol{F}_j) = l_{ij} \end{gathered} \tag{5.1.3}$$

where h_i^2 is called the i^{th} commonality, or common variance, which represents the amount of

variance of the i^{th} measured variable explained by all m factors. φ_i is known as the uniqueness, or specific variance of the i^{th} measured variable, representing the residual variance left unexplained. The i, j^{th} entry of the loading matrix $\boldsymbol{L}$ represents the covariance between the i^{th} measured variable and j^{th} factor [41].

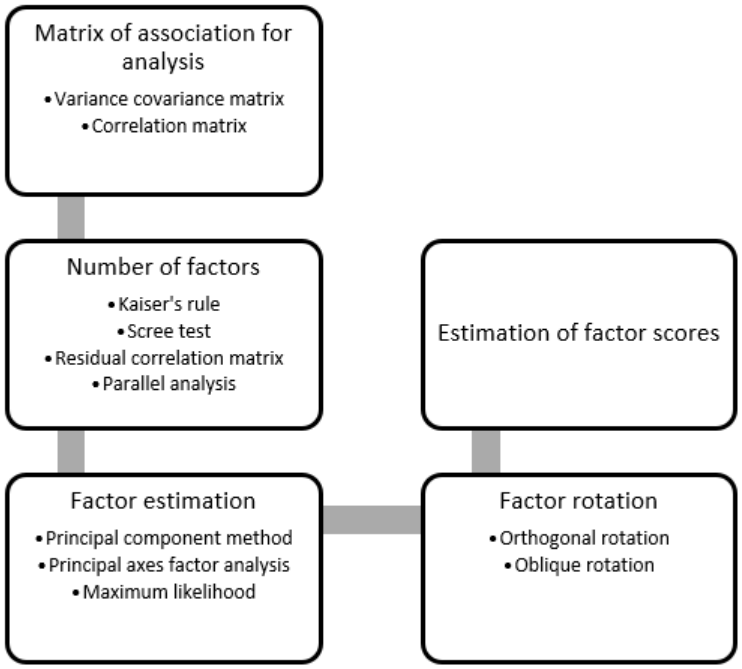


Figure 5.1. Procedure for EFA

The matrix of associations of the dataset could be taken either as the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathrm{p\times p}}$ or the standardized correlation matrix $\boldsymbol{R}_{\mathrm{p\times p}}$. This choice could be made by observing the comparability of the measured variables. If the measured variables have the same unit and are of similar scales, then $\boldsymbol{\Sigma}_{\mathrm{p\times p}}$ could be used; otherwise, $\boldsymbol{R}_{\mathrm{p\times p}}$ is generally preferred as the standardized values are much more comfortable for interpretation.

5.1.3 Number of factors and methods of estimation

The most commonly used approaches for determining the number of factors are shown as follows: (1). Kaiser's rule states that a noteworthy factor should have an eigenvalue of greater than 1. As in factor analysis, the eigenvalues of all possible factors sum to the number of measured variables (p), therefore the importance of a factor can be represented by its eigenvalue divided by p. Also, if the product of the division is greater than 1, then that particular factor could be considered as

significant.

(2). A screen test is a bar plot that shows the percentage variance explained per factor in descending order. Important factors are those that precede the "elbow" of the plot, where factors exhibit sudden drops of significance.

(3). Residual correlation matrix could be found by subtracting from the original correlation matrix with that reconstructed with selected factors. Scenarios with fewer factors with sufficiently small residual correlation entries are preferred.

Most statistical analysis programs use the principal component method as the default method for factor estimation. Also, commonly used methods include the principal axes factor analysis and maximum likelihood estimation.

(1). Principal component method

When the specific variances are set to zero, the implied covariance matrix of the common factor model is similar to the spectral decomposition of the variance-covariance matrix.

$$\begin{aligned}\boldsymbol{\Sigma}_{p\times p} &= \boldsymbol{L}_{p\times m}\boldsymbol{L}^{T}_{m\times p} + \mathbb{0}_{p\times p} \\ &= \begin{bmatrix}\sqrt{\lambda_1}v_1 & \sqrt{\lambda_2}v_2 & \dots & \sqrt{\lambda_p}v_p\end{bmatrix}\begin{bmatrix}\sqrt{\lambda_1}v_1^T \\ \sqrt{\lambda_2}v_2^T \\ \vdots \\ \sqrt{\lambda_p}v_p^T\end{bmatrix} \\ &= \sum_{i=1}^{p}\lambda_i\, v_i v_i^T = \boldsymbol{\Lambda V \Lambda}^{\mathrm{T}}\end{aligned} \tag{5.1.4}$$

The final form is obtained by dropping the last p-m terms:

$$\begin{aligned}\boldsymbol{\Sigma}_{p\times p} &= \boldsymbol{L}_{p\times m}\boldsymbol{L}^{T}_{m\times p} + \boldsymbol{\varphi}_{p\times p} \\ &= \begin{bmatrix}\sqrt{\lambda_1}v_1 & \sqrt{\lambda_2}v_2 & \dots & \sqrt{\lambda_m}v_m\end{bmatrix}\begin{bmatrix}\sqrt{\lambda_1}v_1^T \\ \sqrt{\lambda_2}v_2^T \\ \vdots \\ \sqrt{\lambda_m}v_m^T\end{bmatrix} \\ &+ \begin{bmatrix}\varphi_1 & 0 & \dots & 0 \\ 0 & \varphi_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \varphi_m\end{bmatrix}\end{aligned} \tag{5.1.5}$$

(2). Principal axes factor analysis

Principal axes factor analysis starts from a principal component analysis (PCA) but with the diagonal entries of the analyzed correlation matrix replaced with respective communalities of each

variable. The altered correlation matrix will then have PCA performed on it iteratively until it the communalities converge, or a certain maximum iteration has been reached.

(3). Maximum likelihood

A maximum likelihood estimate could be found by maximizing a likelihood function as shown in Equation (5.1.6).

$$\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-\frac{np}{2}}|\boldsymbol{\Sigma}|^{-\frac{n}{2}}e^{-\frac{1}{2}tr[\boldsymbol{\Sigma}^{-1}(\sum_{j=1}^{n}(x_j-\overline{x})(x_j-\overline{x})^T + n(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T)]} \quad (5.1.6)$$

5.1.4 Factor Rotation

Directly estimated factor loadings might not be interpretable. Consequently, rotations are performed, which redistributes the location of variance within the loadings facilitating the interpretation. Rotation is often performed with the aim of reaching a simple structure, which, in a column perspective, have an approximately equal number of observed variables represented by each factor; or in a row perspective, have most observed variables primarily correlated with only one factor. Factors can be rotated either orthogonally or obliquely. Oblique rotation occurs when the transformation matrix is non-orthogonal and is often performed in order to render factors correlated to account for broad factor generalization and overlapping. In an oblique model, the structure coefficient matrix S is found as the product of factor loadings and inter-factor correlation $\boldsymbol{R}$, with $\boldsymbol{S}_{p\times m} = \boldsymbol{L}_{p\times m}\boldsymbol{R}_{m\times m}$. When $\boldsymbol{R}_{m\times m} \neq \boldsymbol{I}_{m\times m}$, both S and L matrices need to be examined during interpretation. Moreover, higher-order factors could be extracted from $\boldsymbol{R}$, and they need to be interpreted as well [42].

5.1.5 Factor Score Estimation

Factor score estimation is usually carried out using the weighted least squares and regression methods.

(1). Weighted least squares

Specific factors are treated as residuals, the residuals sum of squares is minimized weighted by their respective reciprocal variances. The formulation of the problem and the solution while taking the estimated values as true values are shown as follows:

$$\sum_{i=1}^{p}\frac{\varepsilon_i^2}{\varphi_i} = \boldsymbol{\varepsilon}^T\boldsymbol{\varphi}^{-1}\boldsymbol{\varepsilon} = (\boldsymbol{X}-\boldsymbol{\mu}-\boldsymbol{L}\boldsymbol{f})^T\boldsymbol{\varphi}^{-1}(\boldsymbol{X}-\boldsymbol{\mu}-\boldsymbol{L}\boldsymbol{f}) \quad (5.1.7)$$

$$\hat{\boldsymbol{f}}_j = \left(\hat{\boldsymbol{L}}^T \hat{\boldsymbol{\varphi}}^{-1} \hat{\boldsymbol{L}}\right)^{-1} \hat{\boldsymbol{L}}^T \hat{\boldsymbol{\varphi}}^{-1} (x_j - \bar{x})$$

(2). Regression method

For a factor model specified in Equation (5.1.2), the joint distribution of $(\boldsymbol{X} - \boldsymbol{\mu})$ and $\boldsymbol{F}$ is $\mathcal{N}_{p+m}(0, \boldsymbol{\Sigma}^*)$, where $\boldsymbol{\Sigma}^*_{(p+m)\times(p+m)} = \begin{bmatrix} \boldsymbol{\Sigma}_{p\times p} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\varphi} & \boldsymbol{L}_{p\times m} \\ \boldsymbol{L}^T_{m\times p} & \boldsymbol{I}_{m\times m} \end{bmatrix}$. From the joint distribution, it is possible to find the conditional expectation [43]:

$$\mathbb{E}(\boldsymbol{F}|\boldsymbol{x}) = \boldsymbol{L}^T(\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\varphi})^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \tag{5.1.8}$$

Bibliography The term $\boldsymbol{L}^T(\boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\varphi})^{-1}$ in Equation (5.1.8) is analogous to a regression coefficient, consequently given vector of observation x_j and taking the estimated values as actual values, the estimated factor score can be found as $\hat{\boldsymbol{f}}_j = \hat{\boldsymbol{L}}^T\left(\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\varphi}}\right)^{-1}(x_j - \bar{x})$.

5.1.6 Confirmatory Factor Analysis (CFA)

EFA is a data-driven technique, with the latent variables being *a posteriori* whereas in confirmatory factor analysis the factors and their corresponding loading matrix are determined before analysis, therefore a prior method [39]. A sufficient level of an empirical or theoretical foundation is needed for model specification and evaluation in CFA. Consequently, CFA is often applied when there has already been a level of development in research, where the tentative underlying structure has been identified with analytical techniques such as EFA [44]. EFA and CFA are similar in the sense that they are both built on the common factor model. However, certain differences between the two methods exist and will be summarized later in this chapter.

Test statistics for evaluating the fitness of a CFA model is similar to that of structural equation models and are detailed in the next sections of this chapter.

5.1.7 Structural Equation Models (SEM) with Latent Variables

5.1.7.1 General structural equation model

The general structural equation model can be considered as some combination of multiple regression, which concerns the relationships between observed variables with errors being latent variables, and factor analysis, which finds the link between latent and observed variables but with limited emphasis on the relationships between latent variables. It inherently consists of a measurement model that specifies the relationship between observed and latent variables and a latent variable model that delineates links among latent variables [43]. The general decision sequences for SEM is shown in Figure 5.2.

