UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

PRACTICAL IMPLEMENTATION OF RADIAL BASIS FUNCTIONS IN SPARSE DATASETS IN RESOURCE ESTIMATION

Rui Goncalves

Student number: 0602008V

School of Mining Engineering

University of the Witwatersrand

Johannesburg, South Africa

Supervisor: Professor Christina Dohm

A research report submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Mining Engineering. MSc (Eng 50/50)

2021

DECLARATION

I, Rui Goncalves (student number: 0602008V), am a student registered for a MSc in Engineering in the year 2020 and hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong;
- I confirm that the work submitted for assessment for the above course is my own unaided work, except where I have explicitly indicated otherwise;
- I have followed the required conventions in referencing the thoughts and ideas of others; and
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Rui Goncalves

15 September 2021

ABSTRACT

The choice of an estimation method in resource estimation is often constrained by the availability of data and the experience of the modeller. Moreover, insufficient data coverage or a highly variable deposit may hinder the application of kriging methods. The radial basis function has widely been applied to generate three-dimensional surfaces and volumes for geological modelling purposes. However, it also has application as an estimator, as the methodology is mathematically equivalent to dual kriging.

Using a vein-hosted copper deposit, the focus of this research study report is to investigate the application of the radial basis function as a resource estimation technique. A single vein from the dataset was targeted and, after analysis, spatially subdivided into a high-grade and a low-grade domain. A stochastic simulation was used to generate a representation of "reality" for each domain, with validations made against these. A single simulation was selected randomly and sampled at a 15 m x 15 m x 1.5 m spacing to emulate a grade control drilling grid. These samples constitute the database and were used as input for the estimation of the two domains identified. For each domain, a three-dimensional block model was generated using three estimation methods – the radial basis function, ordinary kriging, and inverse distance weighting. To ascertain which method was superior at predicting grades locally, the estimates were compared against input data and individually against each other using statistics and visual validations, such as grade distribution and swath plots.

The findings demonstrate that the radial basis function performs equally to ordinary kriging and inverse distance weighting when supported by sufficient sample coverage. In areas with poor coverage, the radial basis function outputs resulted in closely comparable estimates. As the radial basis function is a global estimator, unlike ordinary kriging, confidence in the estimates cannot be quantified easily. Therefore, the radial basis function was evaluated against the 99% confidence intervals of the ordinary kriging estimates. The radial basis function estimates were well within these

confidence limits and similar to the ordinary kriging and inverse distance methods.

The conclusions drawn from this report suggest that radial basis function can provide robust estimates, even when sample coverage is inferior. For grade control purposes, where data coverage is abundant, it can be considered as an alternative estimation method. However, due to the lack of metrics that quantify the confidence in the radial basis function estimates, it is unlikely that this method will displace kriging methods in the statement of Mineral Resources.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Christina Dohm for her valuable input and guidance throughout this research project.

I want to acknowledge and thank First Quantum Minerals Ltd and Kansanshi Mining PLC for permission to use the dataset and relevant information. Moreover, I thank Seequent for providing access to Leapfrog Geo® and Leapfrog EDGE®, without which this work would not be possible.

In addition, I want to thank my family for their encouragement, with special mention of my sister India Goncalves, who took time to assist me with the editing of this report.

Finally, I want to thank my partner, Wernich Hein, for encouraging me and supporting me throughout this research project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3D .... *Three-dimensional*
ANN .... *Artificial Neural Network*
CI .... *Confidence Intervals*
CL .... *Confidence Limits*
Cu .... *Copper*
CV .... *Coefficient of Variation*
DD .... *Diamond Drillhole*
DRC .... *Democratic Republic of Congo*
EDA .... *Exploratory Data Analysis*
FQM .... First Quantum Minerals
HG .... *High-grade*
IDW .... *Inverse Distance Weighting*
KE .... *Kriging Efficiency*
KV .... *Kriging Variance*
LCS .... *Lower Calcareous Sequence*
LD .... *Lower Dolomite*
LG .... *Low-grade*
LM .... *Lower Marble*
LPS .... *Lower Pebble Schist*
ML .... *Machine Learning*
OK .... *Ordinary Kriging*
QKNA .... *Quantitative Kriging Neighbourhood Analysis*
Q-Q .... *Quantile-Quantile*
RBF .... *Radial Basis Function*
RC .... *Reverse Circulation*
SGS .... *Sequential Gaussian Simulation*
SK .... *Simple Kriging*
SoR .... *Slope of Regression*
SVM .... *Support Vector Machine*
UD .... *Upper Dolomite*
UM .... *Upper Marble*

UMC.. *Upper Mixed Clastics*

UPS ....*Upper Pebble Schist*

1 INTRODUCTION

Traditionally, in the mining industry, volumes representing geological or grade domains have been defined by manually digitising polylines in cross-sections and then connecting these to create three-dimensional (3D) surfaces or volumes. Referred to as explicit modelling, this process may be time-consuming, depending on the complexity of the mineral deposit being modelled.

Recently, numerous geological modelling software packages have implemented automatic tools for spatial modelling, which use a mathematical volume function to generate closed volumes or contacts. This methodology, referred to as implicit modelling, has reduced the time required to create and update resource models, transforming how geological modelling is done in the mining industry.

Most implicit modelling packages are founded on the radial basis function (RBF), of which Leapfrog Geo® is the most widely used in the industry. Leapfrog® is a geological modelling software that is largely responsible for the adoption of implicit modelling. However, the use of the RBF goes beyond the creation of wireframes, as the method can also be applied in grade estimation. This creates a necessity to understand how best to practically apply this novel estimation method.

1.1 Project Background

Conventional methods, such as ordinary kriging (OK), simple kriging (SK) and inverse distance weighting (IDW), are well-entrenched in resource estimation in the mining industry. These estimation methods are present in most mining software packages by default, but only recently became available in Leapfrog®, with the introduction of an estimation module called Leapfrog EDGE®. This module introduced the RBF as a resource estimation method to a wider audience.