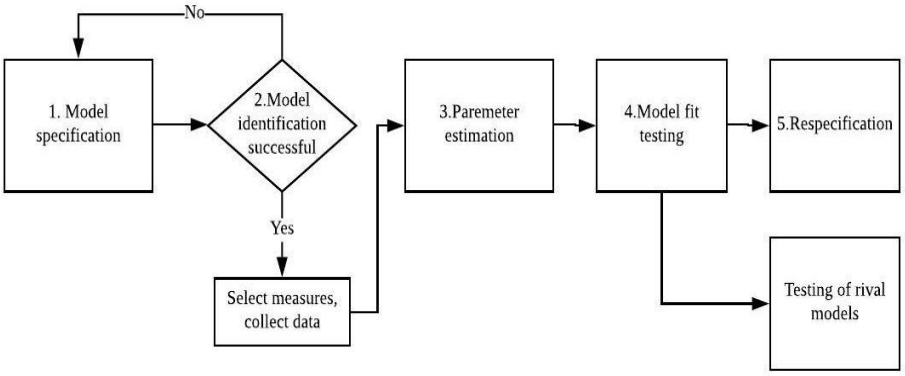


Figure 5.2. Decision sequences for SEM

SEM is a confirmatory technique and not well suited for exploratory identification of relationships in research. Like CFA, the prerequisite for the use of SEM is a prior specification of a model, often from previous research or theory [45]. Both the measurement and the latent variable model need to be specified.

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{B\eta} + \boldsymbol{\Gamma\xi} + \boldsymbol{\zeta} \\ \boldsymbol{y} &= \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon} \\ \boldsymbol{x} &= \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \tag{5.1.9}$$

Equation (5.1.9) shows the SEM measurement model, followed by the latent variable models for endogenous and exogenous latent variables. Assuming $\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\epsilon}, \boldsymbol{\delta}$ (all latent terms) have zero expectation, $(\boldsymbol{I} - \boldsymbol{B})$ is non-singular, $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ uncorrelated, $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta}, \boldsymbol{\xi}$ and $\boldsymbol{\delta}$ are uncorrelated as well as $\boldsymbol{\delta}$ and $\boldsymbol{\xi}$, $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ are uncorrelated. Table 5.1 shows the dimensions and definitions of the parameters in the general structural model [43].

Table 5.1. SEM parameters

Parameter	Dimension	Definition
$\boldsymbol{\eta}$	$m \times 1$	Latent endogenous variables
$\boldsymbol{\xi}$	$n \times 1$	Latent exogenous variables
$\boldsymbol{\zeta}$	$m \times 1$	Error term in the latent variable model
$\boldsymbol{B}$	$m \times m$	Coefficient model for latent endogenous variables
$\boldsymbol{\Gamma}$	$m \times n$	Coefficient model for latent exogenous variables
$\boldsymbol{y}$	$p \times 1$	Observed variables that indicates $\boldsymbol{\eta}$
$\boldsymbol{x}$	$q \times 1$	Observed variables that indicates $\boldsymbol{\xi}$
$\boldsymbol{\epsilon}$	$p \times 1$	Measurement errors for $\boldsymbol{y}$
$\boldsymbol{\delta}$	$q \times 1$	Measurement errors for $\boldsymbol{x}$
$\boldsymbol{\Lambda_y}$	$p \times m$	Coefficients relating $\boldsymbol{y}$ to $\boldsymbol{\eta}$
$\boldsymbol{\Lambda_x}$	$q \times n$	Coefficients relating $\boldsymbol{x}$ to $\boldsymbol{\xi}$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{I}-\boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\phi}\boldsymbol{\Gamma}^T + \boldsymbol{\psi})(\boldsymbol{I}-\boldsymbol{B})^{-T} & (\boldsymbol{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\phi} \\ \boldsymbol{\phi}\boldsymbol{\Gamma}^T(\boldsymbol{I}-\boldsymbol{B})^{-T} & \boldsymbol{\phi} \end{bmatrix} \quad (5.1.10)$$

Equation (5.1.10) is the covariance structure hypothesis for the general structural model, where $\boldsymbol{\Sigma}$ represents the population variance-covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the model implied variance-covariance matrix with θ representing the vector of free model parameters. $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ represent the variance-covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$. One key requirement for model identification is that the number of unknown parameters in $\boldsymbol{\theta}$ must be smaller or equal to the number of nonredundant terms in the implied variance-covariance matrix, known as the t-Rule [43]. The ideal situation in model identification is to have more equations than unknowns, i.e., an overidentified model. An overidentified model has multiple possible solutions, but the one with the best fit to the data could be selected. In contrast, an under-identified model has no unique solution while a just-identified

model has exactly one solution but with large measurement and sampling error. Overidentification could be achieved by constraining some of the parameters to predetermined values [45].

5.1.7.2 Parameters estimation and model fit

The general structural equation parameters are estimated by finding the closest estimate of the implied variance-covariance matrix ($\widehat{\boldsymbol{\Sigma}}$) to the estimated population variance-covariance matrix (*S*), with respect to some minimization criterion $F[\boldsymbol{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})]$, which is a function of $\boldsymbol{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The most commonly used criterions include maximum likelihood (ML), unweighted least squares (ULS) and generalized least squares (GLS).

$$\begin{aligned} F_{ML} &= log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + tr\{\boldsymbol{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\} - log|\boldsymbol{S}| - (p+q) \\ F_{ULS} &= \frac{1}{2} tr\{[\boldsymbol{S} - \boldsymbol{\Sigma}(\boldsymbol{\theta})]^2\} \\ F_{GLS} &= \frac{1}{2} tr\{[\boldsymbol{I} - \boldsymbol{\Sigma}(\boldsymbol{\theta})\boldsymbol{S}^{-1}]^2\} \end{aligned} \tag{5.1.11}$$

Among the three estimators given in Equation (5.1.11) the maximum likelihood estimator is considered to be most consistent when the sample size is large, and the observed variables can be considered as jointly normal. Conversely, when the sample size is large but multivariate normality is in question, the generalized least squares estimator is the most reasonable choice [45].

Evaluating the fitness of a structural equation model could be potentially tricky as any mis-specified model may be obtained to fit the data by adding free parameters. Moreover, a perfect fit could occur when the model degrees of freedom equal to zero (i.e., model is just-identified, and all possible free parameters are estimated), but there is little scientific value in such a model [46]. Besides, some degree of misfit between the observed and implied variance-covariance matrix is expected due to sampling fluctuations; nevertheless, the misfit might just as likely originate from model misspecification. Hence, it is reasonable to assume that a model that fits the sample variance-covariance matrix is just one among many potentially causally different models that are consistent with the data; therefore, it is necessary to evaluate the equivalent or close-to-equivalent models and differentiate between them [47].

In general, three different types of fit indices exist for SEM:

(1). Absolute fit indices evaluate the model's ability to reproduce a variance-covariance matrix close to that of the observed data. The most commonly used absolute fit index being the χ^2 index, with $\chi^2_{model} = (N-1)F_{ML}$ where N is the sample size and F_{ML} is the type of estimator used (could also be F_{GLS} or F_{ULS}). The model fits the data perfectly when $\chi^2_{model} = 0$.

(2). Comparative fit indices are often used to compare the level of fit between that of a theoretically

derived model relative to some baseline model, which is often called a null model that specifies no causal relationships between the variables. One of the comparative fit indices is the normed fit index (NFI), with $NFI = \frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$ and an NFI value of greater than 0.9 is considered to be a good fit [45].

(3). Parsimonious fit indices assess the trade-off between model fit and degrees of freedom. For instance, the parsimonious normed fit index is defined as $PNFI = \frac{df_{model}}{df_{null}} \times \frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$.

While reporting fit indices for a structural equation model, there is no need for researchers to report the values of all fit indices, practices such as only reporting indices that indicate good fits should also be avoided. Notwithstanding, reporting different kinds of test statistics is advised as they evaluate various aspects of model fit [48].

5.1.7.3 Sample size considerations and comparisons among modelling techniques

Structural equation modeling is a large sample technique that requires at least 200 samples for a model of moderate complexity, or 10 samples per estimated parameter [45]. While studying underlying structures in the data, it is important that the researcher can identify the merits and demerits of various methods.

I. While EFA more commonly analyzes the correlation matrix of the data, CFA can be used to study both the correlation and the variance-covariance matrix of the data. The resulting output from CFA could include an unstandardized solution, a standardized solution, and a completely standardized solution.

II. CFA models are considered to be more parsimonious than EFA models. In EFA all observed variables are free to load/covary with all factors after which the factors are rotated in row/column perspective, whereas in CFA simple structure is achieved by specifying linking of observed variables to factors before analysis, eliminating the need for factor rotation. There are often fewer parameters that need to be estimated in CFA than in EFA.

III. Errors (specific variances) can be allowed to covary in CFA and SEM models, which violates basic assumptions of standard linear regression models. Covariance between errors can be justified as additional covariance in observed variables due to assessment methods, which is reasonably common in measurement models based on surveys and questionnaires.

IV. CFA and SEM allow for the direct comparisons between rival near-equivalent models via model fit, which facilitates theory testing.

V. While traditional regression analysis assumes perfect measurement, SEM explicitly accounts for measurement errors and thereby reducing regression dilution bias, which is more desirable for questionnaire-based datasets. The multiple indicators used by SEM correct for unreliability and provide more accurate estimations of parameters.

VI. SEM allows for the study of indirect, mediational effects in statistical models, where a single variable can be both the “cause”, and the “effect” [49].

5.2 Review of Latent Variable Analysis in Mine and Safety Sciences

Cooper and Phillips studied the relationship between safety climate and safety behavior. Surveys were conducted before and after a safety intervention, which improved the safety behavior of the employees [50]. A 50-item questionnaire was used to measure 7 variables that act as the observed variables for EFA, the factors were estimated with PCA and varimax rotation (orthogonal) that was performed for two prominent factors to occur. Based on the EFA results, Cooper and Philips argued that workers could very well discriminate between factors that directly related to safe operations and those that do so in an indirect manner (two factors formed in EFA) [51]. Moreover, as the structure coefficients remain largely the same before and after safety intervention, the test results suggest that the unobserved structures of the safety climate measure are reliable and consistent with the relevant previous literature [51]. The research also proved that the changes in safety behavior and that in the safety climate do not necessarily reflect on one another. Zhang analyzed the causality between coal miners’ errors and life events using SEM and found an influential effect value of 0.7945 [52]. Paul used SEM to study the role of personal factors on work injury in underground mines. The study has found rebelliousness, negative affectivity and job boredom as three key personal factors increasing work injuries [53].

Seo et al. investigated constructing a reliable factor structure for safety climate measures in order to overcome the limitations of traditional safety measures [54]. Over six hundred valid samples were collected from workers in the grain industry, after which EFA and CFA were performed on the data. Eventually, a good fit was achieved with the finalized CFA model based on multiple test-statistics. Having developed a model that is consistent with the data, Seo et al. concluded that it is important to consider the influence of management commitment and supervisor support on other variables studied, as they loaded onto observed variables meant to measure other factors. Additionally, their paper managed to develop a reliable factor structure for safety climate measure,

and therefore the same construct might be used to measure safety climate of workers in the same industry. The developed safety climate measures might also give insight into other industries, due to the inherent nature of the safety climate itself [54].