Since its introduction, the RBF has not seen many practical applications due to kriging being the preferred method for resource estimation. The methodology of applying kriging is well-researched and documented and, as it is a best linear unbiased estimator, it is the method of choice for the estimation of Mineral Resources. The use of the RBF is unlikely to replace kriging, although there may be situations where its application is merited and perhaps more suitable than other more established techniques.

1.2 Research Motivation

Studies on the RBF tend to focus on comparing the outputs with the equivalent of an OK estimate. While this is a useful comparison, there are no guidelines on when and how best to apply the RBF methodology. On the other hand, a wealth of information is available on the application of the various kriging methods, backed up by decades of theory and practical research.

In EDGE®, a user will come across terminology and methods for the RBF analogous to those employed with kriging, including the application of semivariograms. This may create the false perception that the same methodology applied in kriging is transferable to the RBF. Nevertheless, there are many differences that set these two estimation techniques apart, such as the lack of a definable search neighbourhood for the RBF due to it assigning weights to samples globally across the domain.

Furthermore, when quantifying the quality of a kriging estimate, most software packages output metrics, such as slope of regression (SoR), kriging variance (KV) and kriging efficiency (KE), for each estimated block. Similar metrics for RBF do not exist and the absence of statistical basis for measuring confidence in the estimation process makes model validations challenging. This could result in the incorrect application of the RBF or it being overlooked in favour of more established methods.

1.3 Problem Statement

In resource estimation, the choice of the estimator for an ore deposit is often dependent on various factors, such as the geological complexity of the area, the spatial distribution of the sample dataset and the statistical nature of the samples. Traditionally, kriging estimators tend to be the preferred method employed in the industry. However, this is only possible if enough data exist to calculate an experimental semivariogram to allow for the fitting of a model. In cases where data are sparse, this may not be possible as semivariogram models produced on sparse data may introduce unrealistic continuity, and other deterministic methods, such as estimates based on IDW, may be explored.

Estimation of deposits with sparse datasets is often challenging and the techniques available tend to be unreliable and produce biased estimates. This can be problematic in early-stage exploration programmes where insufficient data is often a reality. The RBF may be useful in these situations. Since the RBF is a global estimator, the possibility exists that more robust estimates can be generated compared to a method like IDW. Therefore, with few practical examples to draw knowledge from, this study aims to provide a methodology on how to apply RBF and to identify possible challenges and shortcomings when using this estimation technique.

1.4 Assumptions

The data for this research project were provided under a non-disclosure agreement between the author and Kansanshi Mining PLC, which is 80% owned by First Quantum Minerals (FQM) Ltd. It is assumed that the data were acquired and validated to the highest of standards of quality assurance and quality control and that the alterations done to protect the privacy of the company have not fundamentally changed the nature of the dataset.

1.5 Research Question

The availability of the RBF for grade estimation purposes in mining software packages is a recent development. Due to its novelty, its practical application lags behind more traditional methods largely as a result of the lack of knowledge on how to implement it. Knowing in which cases to apply the RBF becomes important. When data coverage is poor, applying kriging becomes challenging. However, if the RBF delivers an acceptable estimate that is close to the truth, then in the scenario presented here, it could be considered an alternative to IDW. The comparisons used statistical tools and visual observations on the grade outputs of the RBF and compared these against a simulation, representing this truth and against the other two methods.

Consequently, the following question arises: Can the RBF method be used to estimate grades in situations where it is difficult to apply kriging methods and how do these estimates compare against more established methods?

1.6 Research Objectives

Using the dataset available, this research project attempted to quantify and compare the efficiency of an RBF estimate in situations where data coverage is poor. To quantify the effectiveness of the RBF estimate, comparisons were made against a "true" value. These "true" values were created from a stochastic simulation to generate a picture of a "reality" on which the RBF estimates can be compared against. Additionally, IDW and OK comparisons were made against this "reality" to determine if a particular method is superior.

The aim of this study was to identify how sparse data affects the quality of the RBF estimate and when its application is justified for situations where the use of kriging is not possible. If proven to be a reliable estimator, the RBF may be seen as an alternative estimator.

1.7 Research Methodology

The methods applied to the research are rooted in practices applied in geological modelling and resource estimation. This research study used the following methods and tools, in order of the research project flow:

- Literature review of applications of the RBF in resource estimation and how this method relates to conventional estimation methods;
- Manipulation of the data to conceal the location and true grades of the source dataset and honour the consent agreement in place with FQM;
- Statistical analysis of the data;
- Geological modelling of stratigraphic units and vein modelling;
- Exploratory data analysis (EDA) of data, domaining of different data populations, and creation of sample composites;
- Normal score transformation of datasets, geostatistical analysis, and semivariogram modelling;
- Optimisation of estimation and search parameters through quantitative kriging neighbourhood analysis (QKNA);
- Generating multiple conditional simulations using Sequential Gaussian Simulation (SGS);
- Sampling of simulated outputs to create a theoretical grid of samples to be used in estimation;
- Selecting the most appropriate RBF parameters;
- Estimation of grades into a block model using the three estimation techniques: RBF, OK, and IDW;
- Validation of model outputs against inputs using statistical analyses;
- Assessing the three estimation methods against each other using statistical and visual tools; and
- Reaching a conclusion by interpreting the results presented and providing concluding remarks and recommendations on the application of the RBF for estimation purposes.

1.8 Research Study Layout

This research report is laid out in a way that is succinct and easy to follow for the reader. The research report is structured as follows:

- Chapter 1 introduced the research topic, outlining the motivation for this study, framing the research question and objectives to the reader, and providing an overview of the methodology used in this research.
- Chapter 2 summarises the available literature, broaching topics that are relevant to the study, such as applications of the RBF in volume modelling, its relationship to kriging as an estimator, and giving examples in the literature of its application.
- Chapter 3 outlines the geological setting for the source of the data, providing an overview of the regional and mine geology as well as describing the mineralisation.
- Chapter 4 details, step by step, the methodology used in this research study – from data manipulation, through geological modelling and domain definition, to conditional simulation and estimation of the domains.
- Chapter 5 presents the results for the estimations of the RBF, OK, and IDW, and compares these against each other and the simulated data representing the truth. The confidence in the RBF estimates is evaluated against the 99% confidence intervals of the OK estimates.
- Chapter 6 discusses the results presented in the preceding chapter and gives an explanation for the observed similarities and differences between the three methods.
- Chapter 7 draws conclusions from the observations made in this research study and discusses these to answer the research question.