Liu and Li analyzed the latent structures of Firm Safety Management Capability (FSMC) based on data collected from coal mines in northern China [55]. The model they proposed was relatively complex and involved 20 latent variables, including five latent endogenous variables and 15 latent exogenous variables. The latent variables are factors of 76 observed variables, which were based on 999 valid questionnaire surveys answered by miners. Latent variables analyzed in the study by Liu and Li concerns factors that pertain to five main groups related to FSMC, namely relevant safety aspects of the workers, the teams, the firm, and the environment [55].

Their proposed model was based on complex a prior hypothesis on interactions between workers, teams, the firm and working environments. The model proposes some of the worker attributes as endogenous latent variables that are influenced by exogenous latent variables concerning teams, the firm, and working environments. Among the endogenous latent variables, a worker's knowledge and skills have a directional effect on his or her working habits while working habits were also set to covary with responsibility and psychological qualities. Both psychological qualities and working responsibility have a covariational effect on workers knowledge and skill. There also exist various constraints on the directional effects between the endogenous variables and exogenous variables so that the exogenous variables only have an impact on those endogenous variables that are backed up by theories. The model had adequate fitness based on multiple test statistics and theories based on FSMC was found to match the sample data.

5.3 Analysis of Latent Variables in Cognitive Work Analysis

In recent years, researchers have been trying to comprehend the hidden structures, including environmental, organizational and socio-technical factors, that potentially lead to accidents and fatalities in mining complexes [56-58]. In this chapter, it has been proposed that latent variable analysis methods could be used in combination with cognitive work analysis (CWA) in order to better understand the complex and dynamic relationship among the human, environmental and technological factors in sociotechnical systems.

Cognitive work analysis aims to analyze all vital elements of human-work interaction via the application of concepts from various disciplines including engineering, cognitive science, social

science, and psychology. As mining is often deemed as a dynamic, hazardous, automated system that is full of uncertainties, coupled and mediating subsystems and potential disturbances, which agrees with the definition of a complex sociotechnical system by Vicente, it is ideal for the application of cognitive work analysis [59]. It has been stated that understanding the constraints and capabilities of personal, social, organizational, technological elements of a system are generally helpful for finding means to reduce human error factors, reduce the frequency of occurrence of safety-related incidents and increase the overall system performance. CWA typically involves five phases of analyses, i.e., work domain analysis, control task analysis, strategies analysis, social organizational analysis and worker competencies analysis. Demir et al. proposed 11 factors to quantify the overall cognitive quality of a mining operation [57]. Structural equation modeling could be used to analyze the interactions and mediating effects among the factors, to facilitate researchers' understandings of human-work interactions in mining operations. In this chapter, a similar approach is proposed to model the factors in the five levels of cognitive work analysis as latent variables. A list of variables and their descriptions are shown in Table 5.2. The list is partly based on the proposed factors by Demir et al., with several safety-related variables added [57]. Each observed variable in the list could be one singleton as well as several closely related observed variables. Variables generally should have ratings from 0 to 5. Variables related to the work domain should be rated by relevant professionals evaluating the mine, while other variables could be obtained from questionnaires.

As some of the observed variables might be moderately correlated and it is not immediately clear how some of them interact with each other, it might be more helpful to extract factors from them via EFA instead of assigning factors to them *a priori*. The extracted factors, after interpretation, could be fitted to many probable rival structural equation models to compare relative fitness.

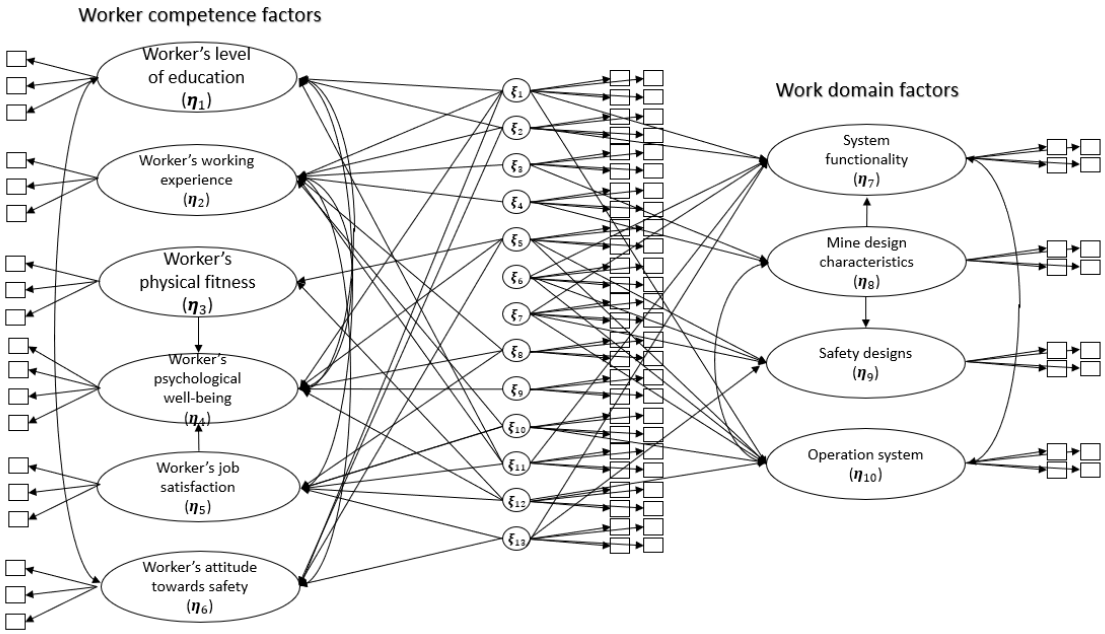


Figure 5.3. SEM path diagram of cognitive work factors

Table 5.2. List of latent and observed variables in cognitive work analysis

Levels of factors of CWA	Latent variables	Notation	Description
Work domain factors	System functionality	η_7	Having a clear idea of the purpose of each subsystem of the mining operation
	Mine design characteristics	η_8	Ratings for the technical aspects of the mine design. Including but not limited to ventilation, rock mechanics, mineral processing, mine planning, etc.
	Safety designs	η_9	The mine has safety structures for emergencies
	Operation system	η_{10}	Design rating for ventilation system of the mine
Control task factors	Task guideline	ξ_1	Having a clear guideline or standard operating procedures for a particular task
	Performance criteria	ξ_2	Having a clear method of evaluation for the performance of a job
	Equipment compatibility	ξ_3	The compatibility and efficiency of current equipment for completing the desired task
	Equipment availability	ξ_4	If the equipment is readily available and easily accessible for utilization
Strategy factors	Risk potential	ξ_5	If the tasks to be performed are subject to potential hazardous outcomes
	Preparation for uncertainties	ξ_6	Level of preparedness for unintended events, precautions taken
	Strategical planning	ξ_7	Alternative plans or methods in the case of disturbances or emergencies
Social and organizational factors	Supervisor support	ξ_8	The supervisor's attitude towards safety, holdings of safety meetings
	Time management	ξ_9	Deadlines set for tasks to ensure completion in time
	Supervisor communication	ξ_{10}	Effectiveness of communication between supervisor and supervisees
	Roles and responsibilities	ξ_{11}	Clearly defined roles and responsibilities for workers
	Co-worker support	ξ_{12}	Level of mutual aid between workers
	Aspects of safety culture	ξ_{13}	The presence of safety department, management of safety-related issues
Worker competency factors	Level of education	η_1	The highest education level of a worker
	Working experience	η_2	Number of years of experience working in a related position
	Physical fitness	η_3	Level of physical fitness and health of the worker, presence of past injuries
	Aspects of psychological well-being	η_4	Worker's ability to handle stress and general state of mental health
	Job satisfaction	η_5	How satisfied the worker is with job position, salary, etc.
	Attitude towards safety	η_6	Worker's willingness to follow safety-related regulations

One probable model depicting interactions between the five factors studied in cognitive work

analysis is shown in Figure 5.3. The model hypothesis is *a priori* and proposes that the factors in the work domain and worker's competence domain are endogenous factors while factors in the other three layers of cognitive work analysis generally function as exogenous factors. The reason being factors related to the qualities of the mine employees and technical aspects of the mine should cause the other latent variables, for instance, the social and organizational factors should depend on the competencies of the workers, and to some extent the technical aspects of the mine design.

The proposed SEM model effectively models mediating effects among variables, which is superior to standard regression models. One example is that although the exogenous variable named "risk potential (ξ_5)" does not directly depend on the endogenous variable "Worker's level of education (η_1)", it receives a mediating effect as η_1 is correlated with "Worker's attitude towards safety (η_6)", which has a direct regression effect on ξ_5, as it is assumed that better-educated employees generally focuses more on safety-related issues. The explicit modeling of many instances of these indirect effects among variables ought to make the model more theoretically sound. Each factor, no matter exogenous or endogenous should be measured by a few observed variables that could be gathered from questionnaire responses. In path analysis notations, an ellipse characterizes an endogenous variable whereas a circle represents an exogenous variable. Squares are observed variables and arrows, and double-sided arrows represent directional effect and covariance respectively.

Studying the interactions among the cognitive work factors could provide new insights into the human-machine-environment interactions in complex mining operations and facilitate the utilization of cognitive work analysis as well. However, questionnaire design, data collection, and modeling of such complex causal relationships could prove challenging. It is often more convenient to start from relatively simple questionnaires with EFA and CFA models. Figure 5.4 is a proposed CFA model on the Social-Organizational factors, which is a subset of the cognitive work analysis factors.

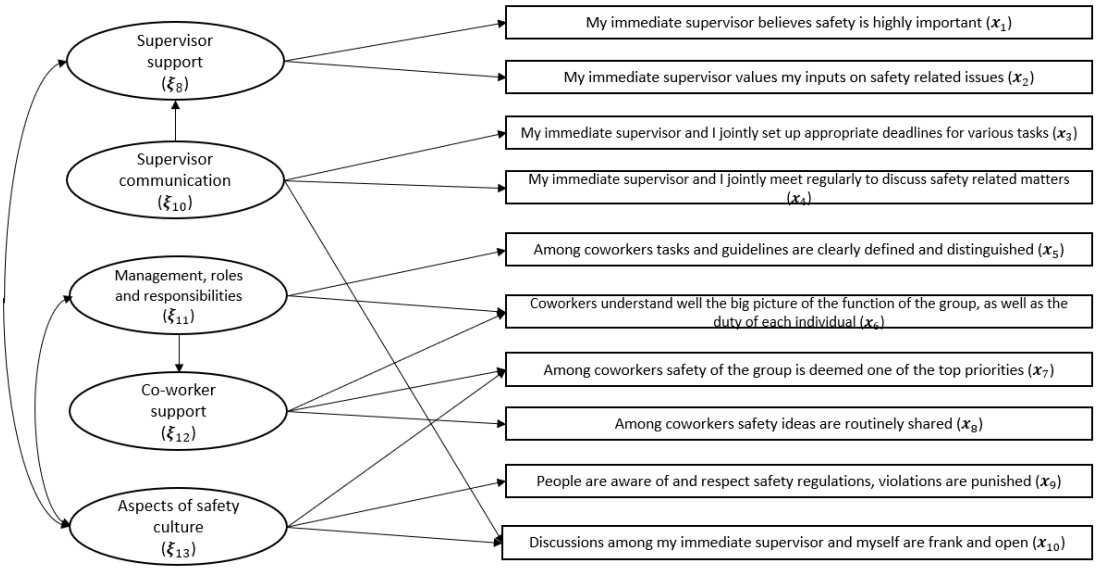


Figure 5.4. CFA path diagram of socio-organizational factors

The proposed model contains only exogenous latent variables and observed variables. The observed variables, i.e., questionnaire items, are designed to measure a latent variable, and a certain degree of overlapping is allowed. While only ten observed variables are shown in Figure 5.4, often a lot more is needed to represent a latent variable accurately. Depending on the fitness of the model, latent variables could be removed or even dropped, as despite being based on theory, the proposed model might not be entirely consistent with all data. Adjacent CFA models with different proposed relationships among the exogenous latent variables could also be tested to improve people's understanding of the socio-organizational factors. If decent fitness could be achieved, it could be implied that the proposed model is consistent with data collected from a certain mining facility. Therefore, an inference could be made on certain variables so that the mine manager could identify which observed variables to work on in order to improve the overall safety or management of the mine. Similar techniques could be applied to other subsets of CWA factors before a comprehensive SEM model is evaluated so that each sub-branch of the complex interactions among human, machine and the environment could be analyzed and understood.