2 LITERATURE REVIEW

2.1 Introduction

Historically, the mining industry has always faced challenges in quantifying the in-situ grades and tonnages of an orebody. Before the late modern period of the 20th century, estimates were based on direct observations, taken as the orebody became exposed during mining operations. Often, mining progressed with little or no information about the orebody, guided by luck and ingenuity of individuals seeking fortunes in the ground. Frequently, poor understanding of the geology and orebody would result in dwindling returns and the abrupt closure of a mine. History is dotted with these boom-bust cycles, such as the iconic gold rushes of the 1800s. Thus, the importance of estimating resources with a high degree of confidence cannot be overstated for the successful operation of a mine.

As time progressed, mining companies and mining operations became more organised and, with this, the collection of information to assist operations became more structured and abundant. The first resource estimations were created from data collected for grade control purposes, used to monitor mining operations (Glacken & Snowden, 2001). Glacken and Snowden (2001) noted that these first estimates, which can be considered reserves, had dilution and recovery factors built into the estimates and used a simplistic geometric method, such as polygonal estimation.

Geometric methods, such as polygonal estimation, triangulation, and estimation by cross-sections, which are all variants of a nearest-neighbour estimation, gave way for deterministic methods concerned with assigning weights to samples as a function of their distance, such as inverse distance estimation. In the 1940s, work by Bertil Matérn applied in forestry was used to compute the standard errors for spatial surveys, which formed part of his 1960 thesis "Spatial Variation" and is credited with laying the mathematical foundations for spatial statistics (Guttorp & Lindgren, 2018). From the mid-20th century, work by Herbert Sichel, Danie Krige, and Georges Matheron introduced statistical methods into the mining industry, which used the spatial correlation between samples to optimally assign weights during estimation, as applied in kriging.

Technological advancements in personal computing in the latter half of the previous century meant that these complex methods could be applied easily to block modelling software for resource and reserve estimation. However, Glacken and Snowden (2001) pointed out that during the latter half of the 20th century, lack of knowledge by practitioners led to poor application of geostatistical methods in the industry, resulting in a tarnished reputation. While this slowed the adoption of kriging as an estimator in some parts of the world, the method eventually gained almost global acceptance in the industry, with applications to transformed data, such as indicator kriging, in nonlinear estimation methods and in stochastic simulations.

The 21st century has seen the rapid development of more powerful personal computers, which have allowed a range of new techniques to find their way into mining software packages. Computational learning techniques, such as machine learning (ML), support vector machine (SVM), and artificial neural networks (ANN), fall outside the topic of discussion for this research project, but are worth mentioning, as recently they have become popular topics of discussion. Presented as data-driven alternatives to geostatistical methods, these methods often require little knowledge of the mathematics underpinning them, which is seen as an advantage over geostatistical methods. Furthermore, computational learning techniques are data-driven, which do not rely on the same assumptions of stationarity as kriging for the input variables and are comparatively easier to modify and apply to different deposits (Kaplan & Topal, 2020).

A method that has gained popularity over the past two decades is the RBF. Used as an algorithm for creating surfaces and volumes, the RBF has seen practical applications as an estimation technique recently. In ML, the RBF is popular as an ANN or a kernel used in SVM techniques. According to

Afonja (2017), a kernel is defined as a class of algorithm utilised for pattern analysis that employs a linear classifier to solve non-linear problems.

This research project will focus on the RBF and its application as a linear estimator. The RBF use in ML techniques is discussed briefly but does not form part of this research project. The literature review begins with an introduction of the RBF as a modelling tool for 3D volumes and surfaces. Thereafter, how this method relates to more established estimation methods are discussed and examples of its application as an estimator are given.

2.2 Radial Basis Function – An Implicit Modeller

Implicit modelling is a recent addition to the lexicon of the mining industry. The term was popularised by Jun Cowan (Cowan et al., 2002; Cowan et al., 2003) as an alternative to the traditional, manual method of creating 3D modelling wireframes. Consequently, this traditional method has been labelled as explicit modelling.

Explicit modelling utilises computer-aided design tools to create surfaces that represent geological and orebody contacts. This method tends to be laborious and time-consuming, requiring a user to digitise strings, representing geological contacts, in sections along drillhole lines, which are then linked up to create a volume or surface in 3D. Cowan et al. (2002) stated that the main disadvantage of this method is the uniqueness of the model, which is dependent on the interpretation of the geologist at the time of wireframing, thus preventing its replication. This has implications when successive versions of a model need to be reconciled to its predecessors.

Implicit modelling was introduced as a faster alternative to the manual, explicit methods that have dominated the industry. This methodology uses a volume function that, on the condition that the same data and parameters are applied, ensures the replicability of the volume model. According to Cowan et al. (2003), the advantages of this method include the ability to process small and large datasets, to model structurally complex geology, and to dynamically update models when new data are introduced. Importantly, the method allows for geological models to be generated at greater speeds. This time-saving aspect means that multiple models can be generated quickly, whereas explicit methods often rely on a single interpretation of the geology, with limited iterations allowed for, due to time constraints.

Cowan et al. (2003) presented the basic definition of an implicit model as a function defined throughout space that describes a solid. This function is modelled by spatially interpolating drillhole data from which the surface of the solid is extracted. The surfaces are said to exist implicitly in a continuous volume function.

A volume function generates an isosurface that includes the sample points and interpolates throughout space points where no information is available. Contact points from the sample data are assigned the same value to ensure the output surface honours all the input data – typically this is given a value of zero. Lithology codes are converted into numeric values and assigned positive values when representing the interior of the lithology unit being modelled, while exterior lithologies to this unit are assigned negative values, which ensures that an isosurface conformed to the contact points (zero values) can be interpolated correctly (see Figure 2.1).