5.4 Chapter Summary

This chapter overviews the basic concepts and applications of causal modeling of theories regarding mine safety and safety science in the mining industry using techniques that involve latent variables. So far there has been a rather limited amount of research that analyzes latent constructs in occupational safety climate and behaviors in mining operations. Nevertheless, latent variable analysis techniques including SEM, CFA, and EFA have been proven in numerous past researches as effective and promising approaches for testing of causal theories in these areas with clear advantages over other quantitative methods including regression. In order to facilitate the adaptation of SEM, CFA and EFA in mine safety analysis, the concepts and their applications, result evaluations and interpretations are explained thoroughly as well as providing critical comparison and including examples of possible usage. In future studies, researchers could incorporate the methods detailed in this chapter in combination with CWA for real life mining applications to better understand and enhance occupational health and safety in the mining industry.

Chapter 6

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

6.1 Mining-Mineral Processing Integration and Target Grades

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 [60-63]. 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 [64], a quadratic function that penalizes deviation from a certain target [65]. It has been proposed that every processing stream maintains a target grade where blocks with the same grade receive no loss from processing, but those with grades different from the target get penalized based on their deviations. An illustration of this idea is shown in Figure 6.1. 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.

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 optimizing mining and mineral processing systems [66-73], but few 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 [74-76].

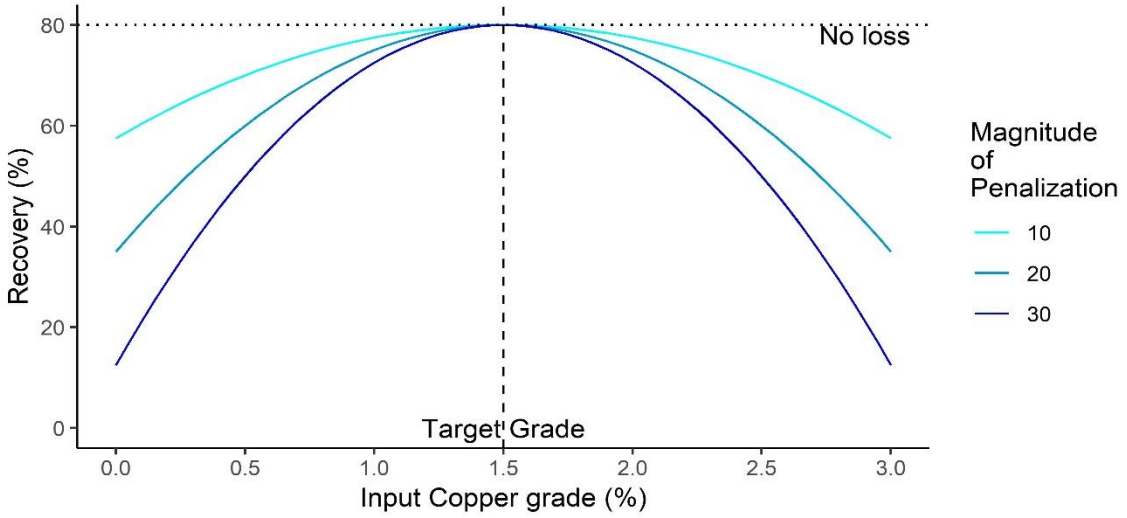


Figure 6.1. Relationship between input grade and recovery modelled by Taguchi loss function

After deciding the optimal number of processing streams through clustering, 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 generally the companies 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. 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 modelling penalized recovery. Moreover, CLARA, which is a robust clustering algorithm for large datasets are used and total revenues from different scenarios are compared and optimized. In addition, block destinations are tweaked according to sequential Gaussian simulations and capacities of processing

streams can be further smoothed across the LoM.

6.2 Methodology of Clustering Algorithms and Economic Evaluations

6.2.1 Clustering algorithms for optimal process design

6.2.1.1 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 prespecified number of clusters in such a way that minimizes the within cluster dissimilarity and maximizes the inter-cluster dissimilarity. 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 as it considers exactly the spatial distance between points. A brief summary of the k-means algorithm is shown in Algorithm 1 [77].

Algorithm 1 K-means Clustering

Input:

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

Number of clusters k

Maximum number of iterations N

Output:

k clusters

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

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 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 a k-means clustering is the total within cluster sum-of-squares (TWSS); its formula is shown in Equation (6.2.1).

$$TWSS = \sum_{i=1}^{k} WSS(i) = \sum_{i=1}^{k} \sum_{j \in C_i} \|\boldsymbol{x}_j - \boldsymbol{\mu}_i\|_2^2 \qquad (6.2.1)$$

Where $\boldsymbol{x}_j$ refers to the jth data point, $\boldsymbol{\mu}_i$ 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; hence it is a common practice to start with many different initial

allocations and 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 could be determined by plotting the TWSS against the number of clusters [78]. Ideally the TWSS value should be minimized, but as the value will always tend to zero when the number of clusters tend 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 would lead to significant diminishing benefit in the reduction of TWSS, identifying the optimal number of clusters in this manner is also more commonly known as the elbow method.

6.2.1.2 Partitioning Around Medoids (PAM) and CLARA

The k-means clustering algorithm has numerous drawbacks, including:

- The number of clusters must be chosen manually
- The final output is dependent on the initial random assignment of cluster numbers
- The algorithm shows sensitivity to noise and outliers due to the use of means

The first and second problem could 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) could be considered similar to a robust-form of k-means clustering. At the cost of $O(k(n-k)^2)$, PAM is still too cumbersome to be applied to truly large datasets. Hence of a modified version of PAM based on resampling named CLARA (Clustering LARge Applications) was selected to optimize processing options.

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 [77].

Algorithm 2 Partition Around Medoids (PAM)

Input:

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

Number of clusters k

Maximum number of iterations N

Output:

k clusters

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

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

The CLARA algorithm is a modified version PAM designed for large datasets, 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 [79].

Algorithm 3 CLARA

Input:

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

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 $O(km^2 + k(n-k))$, which is a significant improvement from PAM. The downside of CLARA is that if the best k medoids are 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.

6.2.1.3 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 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 [80].

Algorithm 4 Spectral Clustering

Input:

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

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 $\boldsymbol{L}$, $\boldsymbol{L} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$, then find the k eigenvectors ($\mathbf{Z} \in \mathbb{R}^{n \times k}$) corresponding to the k smallest eigenvalues of $\boldsymbol{L}$
5: Use k-means to cluster $\mathbf{Z}$ into k clusters, assign the rows of $\boldsymbol{X}$ to the same clusters as rows of $\mathbf{Z}$

Unfortunately, the spectral clustering algorithm is computationally expensive at a complexity of $O(n^3)$, largely due to the need to explicitly construct 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., which functions by first compressing the data into l representative observations, then applying spectral clustering to the compressed data [81]. The ratio $\frac{l}{k}$ is referred to as the compression ratio, a brief summary of the KASP algorithm is shown in Algorithm 5.

Algorithm 5 KASP

Input:

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

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

6.2.2 Economic Evaluations of Processing Scenarios

After using the k-means algorithm to group the data points into *k* different clusters, the clusters are sorted in ascending order of average grades. A *k* number of different processing streams are then sampled from *n* number of total 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 get sent to processing streams designed to have higher recovery. Therefore, for a given number of *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 Equation (6.2.2).

$$Number\ of\ scenarios = \sum_{k=1}^{m} C_k^n \tag{6.2.2}$$

The idea of 'target grade' is applied in this chapter, such that grade deviation from the mean will receive a penalized recovery during processing can be modeled with the Taguchi loss function [64].

$$L\left(x_j^{\gamma}\right) = c\left(x_j^{\gamma} - \mu_i^{\gamma}\right)^2 \quad \forall x_j \in C_i \tag{6.2.3}$$

Where x_j^{γ} is the value of attribute γ (in the polymetallic case) of the j[th] block in the i[th] 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 calculations are performed on each scenario and the one that maximizes profit is deemed as optimal. Formulas for calculations of revenue and cost are shown in Equation (6.2.4). The representations of the parameters are shown in Table 6.1.

Table 6.1. List of parameters for revenue and cost calculations

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

$$\text{Total revenue} = \sum_{j=1}^{N} R(x_j) = \sum_{j=1}^{N} \sum_{\gamma \in \boldsymbol{a}} x_j^\gamma \times \rho \times V \times P_\gamma \times [r_i^\gamma - L(x_j^\gamma)]$$
$$\text{Total cost} = Total\ construction\ cost\ for\ processing\ streams + Total\ mining\ cost + Total\ processing\ cost \quad (6.2.4)$$
$$\text{Total cost} = \sum_{i=1}^{m} M_i + N \times m_c \times \rho \times V + \sum_{j=1}^{N} \sum_{i=1}^{m} y_{ji} \times \rho \times V \times p_i$$

6.2.3 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 could 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 are sent to their most likely destination to reduce variation in processing capacities. When the high-grade processing is over the mean capacity and low-grade processing is under the mean capacity, marginal low-grade blocks currently sent to high-grade processing are switched to low-grade processing to fill the gap, and if there are not sufficient blocks, then marginal low blocks currently sent to waste will also be switched to low processing. When low-grade processing is over the mean capacity and high-grade processing is under, then marginal low-grade blocks will be sent from waste to low grade processing and marginal high-grade blocks from low grade processing to high grade processing as well. Similarly, if both processing streams are over or under the mean capacity, then the marginal waste blocks currently in low and high processing are sent to waste or marginal low and high blocks are sent respectively to low and high processing. While switching destinations, marginal blocks are ranked according to the descending order of likelihood, hence blocks with highest likelihoods are switched first. A schematic of the process is shown in Figure 6.2. By switching the destination of the marginal blocks to their corresponding most likely destinations, variation in processing capacities across the mine life can be effectively reduced at a minimum level of risk.

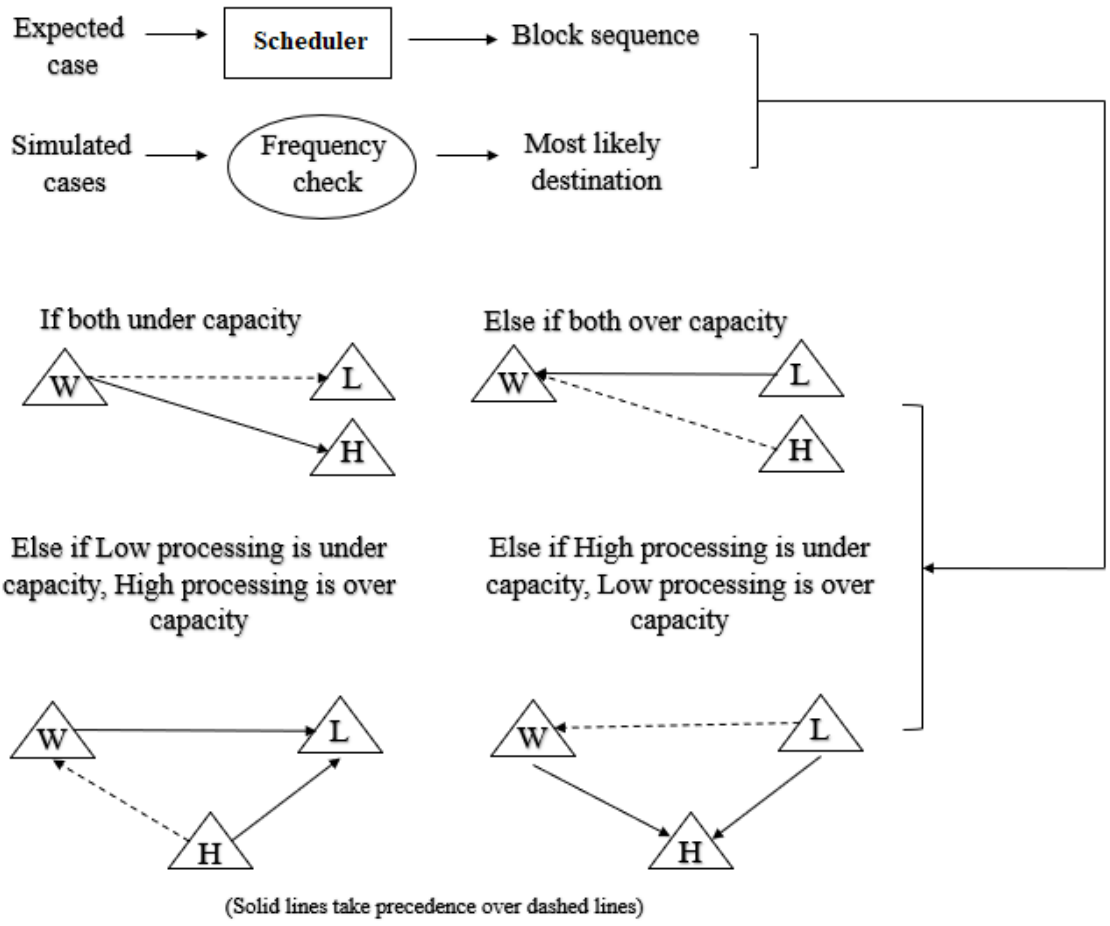


Figure 6.2. Identification and changing destination of marginal blocks

6.3 Mining-Mineral Processing Integration Case Study

6.3.1 Determination of Optimal Processing Scenario

A relatively large data set related to a copper deposit was used in this study. The data set contains 145,800 blocks with 15 equally likely geostatistical simulations generated with sequential Gaussian simulations [82, 83]. The simulations are realized on the nodes or locations of a random grid. In this simulation, conditioning data are converted to equivalent normal values, and the variography of the converted values is computed. Using conditioning and previously simulated values, the value is then estimated (kriged) at the simulation location of the grid. A random sample is finally taken from the distribution characterized by the estimated kriged value and its variance at the simulation location on the grid. This process is repeated for all locations on the grid. In addition to generating multiple realizations of grade uncertainty, Sequential Gaussian simulation

were also used to reproduce variability in various engineering phenomena such as soil water content [84], the standard penetration tests to characterize soil exploration [85], nickel contamination [86] and appraising geochemical anomaly [87]. As expressed by Dowd [88], 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, (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 [89] discussed the effect of the number of simulations on transfer functions and concluded that sequential Gaussian simulation produced more accurate outcomes. He also emphasized that having more than 20 simulations has not much effect on accuracy.

In order to compensate for computational complexities, a relatively small number of possible processing stream options are considered. Detailed information regarding those processing options is shown in Table 6.4. A list of profitability parameters used in this case study is detailed in Table 6.3. A histogram depicting the expected average of the 15 simulations are shown in Figure 6.3.

Table 6.2. 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 6.3. List of profitability parameters

Parameter	Representation	Unit
P_{copper}	Price of copper per ton	$5939.1
m_c	Mining cost per ton	$1.75
c	Magnitude of penalization	30
V	Block volume (Block size $5m \times 5m \times 10m$)	250 m^3
ρ	Block bulk density	4 ton/m^3

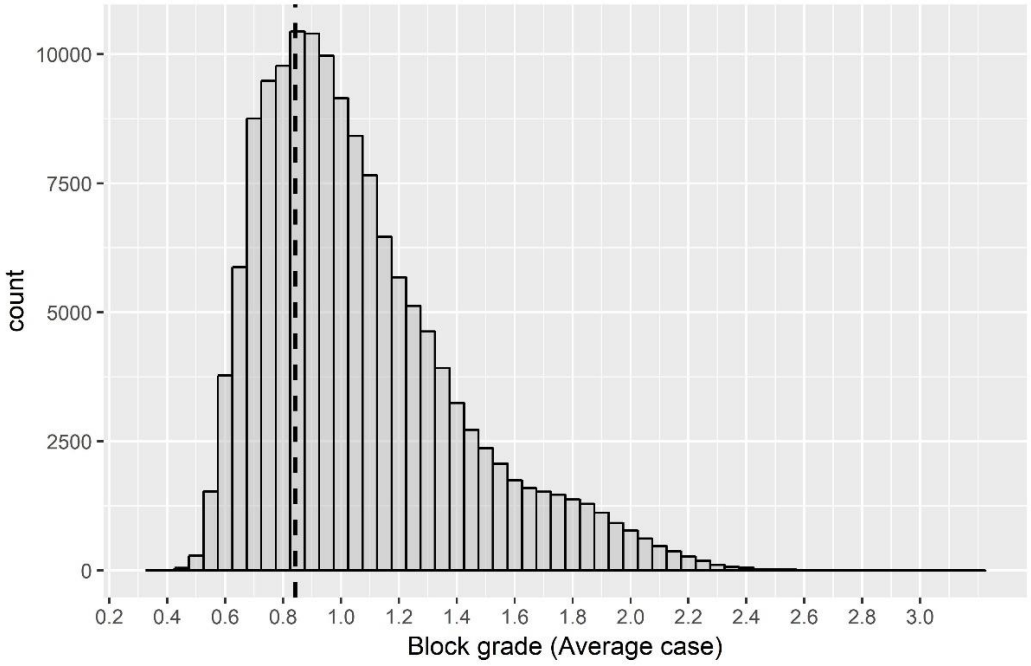


Figure 6.3. Simulated average block grades

In this case study, blocks with grades lower than the lowest possible cut-off 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 were 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 6.4. 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 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 could be observed that k-means has only marginally better TWSS when compared with CLARA, even when medoids are used as cluster centres instead of means in CLARA, while KASP performed similarly to K-means before 3 clusters, but fluctuated with more clusters, possibly due to the small compression ratio used.

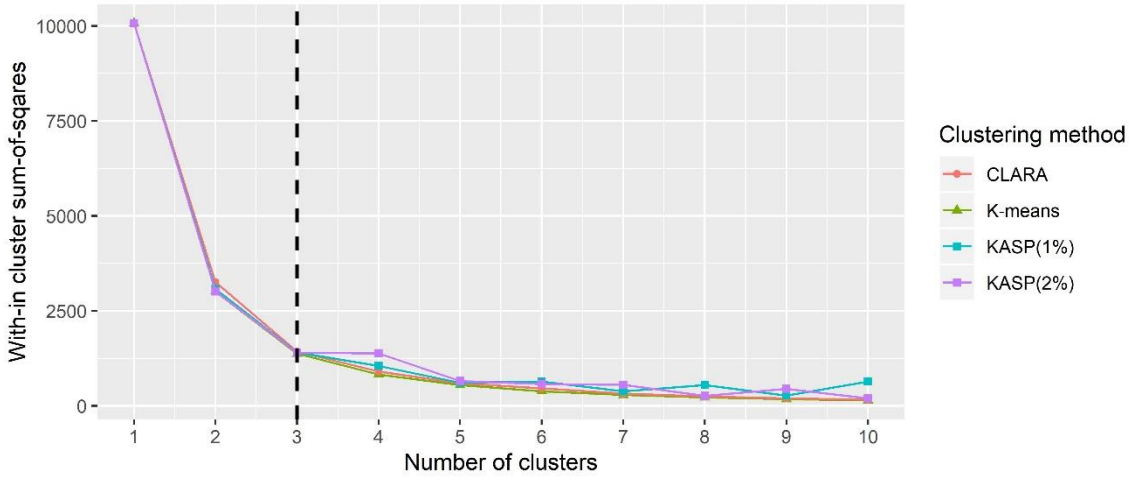


Figure 6.4. TWSS plot for CLARA and k-means clustering

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 cut-off grade, respectively. Table 6.4 details the possible processing scenarios, where in each scenario the clusters are mapped with different corresponding processing destinations. For instance, in processing scenario 7, the data set was partitioned into 2 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 are computed exhaustively and shown in Figure 6.5.

Table 6.4. 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

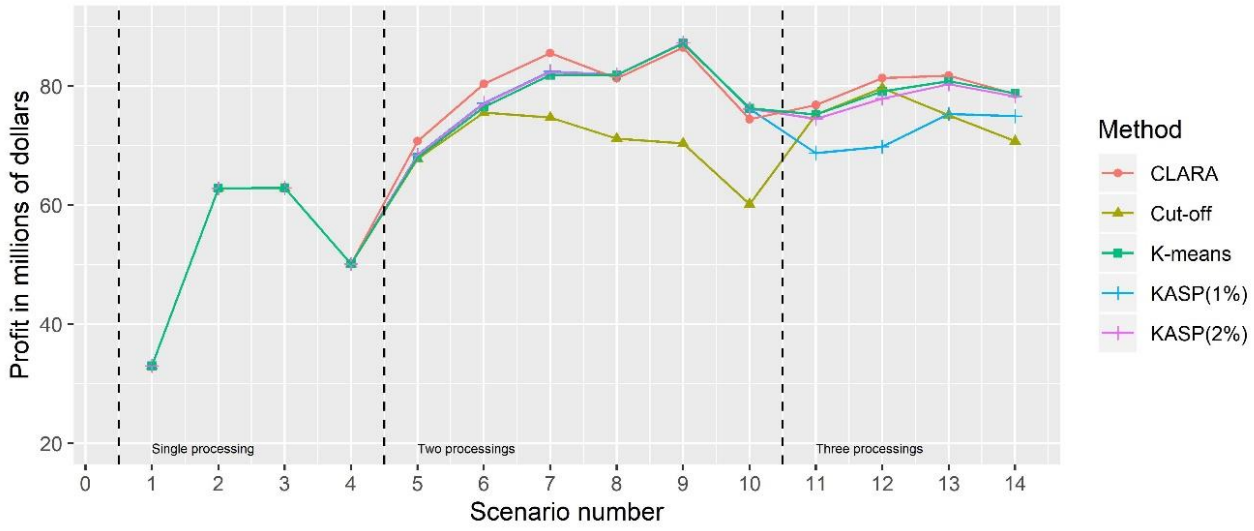


Figure 6.5. Comparison of profits for different clustering results at various scenarios

As can be seen from Figure 6.5, 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 $87.25M. In general, when the deviations from target grades are penalized in mineral processing, determining block destinations via clustering algorithms generate higher profits when compared to using marginal cut-off 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 than CLARA at its scenario with maximum profit, while CLARA at processing scenario 9 resulted in a profit of $86.47M. It could also be seen that KASP performed better in all scenarios when the compression ratio was increased. If higher computational power were available KASP could be projected to yield even better results.