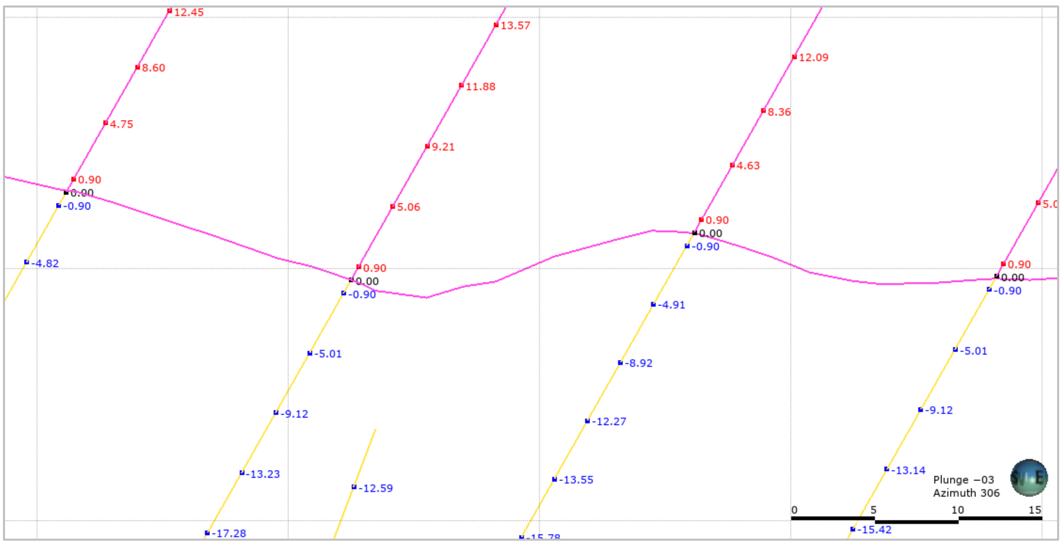


Figure 2.1: Example of an implicit surface, showing contact values (0), interior lithology as positive, and exterior as negative values. The values represent distance to the contact being modelled. The implicit surface was created in Leapfrog Geo®

As a modelling method in the mining industry, implicit modelling was driven initially by the geological modelling software, Leapfrog®. Since then, effectively all competing software brands have introduced an implicit modelling tool. While a variety of algorithms can be used for implicit modelling, including numerous forms of kriging, the RBF is the algorithm of choice for implicit modelling.

The RBF is an isotropic function, with its values dependent on the Euclidian distance from a fixed origin (Chilès & Delfiner, 2012). It consists of a series of interpolation functions, of which the multiquadric functions were shown early on to be useful at approximating topography with relatively few data points available (Hardy, 1971). This is considered by Cowan et al. (2003) as the first introduction of the RBF to geological literature, although its application was constrained by computer hardware at that time, as excessive processing power was needed for calculating and storing the coefficients (i.e., weights) of what was considered a modest number of data points. In a mining operation, tens of thousands of data points are typically available, making the method unsuitable unless computing speeds were improved. Thus, a fast form of RBF, named FastRBF™ was developed by Seequent, based on work by Beatson et al. (1999), which greatly increased processing speeds for large datasets, making it a viable method used in the mining industry for volume modelling.

2.3 The Radial Basis Function as an Estimator

The use of the RBF in the mining industry has been exclusively as a surface interpolator for the past two decades. Recently, through its implementation in mining software packages and in various programming languages, such as Python and R, the RBF has been applied as an estimator for Mineral Resources.

The use of the RBF as an estimator is described mathematically by Chilès and Delfiner (2012), who showed that the solution for the RBF corresponds with dual kriging. The authors explained that the dual term originates from an alternative derivation of the kriging equations by minimisation of space, similar to splines. A spline is a set of low-degree polynomials used in curve design built from many individual segments, which yields a tighter curve than fitting a high-degree polynomial through a set of given points (Salomon, 2006).

A dual kriging estimator, $z^*(x)$ is given by Chilès and Delfiner (2012) in the form:

$$z^*(x) = \sum_\alpha b_\alpha \, \sigma(x_\alpha, x) + \sum_l c_l f^l(x)$$

The first component on the left, ($\sum_\alpha b_\alpha \, \sigma(x_\alpha, x)$), is probabilistic, describing the local influence of samples around a point, and is the weighted sum of covariances between the samples x_α, and the point x. The second component on the right, $\sum_l c_l f^l(x)$ is the drift component, is deterministic, and is locally a linear combination of known polynomial functions $f^l(x)$ of unknown coefficients c_l, affecting the estimate away from data, throughout space (Stewart et al., 2014).

The RBF estimator is given by Chilès and Delfiner (2012) in the form:

$$\Phi^*(\mathrm{x}) = \sum_\alpha b_\alpha \, \varphi(|x - x_\alpha|) + \sum_l c_l f^l(\mathrm{x})$$

Where x_α are the points where the interpolant is to be constructed; b_α are the weights; and $\varphi(|x - x_\alpha|)$ is the isotropic, spatial distance function from which the RBF derives its name.

The two estimators shown above illustrate how similar the RBF is to kriging. The derivation of the dual kriging equations and RBF fall outside of the scope of this research study, but are discussed in detail in Chilès and Delfiner (2012), Costa et al. (1999), Stewart et al. (2014), and Yamamoto (2002).

Stewart et al. (2014) highlights two important parameters that are user-defined when applying RBFs. The first is the choice of spatial function, which performs a similar role to the variogram model in kriging. The spherical variogram model, typically used in kriging, is presented as an option, but it was observed to cause artefacts when generating grade shells. Therefore, Seequent developed the spheroidal model, which combines a linear behaviour near the origin with a function that approaches the sill value asymptotically. The spheroidal variogram has since also been introduced in other resource estimation software, such as in Datamine Supervisor® (Stewart, 2019).