6.3.2 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, Whittle output a total mine life of 10 periods (years). After identifying the borderline blocks, in each period, depending on the situation, borderline blocks are sent to their most likely destination to reduce variation in processing capacities. When the high-grade processing is over mean capacity and low-grade processing is under mean capacity, borderline low-grade blocks currently sent to high-grade processing are switched to low-grade processing to fill the gap, and if there are not sufficient blocks, then borderline low blocks currently sent to waste will also be switched to low processing. Vice versa when low-grade processing is over mean capacity and high-grade processing is under. Similarly, if both processing streams are over or under mean capacity, then the borderline waste blocks currently in low and high processing are sent to waste or borderline low and high blocks are sent respectively to low and high processing. While switching destinations, borderline blocks are ranked according to the descending order of likelihood, hence blocks with highest likelihoods are switched first. The processing capacities of processing streams across the mine life before and after the switching are shown in Table 6.5. The final year of mine life was intentionally left out as most of the valuable ores have been mined out and there is not sufficient among out material left to be mined. The details of mean and variances of processing capacities across the mine life are shown in Figure 6.6. 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 borderline blocks managed to lower the variance in low-grade processing capacity by 31% and that of high-grade processing by 17%. As a result of the re-classification of blocks, a smoothing effect on the processing volumes throughout the periods can be observed.

Table 6.5. Processing capacities for old and new sequencing

Low processing	Mean	Variance
Old sequencing (before switching)	7948	420622
New sequencing (after switching)	7904	289407
High processing	**Mean**	**Variance**
Old sequencing (before switching)	2891	1442632
New sequencing (after switching)	2779	1187592

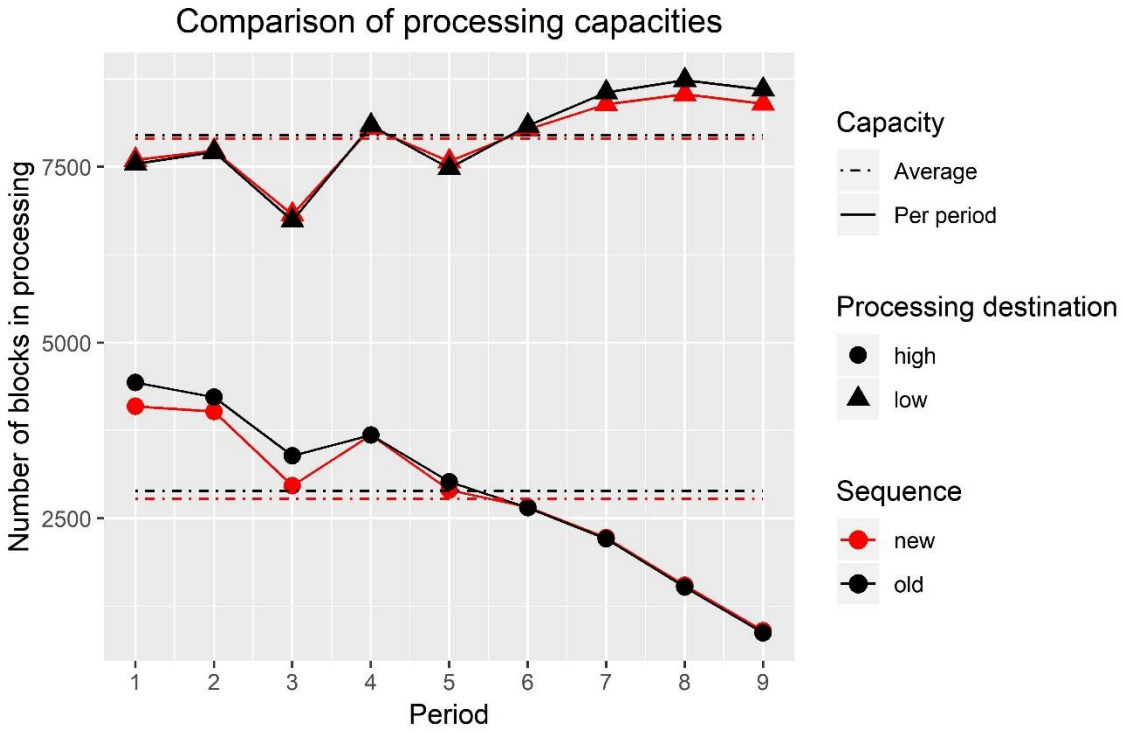
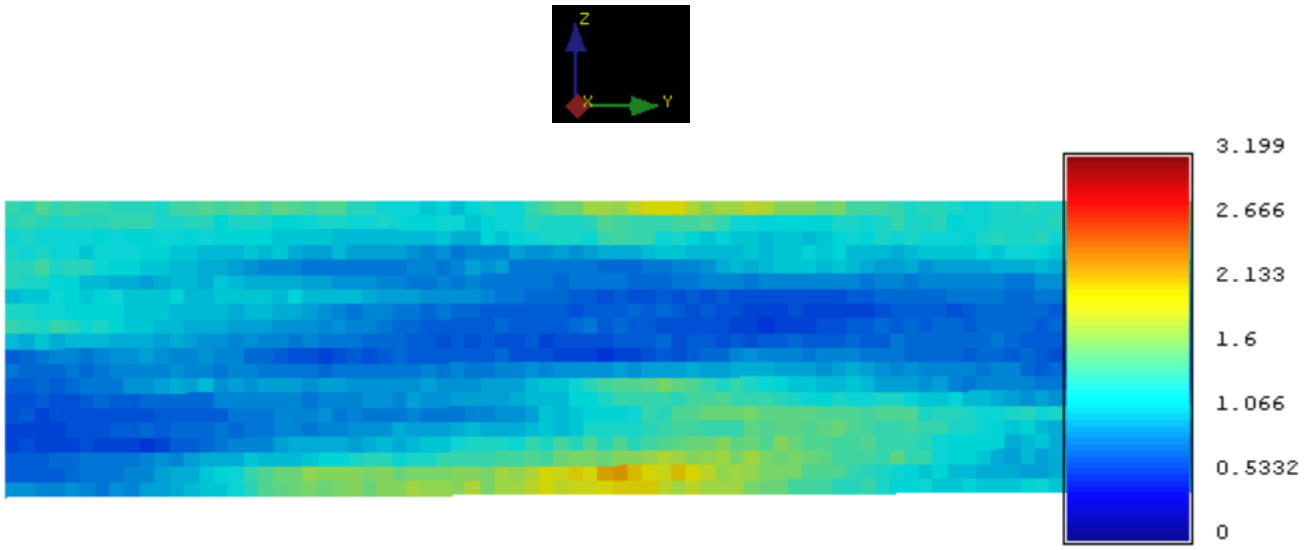
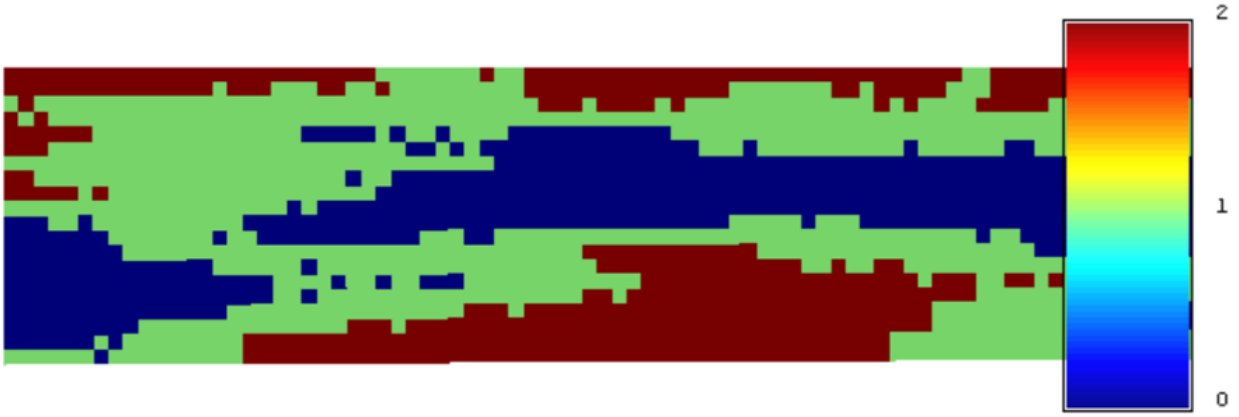


Figure 6.6. Processing capacity across mine life for old and new sequencing

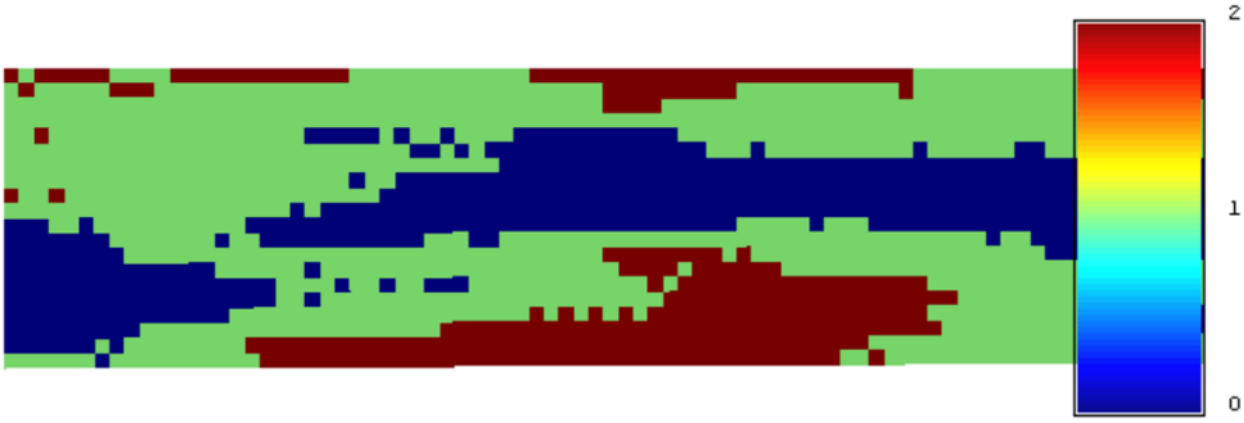
Figure 7 shows in-situ grades and the outcomes of CLARA and KASP for three destinations. In this figure, the blocks shown in navy blue, green and claret red are routed to waste dump, low-grade and high-grade processing, respectively. The consistency between the grades and block destinations can be seen easily in the figure. As also seen from the figure, the number of blocks to be sent to high grade processing is slightly more in CLARA's results compared to KASP's.



(a) Block grades (%)



(b) Results of CLARA



(c) Results of KASP

Figure 6.7. Block grades (a) and Process destinations of blocks (b: CLARA and c: KASP)

6.4 Chapter Conclusions

This chapter 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 to be 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, both clustering methods should be considered in grouping blocks with similar grades. Furthermore, by identifying borderline 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 mine life can be reduced at minimum risk and cost. Although the simulated grades may differ than 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. After the settling of the processing capacities, the mine schedule can be generated with the new parameters. The other extension will be incorporation of rock and metallurgical characteristics affecting processing performance into the process design.

Chapter 7 Final Conclusions and Future Works

Various concepts from other fields such as statistics, mathematics, computer science and machine learning could be used to further optimize mining related problems. In particular, as more and more data are collected and stored about mining operations, it is increasingly important that people are able to take advantage of the heightened amount of information and make more informed, data-driven decisions. Machine learning has significant potential to carry additional value to mining operations.

In Chapter 3, it was shown that PCA based solution helped reducing the problem while maintaining most information in the original data. The PCA based design for stockpiles could be especially useful in polymetallic cases, and benefit of applying PCA could increase with the number of material-grade variables involved in the design. Also determined from the simulation was that Chevron and Windrow stockpiles with same dimensions had very similar effectiveness in terms of VRR. For future analyses, more advanced modelling techniques could used to simulate the blending process, as opposed to the linear model used in this thesis.

The abilities of GLMs to model data that are discrete in nature were shown in chapter 4, possible future works regarding GLMs include more advanced methods such as mixed models, it could be demonstrated that using statistical techniques could adequately model data related to mining, quality control and reliability engineering, and quantitative models such as GLMs could give researchers a more nuanced understanding of the relationships among variables. possible future works regarding GLMs include more advanced methods such as mixed models, quasi-likelihood methods and Bayesian inference.

Chapter 5 showed that factor analysis techniques such as EFA, CFA and SEM could be used as quantitative tools in cognitive work analysis of mine safety, the benefits of factor analysis techniques over standard multiple regression methods were discussed but carefully designed questionnaires and data collection are still required in order for it to be applied in real-life mining scenarios. Nevertheless, factor analysis techniques have very well-established theoretical backgrounds and are ideal tools to model organizational, psychological data in similar scenarios.

Chapter 6 takes into consideration realistic concerns such as deviation from target grades and capacities in processing facilities and penalize recovery of blocks via the Taguchi loss function. It was shown that clustering-based destination policies in general performed better than marginal cut-off-grade based methods. And in terms of clustering methods, KASP generated the optimal

scenario while CLARA also performed better than k-means in all except the optimal scenario. Future works regarding material from this chapter include conducting case studies on polymetallic, multivariate datasets, using more advanced clustering methods. It was shown that KASP with just 2% compression ratio already outperformed k-means in most scenarios, a parallelizable spectral clustering algorithm can almost certainly be expected to have better performance. It is also worth mentioning that the determination of the type of loss function, as well as the magnitude of penalization in recovery for blocks that deviate from target grades are of great practical significance in future works.

There are also various other potential applications of machine learning techniques in mining engineering that are yet to be explored. For instance, with the inclusion of large amount of relevant data on equipment reliability, neural networks and support vector machines could be used to accurately predict failures and optimize maintenance schedules. Also worth further investigations are the possible utilization of long short-term memory neural networks in the prediction of commodity prices in mineral economics, or using reinforcement learning and graph representation learning in the better optimization of mining complexes and decision making.

References

[1] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations* (Springer series in statistics). New York: Springer, 2001.

[2] A. Agresti, Foundations of linear and generalized linear models, Hoboken, New Jersey: John Wiley & Sons Inc., 2015. [Online]. Available.

[3] H. Zou and T. Hastie, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology),* vol. 67, no. 2, pp. 301-320, 2005.

[4] M. Sauvageau and M. Kumral, "Analysis of Mining Engineering Data Using Robust Estimators in the Presence of Outliers," *Natural Resources Research,* journal article vol. 24, no. 3, pp. 305-316, September 01 2015.

[5] C. Wang, J. Wang, X. Wang, H. Yu, L. Bai, and Q. Sun, "Exploring the impacts of factors contributing to unsafe behavior of coal miners," *Safety Science,* vol. 115, pp. 339-348, 2019.

[6] M. Rezania, A. A. Javadi, and O. Giustolisi, "An evolutionary-based data mining technique for assessment of civil engineering systems," *Engineering Computations,* vol. 25, no. 6, pp. 500-17, / 2008.

[7] A. K. Mishra, S. V. Ramteke, P. Sen, and A. K. Verma, "Random Forest Tree Based Approach for Blast Design in Surface Mine," *Geotechnical and Geological Engineering,* journal article vol. 36, no. 3, pp. 1647-1664, June 01 2018.

[8] T. Wen and B. Zhang, "Prediction Model for Open-pit Coal Mine Slope Stability Based on Random Forest," *Science & Technology Review,* vol. 32, no. 4-5, pp. 105-9, 02/18 2014.

[9] W. Li, Y. Zhao, W. Meng, and S. Xu, "Output Prediction Model in Fully Mechanized Mining Face Based on Support Vector Machine," in *2009 Second International Workshop on Knowledge Discovery and Data Mining*, 2009, pp. 171-174.

[10] S. Chatterjee, A. Dash, and S. Bandopadhyay, "Ensemble Support Vector Machine Algorithm for Reliability Estimation of a Mining Machine," *Quality and Reliability Engineering International,* vol. 31, no. 8, pp. 1503-1516, 2015.

[11] B.-r. Chen, H.-b. Zhao, Z.-l. Ru, and X. Li, "Probabilistic back analysis for geotechnical engineering based on Bayesian and support vector machine," *Journal of Central South University,* journal article vol. 22, no. 12, pp. 4778-4786, December 01 2015.

[12] L. Wenyu, W. Meng, and Z. Xinguang, "Model for prediction of surface subsidence coefficient in backfilled coal mining areas based on genetic algorithm and BP neural network," *Journal of Computational Methods in Sciences and Engineering,* vol. 16, no. 4, pp. 745-53, / 2016.

[13] G. Acuna *et al.*, "NARX neural network model for predicting availability of a heavy duty mining equipment," in *2nd Latin-America Congress on Computational Intelligence, LA-CCI 2015, October 13, 2015 - October 16, 2015*, Curitiba, Brazil, 2016, p. Associacao Brasileira de Inteligencia Computacional (ABRICOM); Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq); Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES); Universidade Tecnologica Federal do Parana (UTFPR): Institute of Electrical and Electronics Engineers Inc.

[14] Y. Shao, L. Xu, Y. Hu, and X. Ai, "Pressure Prediction in Natural Gas Desulfurization Process Based on PCA and SVR," *Advanced Materials Research,* vol. 962-965, pp. 564-9,

/ 2014.
[15] X. Xu, S. Ding, S. Yang, Z. Zhao, and X. Wu, "Model for predicting coal and gas outburst based on PCA and BP neural network," *Computer Engineering and Applications,* vol. 47, no. 28, pp. 219-22, 10/01 2011.
[16] P. M. Gy, "Sampling of ores and Metallurgical Products during Continuous Transport," 1965.
[17] P. M. Gy, "A new theory of bed-blending derived from the theory of sampling — Development and full-scale experimental check," *International Journal of Mineral Processing,* vol. 8, no. 3, pp. 201-238, 1981/07/01/ 1981.
[18] P. M. Gy, "Optimizing the operational strategy of a mine-metallurgy or quarry-cement works complex," *Canadian Metallurgical Quarterly,* vol. 38, no. 3, pp. 157-163, 1999/07/01/ 1999.
[19] S. Zhao, T.-F. Lu, B. Koch, and A. Hurdsman, "Automatic quality estimation in blending using a 3D stockpile management model," *Advanced Engineering Informatics,* vol. 29, no. 3, pp. 680-695, 2015/08/01/ 2015.
[20] S. Zhao, T.-F. Lu, B. Koch, and A. Hurdsman, "3D stockpile modelling and quality calculation for continuous stockpile management," *International Journal of Mineral Processing,* vol. 140, pp. 32-42, 2015/07/10/ 2015.
[21] G. K. Robinson, "How much would a blending stockpile reduce variation?," *Chemometrics and Intelligent Laboratory Systems,* vol. 74, no. 1, pp. 121-133, 2004/11/28/ 2004.
[22] P. M. Gy, "Sampling of Heterogeneous and Dynamic Material Systems," vol. 10, 1992.
[23] P. A. Dowd, "The design of a rock homogenizing stockpile," *Mineral Processing in the UK, IMM,* p. 63, 1989.
[24] M. Kumral, "Bed blending design incorporating multiple regression modelling and genetic algorithms," *Journal of the Southern African Institute of Mining and Metallurgy,* vol. 106, pp. 229-236, 03/01 2006.
[25] J. Sreejith and S. Ilangovan, "Optimization of wear parameters of binary Al−25Zn and Al−3Cu alloys using design of experiments," *International Journal of Minerals, Metallurgy, and Materials,* vol. 25, no. 12, pp. 1465-1472, 2018/12/01 2018.
[26] E. Hosseini, F. Rashchi, and A. Ataie, "Ti leaching from activated ilmenite–Fe mixture at different milling energy levels," *International Journal of Minerals, Metallurgy, and Materials,* vol. 25, no. 11, pp. 1263-1274, 2018/11/01 2018.
[27] M. d. Werk, "Trade-off between cost and performance in Chevron bed-blending," M.Eng, McGill University, Montreal, 2017.
[28] D. G. Paterson, M. N. Mushia, and S. D. Mkula, "Effects of stockpiling on selected properties of opencast coal mine soils," *South African Journal of Plant and Soil,* vol. 36, no. 2, pp. 101-106, 2019/03/15 2019.
[29] N. Li, J. Li, H. Long, T. Chun, G. Mu, and Z. Yu, "Optimization Method for Iron Ore Blending Based on the Sintering Basic Characteristics of Blended Ore," in *9th International Symposium on High-Temperature Metallurgical Processing*, Cham, 2018, pp. 455-464: Springer International Publishing.
[30] V. Singh, A. Biswas, S. K. Tripathy, S. Chatterjee, and T. K. Chakerborthy, "Smart ore blending methodology for ferromanganese production process," *Ironmaking & Steelmaking,* vol. 43, no. 7, pp. 481-487, 2016/08/08 2016.
[31] D. M. Marques and J. F. C. L. Costa, "An algorithm to simulate ore grade variability in blending and homogenization piles," *International Journal of Mineral Processing,* vol. 120,