The second parameter relates to the choice of degree of the polynomial in the drift model, with four options to choose from: none, constant, linear, and quadratic. Stewart et al. (2014) noted that all models, except for the “none” option, perform similarly within areas that are well informed. The use of no drift is similar to SK with a zero mean and is unsuitable for grade models. The authors cautioned against the use of the linear and quadratic drift models due to concerns with extrapolation, suggesting the constant option as being the safest for grade interpolation. A constant drift performs similarly to SK with a mean value equal to the declustered sample mean.

Literature on the application of the RBF in Mineral Resource estimation tends to focus on its use in ANNs and SVMs, with the outputs often compared against OK. Using a simulated dataset of unspecified source, Costa et al. (1999) compared the outputs of an RBF network to kriging and concluded that, due to the use of a variogram to define the spatial relationship, kriging was the more accurate method. Similar comparisons using an RBF ANN can be found in a study by Samanta (2010), who used data from a gold-placer deposit to observe that the RBF had a higher predictive efficiency relative to kriging, while noting that the RBF resulted in higher bias due to overestimated grades for low sample values. Samanta (2010), suggested that the RBF was superior at predicting higher sample values, resulting in estimates with a lower mean squared error, a measurement of estimator quality, even though kriging resulted in the least biased estimates. More recently, RBF neural networks were applied for predicting mineral prospectivity and were found to be more reliable than using a model produced from a deep learning SVM algorithm (Ghezelbash et al., 2019).

The available literature tends to focus overwhelmingly on the application of the RBF in ML techniques, which often requires a strong understanding of computer programming and a theoretical background that falls outside the scope of this research study. However, a few examples exist on the direct application of the RBF as an estimator. Yamamoto (2002), took a more direct approach in applying the RBF in block modelling to a copper deposit in Brazil, detailing a methodology for choosing and applying an RBF to estimation. The author concluded that the method is reliable and comparable to OK, offering an alternative when there is insufficient data for the modelling of experimental semivariograms. Furthermore, Yamamoto (2002), concluded that, similarly to kriging, the RBF can assign adequate weights to samples in clustered datasets, making it a superior method to IDW.

Stewart et al. (2014), used a simulated low-grade iron ore deposit to compare the RBF and OK estimates. Two scenarios were tested, a regular 10 m x 10 m drilling grid and a semi-regular 20 m x 20 m drilling grid. The RBF rendered results essentially identical to the OK estimates in areas informed by data, where data spacing was less than the variogram range.

In response to the work of Stewart et al. (2014), which used a negatively skewed example, the study by Kentwell (2014) utilised a positively skewed gold dataset. Kentwell (2014), concluded that, when used as a point estimator and regularised into blocks, the RBF was comparable to OK and did not need data transformation typically required from skewed datasets. On the other hand, as a direct block estimator, the RBF was found to be inferior and to not account for block support correctly.

In a recent study by Dos Santos and Yamamoto (2019), the RBF was applied successfully to a structurally controlled gold deposit. This dataset, characterised as highly variable, with a pure nugget effect in the main orientation of the deposit, meant that experimental semivariograms could not be modelled for the three orthogonal directions. In this case, the application of kriging was impossible and the RBF provided robust estimates.

2.4 Chapter Summary

The RBF has seen wide adoption as a modelling tool for volumes and surfaces. Applications of the RBF in computational learning methods and as an estimator are largely academic and the method is not used commonly for resource estimation.

Where semivariogram models cannot be fitted to the experimental data, due to either highly variable datasets or sparse information, the RBF offers a viable alternative to kriging. Stewart et al. (2014) stated that the RBF does not require the same assumption of stationarity as in kriging, although this will result in more reliable estimates. Additionally, as an estimator that uses global drift coefficients and sample weights, the RBF lacks the need for various checks and optimisations to a search neighbourhood as required for estimation by kriging and minimisation of the estimation variance.

As shown by Chilès and Delfiner (2012), the use of a global neighbourhood in the formulation of dual kriging results in a system of equations in which the coefficients do not change with location, thus they only need to be calculated once for all locations. However, this does not allow for the calculation of the KV – that is, the minimised estimation variance – which is minimised during a kriging estimation. The KV is used in resource estimation as a measure of confidence, which is based on the sample to block configuration and the variogram within the search neighbourhood and is independent of the grade values (Coombes, 2008). Moreover, the KV is utilised to calculate two metrics used to measure the conditional bias of a kriged estimate, namely KE and SoR. These metrics play a role in the classification of Mineral Resources and are integral to expressing confidence in the kriging estimate of a Mineral Resource.

Since no equivalent metric is available for the RBF, assessing the quality of an estimate becomes challenging. Therefore, any assessment is reliant on comparisons against kriging methods. Based on the literature review, the application of the RBF has been shown to be viable where estimates are informed by dense datasets. This study will look at how the method performs in scenarios where local estimates are informed by sparse data. As this can be a common scenario in early-stage exploration projects and, to a lesser extent, mining operations, having an alternative method to kriging that performs in a comparable way can be of great value.

3 GEOLOGY

Located in north-western Zambia, the Kansanshi copper deposit lies in the Central African Copperbelt, a fold belt that straddles the Zambian and Democratic Republic of Congo (DRC) border (Chinyuku, 2013). The Central African Copperbelt is the world's largest and highest-grade sediment-hosted copper-cobalt province in the world (Broughton, 2014), with the DRC contributing 1.4 million tonnes of finished copper per annum and Zambia contributing approximately 750 000 tonnes of the metal, based on 2019 production figures (Pillay, 2021).

3.1 Regional Geology

The Central African Copperbelt is a sequence of Neoproterozoic age (1 000–541 million years ago) metasedimentary rocks belonging to the Katangan Supergroup, which were deposited onto older basement rocks of the Palaeoproterozoic age consisting of schists and granitoids (Chinyuku, 2013). The belt, as illustrated in Figure 3.1, is subdivided into the DRC Copperbelt representing a basement-distal deposition setting and the Zambia portion representing a basement-proximal setting (Theron, 2013).