pp. 48-55, 2013/04/10/ 2013.
[32] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 2, no. 4, pp. 433-459, 2010.
[33] J. D. Brown, J. A. Dille, and J. P. Hand, "Quality Control And Shipment Planning at The Carter Mining Co," *Mining Engineering,* vol. 38, no. 12, pp. 1115-1119, 1986.
[34] M. Galetakis, G. Alevizos, F. Pavloudakis, C. Roumpos, and C. Kavouridis, "Prediction of the performance of on-line ash analyzers used in the quality control process of a coal mining system," *Energy Sources, Part A: Recovery, Utilization and Environmental Effects,* vol. 31, no. 13, pp. 1115-1130, 2009.
[35] S. R. Dindarloo and E. Siami-Irdemoosa, "Data mining in mining engineering: results of classification and clustering of shovels failures data," *International Journal of Mining, Reclamation and Environment,* vol. 31, no. 2, pp. 105-18, / 2017.
[36] D. Saha, P. Alluri, and A. Gan, "Prioritizing Highway Safety Manual's crash prediction variables using boosted regression trees," *Accident Analysis & Prevention,* vol. 79, pp. 133-144, 2015/06/01/ 2015.
[37] V. J. Kurian, M. C. Voon, M. M. A. Wahab, N. A. Iskandar, and M. S. Liew, "Multivariate Regression Analysis For Screening Process of Reliability Assessment," *Applied Mechanics and Materials,* vol. 567, pp. 271-6, / 2014.
[38] K. Hwang-Dae, T. J. Robinson, S. S. Wulff, and P. A. Parker, "Comparison of parametric, nonparametric and semiparametric modeling of wind tunnel data," *Quality Engineering,* vol. 19, no. 3, pp. 179-90, 07/ 2007.
[39] K. Bollen, "Latent Variables In Psychology And The Social Sciences," *Annual review of psychology,* vol. 53, pp. 605-34, 02/01 2002.
[40] R. L. Gorsuch, "Factor analysis," in *Handbook of psychology: Research methods in psychology, Vol. 2.*Hoboken, NJ, US: John Wiley & Sons Inc, 2003, pp. 143-164.
[41] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*, Sixth edition [Pearson modern classic edition]. ed. (Pearson modern classic). Upper Saddle River, New Jersey: Pearson, 2019.
[42] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*. 2019.
[43] K. A. Bollen, *Structural equations with latent variables*. Oxford, England: John Wiley & Sons, 1989, pp. xiv, 514-xiv, 514.
[44] T. A. Brown, "Confirmatory factor analysis for applied research, 2nd ed," in *Confirmatory factor analysis for applied research, 2nd ed.*, New York, NY, US, pp. xvii, 462-xvii, 462: The Guilford Press.
[45] E. K. Kelloway, "Using LISREL for structural equation modeling: A researcher's guide," in *Using LISREL for structural equation modeling: A researcher's guide.*, ed. Thousand Oaks, CA, US: Sage Publications, Inc, 1998, pp. ix, 147-ix, 147.
[46] R. Kline, "Principles and Practice of Structural Equation Modeling (2nd Edition)," 01/01 2005.
[47] L. Hayduk, G. Cummings, K. Boadu, H. Pazderka-Robinson, and S. Boulianne, "Testing! testing! one, two, three – Testing the theory in structural equation models!," *Personality and Individual Differences,* vol. 42, no. 5, pp. 841-850, 2007/05/01/ 2007.
[48] D. Hooper, J. Coughlan, and M. Mullen, "Structural Equation Modeling: Guidelines for Determining Model Fit," *The Electronic Journal of Business Research Methods,* vol. 6, 11/30 2007.
[49] D. B. McCoach, A. C. Black, and A. A. O'Connell, "Errors of inference in structural

equation modeling," *Psychology in the Schools,* vol. 44, no. 5, pp. 461-470, 2007.
[50] M. D. Cooper and R. A. Phillips, "Exploratory analysis of the safety climate and safety behavior relationship," *Journal of Safety Research,* vol. 35, no. 5, pp. 497-512, 2004/01/01/ 2004.
[51] R. R. J, *Applied factor analysis*. Northwestern University Press, 1970.
[52] W. Zhang, "Causation mechanism of coal miners' human errors in the perspective of life events," *International Journal of Mining Science and Technology,* vol. 24, 07/01 2014.
[53] P. S. Paul, "Investigation of the role of personal factors on work injury in underground mines using structural equation modeling," *International Journal of Mining Science and Technology,* vol. 23, no. 6, pp. 815-819, 2013/11/01/ 2013.
[54] D.-C. Seo, M. R. Torabi, E. H. Blair, and N. T. Ellis, "A cross-validation of safety climate scale using confirmatory factor analytic approach," *Journal of Safety Research,* vol. 35, no. 4, pp. 427-445, 2004/01/01/ 2004.
[55] T. Liu and Z. Li, "Structural Equation Model for the Affecting Factors of Safety Management Capability of Coal Mine," in *2008 International Workshop on Modelling, Simulation and Optimization*, 2008, pp. 74-77.
[56] D. Komljenovic, G. Loiselle, and M. Kumral, "Organization: A new focus on mine safety improvement in a complex operational and business environment," *International Journal of Mining Science and Technology,* vol. 27, no. 4, pp. 617-625, 2017/07/01/ 2017.
[57] S. Demir, E. Abou-Jaoude, and M. Kumral, "Cognitive work analysis to comprehend operations and organizations in the mining industry," *International Journal of Mining Science and Technology,* vol. 27, no. 4, pp. 605-609, 2017/07/01/ 2017.
[58] H. H. Erdogan, H. S. Duzgun, and A. S. Selcuk-Kestel, "Quantitative hazard assessment for Zonguldak Coal Basin underground mines," *International Journal of Mining Science and Technology,* vol. 29, no. 3, pp. 453-467, 2019/05/01/ 2019.
[59] K. J. Vicente, "Cognitive Work Analysis Toward Safe, Productive, and Healthy Computer-Based Work [Book Review]," *Professional Communication, IEEE Transactions on,* vol. 46, pp. 63-65, 04/01 2003.
[60] G. Verly, "Grade Control Classification of Ore and Waste: A Critical Review of Estimation and Simulation Based Procedures," *Mathematical Geology,* journal article vol. 37, no. 5, pp. 451-475, July 01 2005.
[61] I. M Glacken, "Change of Support and Use of Economic Parameters for Block Selection," 1997.
[62] R. M. Srivastava, "Minimum variance or maximum profitability," CIM Bulletin, vol. 80, no 901, pp. 63 - 98 1987
[63] E. H. Isaaks, "The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data," Stanford University, 1990.
[64] M. Kumral, "Grade control in multi-variable ore deposits as a quality management problem under uncertainty," *International Journal of Quality & Reliability Management,* vol. 32, no. 4, pp. 334-345, 2015.
[65] G. Taguchi, *Introduction to quality engineering: designing quality into products and processes*. The Organization, 1986.
[66] R. C. Goodfellow and R. Dimitrakopoulos, "Global optimization of open pit mining complexes with uncertainty," *Applied Soft Computing,* vol. 40, pp. 292-304, 2016/03/01/ 2016.
[67] M. Yanyan, F. Ferrie, and R. Dimitrakopoulos, "Sparse image reconstruction by two phase

RBM learning: application to mine planning," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA), 18-22 May 2015*, Piscataway, NJ, USA, 2015, pp. 316-20: IEEE.

[68] G.-y. Zhang, G.-z. Liu, and H. Zhu, "Segmentation algorithm of complex ore images based on templates transformation and reconstruction," *International Journal of Minerals, Metallurgy, and Materials,* journal article vol. 18, no. 4, p. 385, July 31 2011.

[69] M. E. Villalba Matamoros and M. Kumral, "Calibration of Genetic Algorithm Parameters for Mining-Related Optimization Problems," *Natural Resources Research,* journal article vol. 28, no. 2, pp. 443-456, April 01 2019.

[70] J. R. Ruiseco, J. Williams, and M. Kumral, "Optimizing Ore–Waste Dig-Limits as Part of Operational Mine Planning Through Genetic Algorithms," *Natural Resources Research,* journal article vol. 25, no. 4, pp. 473-485, December 01 2016.

[71] H. Nguyen, C. Drebenstedt, X.-N. Bui, and D. T. Bui, "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,* journal article March 01 2019.

[72] G. T. Nwaila *et al.*, "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,* journal article May 29 2019.

[73] B. Rajabinasab and O. Asghari, "Geometallurgical Domaining by Cluster Analysis: Iron Ore Deposit Case Study," *Natural Resources Research,* journal article vol. 28, no. 3, pp. 665-684, July 01 2019.

[74] E. Sepúlveda, P. Dowd, and C. Xu, "Fuzzy Clustering with Spatial Correction and Its Application to Geometallurgical Domaining," *Mathematical geosciences,* 07/25 2018.

[75] J. Ruiseco and M. Kumral, "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,* vol. 26, 06/22 2016.

[76] Y. A. Sari and M. Kumral, "Dig-limits optimization through mixed-integer linear programming in open-pit mines," *Journal of the Operational Research Society,* vol. 69, no. 2, pp. 171-182, 2018/02/01 2018.

[77] A. Kassambara, *Unsupervised Machine Learning: Practical Guide to Cluster Analysis in R* (Multivariate Analysis I). STHDA, 2017.

[78] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. Springer New York, 2013.

[79] L. a. R. Kaufman, P.J., "Clustering Large Applications (Program CLARA)," in *Finding Groups in Data*, 2008, pp. 126-163.

[80] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing,* journal article vol. 17, no. 4, pp. 395-416, December 01 2007.

[81] D. Yan, L. Huang, and M. I. Jordan, "Fast approximate spectral clustering," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.

[82] A. G. Journal and F. Alabert, "Non-Gaussian data expansion in the Earth Sciences," *Terra Nova,* vol. 1, no. 2, pp. 123-134, 1989.

[83] J.-P. Chilès and P. Delfiner, "Geostatistics: Modeling Spatial Uncertainty," *Wiley Series In Probability and Statistics,* 01/01 2012.

[84] M. Delbari, P. Afrasiab, and W. Loiskandl, "Using sequential Gaussian simulation to assess the field-scale spatial uncertainty of soil water content," *CATENA,* vol. 79, no. 2, pp. 163-

169, 2009/11/15/ 2009.
[85] H. Basarir, M. Kumral, C. Karpuz, and L. Tutluoglu, "Geostatistical modeling of spatial variability of SPT data for a borax stockpile site," *Engineering Geology,* vol. 114, no. 3, pp. 154-163, 2010/08/10/ 2010.
[86] M. Qu, W. Li, and C. Zhang, "Assessing the risk costs in delineating soil nickel contamination using sequential Gaussian simulation and transfer functions," *Ecological Informatics,* vol. 13, pp. 99-105, 2013/01/01/ 2013.
[87] Y. Liu, Q. Cheng, E. J. Carranza, and K. Zhou, "Assessment of Geochemical Anomaly Uncertainty Through Geostatistical Simulation and Singularity Analysis," *Natural Resources Research,* 06/11 2018.
[88] P. A. Dowd, *Geostatistical Simulation, Course notes for the MSc. in Mineral Resources and Environmental Geostatistics*. Leeds, 1993, p. 123.
[89] P. Goovaerts, "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,* journal article vol. 13, no. 3, pp. 161-182, June 01 1999.