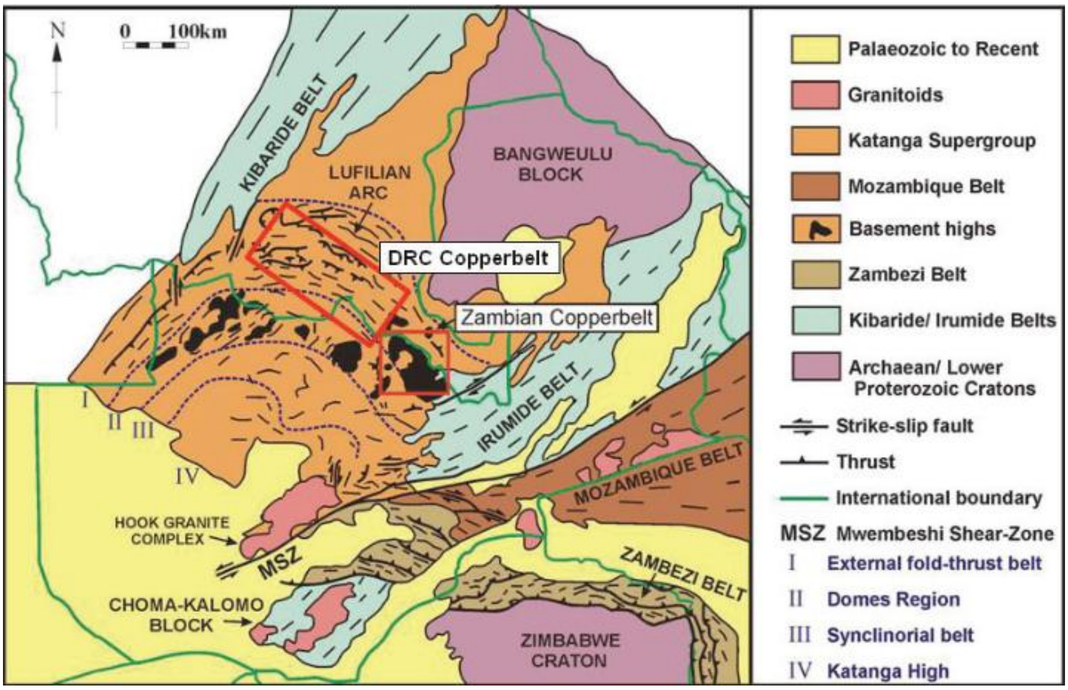


Figure 3.1: Regional setting of the African Copperbelt showing relative locations of the DRC and Zambian belts (Theron, 2013)

Subsequently, this sequence of rocks was deformed and metamorphosed during the pan-African Damaran-Lufilian orogenic event 700–500 million years ago (Torrealday et al., 2000). This event is identified by a complex deformational history that resulted in a fold and thrust belt that, in the Kansanshi area, is characterised by the presence of pre-Katangan basement inliers (see Figure 3.2), which give the name to the Domes Region (Chinyuku, 2013; Torrealday et al., 2000).

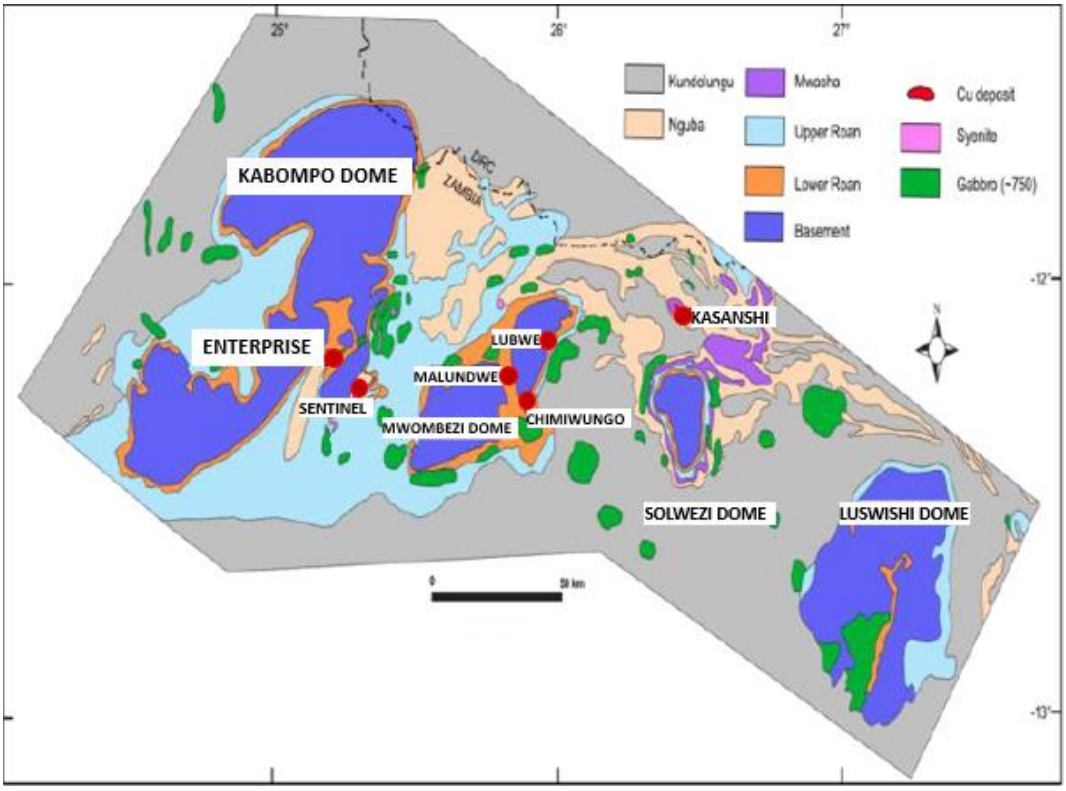


Figure 3.2: Regional geological map of the Domes Region showing the location of Kansanshi mine (Hitzman et al. as cited in Chinyuku, 2013)

3.2 Mine Geology

The Kansanshi deposit is located 15 km north of the Solwezi Dome (refer to Figure 3.2), an inlier of basement rock part of a series of pre-Katangan basement rocks that trend east-west along the inner Lufilian fold belt (Chinyuku, 2013). The stratigraphic sequence at Kansanshi is well-understood from the abundant drilling campaigns and geophysical surveys. The mine stratigraphy has been grouped into a tectonostratigraphic sequence of units that, from top to bottom, consist of the Upper Dolomite (UD) and the Upper Pebble Schist (UPS) Formations. Underlying these units are the Kansanshi Mine Formations, which comprise the Upper Marble (UM), Upper Mixed Clastics (UMC) and Middle Mixed Clastics (MMC), Lower Calcareous Sequence (LCS), and Lower Marble (LM) Formations. Beneath the Kansanshi Mine Formation are the Lower Pebble Schist (LPS) and the Lower Dolomite (LD) Formations (Chinyuku, 2013). The stratigraphic sequence is illustrated in Figure 3.3.

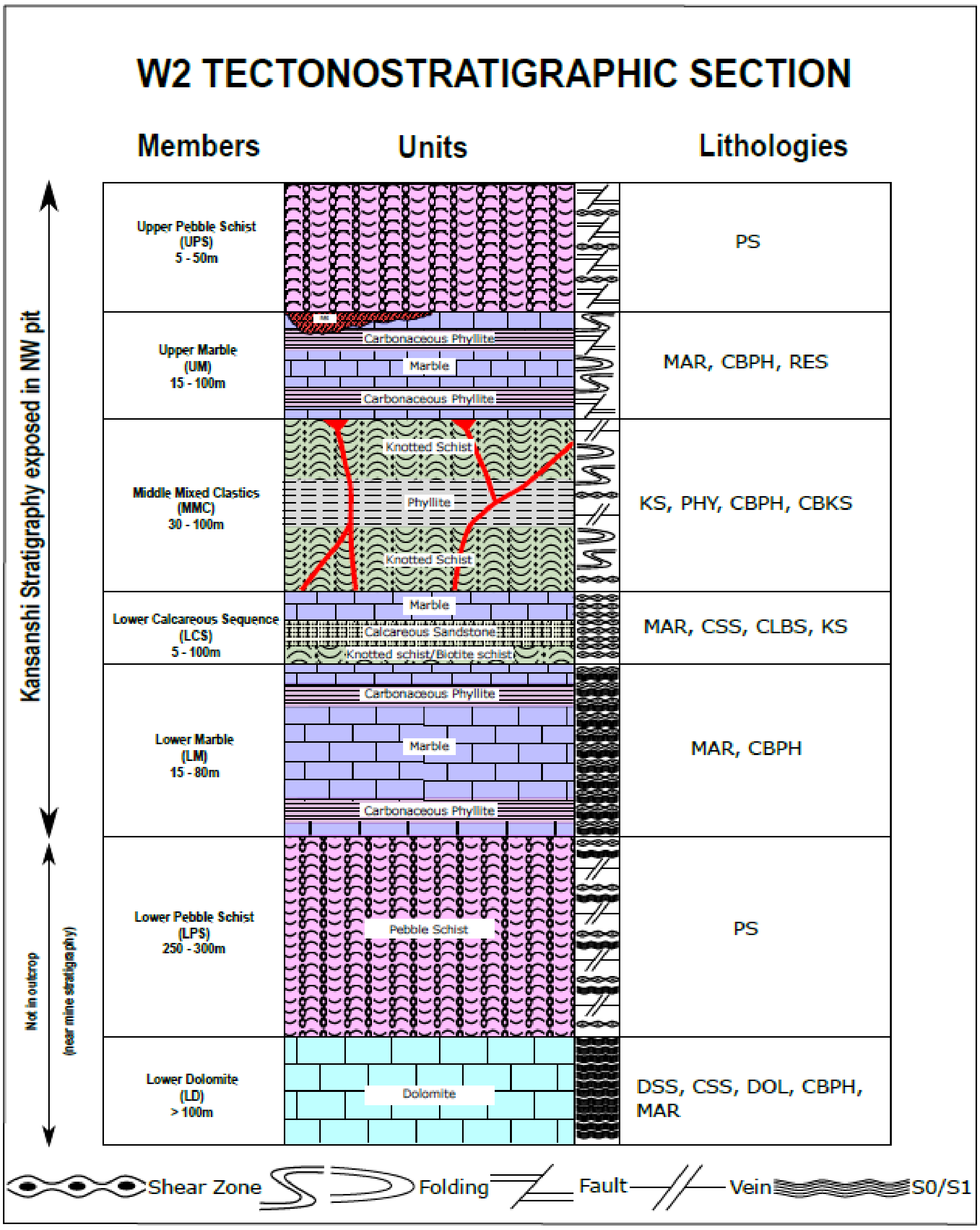


Figure 3.3: Tectonostratigraphic sequence used at Kansanshi Mine (obtained via R. Bartman, 2019, pers. comm.)

3.3 Mineralisation

The mineralisation at Kansanshi is associated strongly with steeply dipping, undeformed quartz dolomite-calcite-chalcopyrite-pyrite veins of variable thickness (Chinyuku, 2013; Torrealday et al., 2000). Stratiform mineralisation occurs as thin bands that run parallel to bedding, while disseminated mineralisation tends to be associated with albite or carbonate alteration (Chinyuku, 2013).

The mineralised veins occur as a stockwork, with undeformed veins tending to strike north-northwest and sub-perpendicular to bedding. While vein-hosted mineralisation occurs across all tectonostratigraphic units of the mine, the MMC hosts a large proportion of these structures (Chinyuku, 2013).

4 METHODOLOGY

4.1 Data Preparation

A dataset was provided by FQM consisting of 550 diamond drillholes (DD) and 518 160 reverse circulation (RC) drillholes used for grade control purposes. In addition, ancillary data, including geology, alteration, weathering, and structural information, as well as topographical data were provided.

To preserve the privacy and nature of the data, the drillholes were transformed into a different spatial setting by translating the X and Y coordinates of the drillhole collars into a new location. Moreover, the assay data provided for this project were altered by multiplying the grade values by a factor, without having an impact on the underlying sample distribution. As a result, the original grade values were not used in this research project.

4.2 Research Area Selection

The data provided covered a very large area. Therefore, a smaller subset of the data was selected as the focus of this research project. The data inside the area selected for this project are shown in Figure 4.1, with DD and RC drillholes identified in the legend.

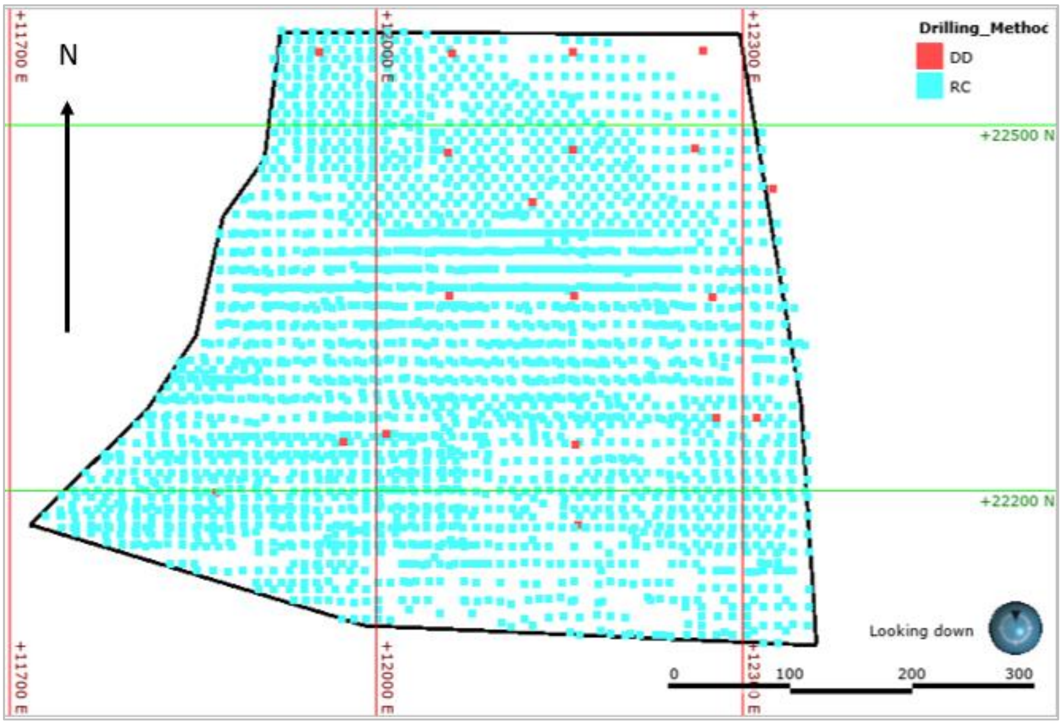


Figure 4.1: Spatial distribution of RC and DD drilling inside the selected area

The definition of this smaller area resulted in a reduced dataset, details of which are listed in Table 4.1.

Table 4.1: Data selected inside the area of interest

Data type	Reverse circulation	Diamond drillholes	Total data
Collars	2 856	20	2 876
Survey	4 367	163	4 530
Assays	66 704	3 664	70 368
Lithology	68 446	1 021	69 467

4.3 Geological Modelling

4.3.1 Stratigraphic model

Implicit modelling was used to create 3D meshes representing geological contacts and weathering surfaces for the area of interest. Implicit modelling has gained popularity over the past 20 years due to its ease of use and ability to dynamically update models built with large datasets.

When building models for geologically complex deposits, it is best to always start with large-scale units, such as stratigraphic sequences or structural domains. This assists the modeller with simplifying the information available and making the modelling process more manageable. Once these broader units are defined, smaller-scale units can be modelled inside the larger ones, thus adding complexity and higher resolution to the model as required.

For the area of interest, 36 lithological units were aggregated into broader, tectonostratigraphic zones guided by modelling practices followed by the mine site and the sequence illustrated in Figure 3.3. This resulted in six units from the stratigraphic sequence being modelled, as listed in Table 4.2.

Table 4.2: Lithological grouping used for modelling the tectonostratigraphic sequence in the area of interest

Model code	**Unit name**	**Lithologies**
UM	Upper Marble	Marbles and carbonaceous phyllite
MMC	Middle Mixed Clastics	Schists and phyllites, some of which are carbonaceous
LCS	Lower Calcareous Sequence	Marble, calcareous sandstone, calcareous biotite schist and knotted schist
LM	Lower Marble	Marble and carbonaceous phyllite
LPS	Lower Pebble Schist	Pebble schist
LD	Lower Dolomite	Dolomitic sandstone, calcareous sandstone, dolomite, marble and carbonaceous phyllite

Leapfrog Geo® provides a range of modelling tools for creating geological contacts, including:

- Deposit
- Erosion
- Intrusion
- Vein/Vein System
- Structural Surfaces

The application of these tools is subject to the type of contact being modelled and the user's preference, as similar outputs can be achieved using different tools. The names given to these tools may allude to specific geological formations, but their applications can be adapted depending on the complexity and structure of the deposit being modelled.

The Deposit and Erosion surfaces are very similar and are best used for tabular or gently dipping contacts – the major difference between them is the ability of the Erosion surface to cut chronologically older units. As the name suggests, the Intrusion tool is best suited for creating closed volumes, but it lends itself well to structurally complex, folded units. The Vein tool is best used for units with clearly defined hanging wall and footwall boundaries, such as dykes, or shear zones. Structural Surfaces are created from structural data measured along contacts or from non-contacting bedding planes. The outputs created from this method incorporates the strike and dip measurements in the orientation of the mesh. Due to the folded nature of the stratigraphy in the area of interest, the Intrusion tool was used to model the tectonostratigraphic package of rocks. The main units – UM, MMC, LCS, LM, and LPS – were modelled from top to bottom, out of preference. After the contacts were defined, output volumes representing each unit were generated as shown in the cross-section in Figure 4.2.

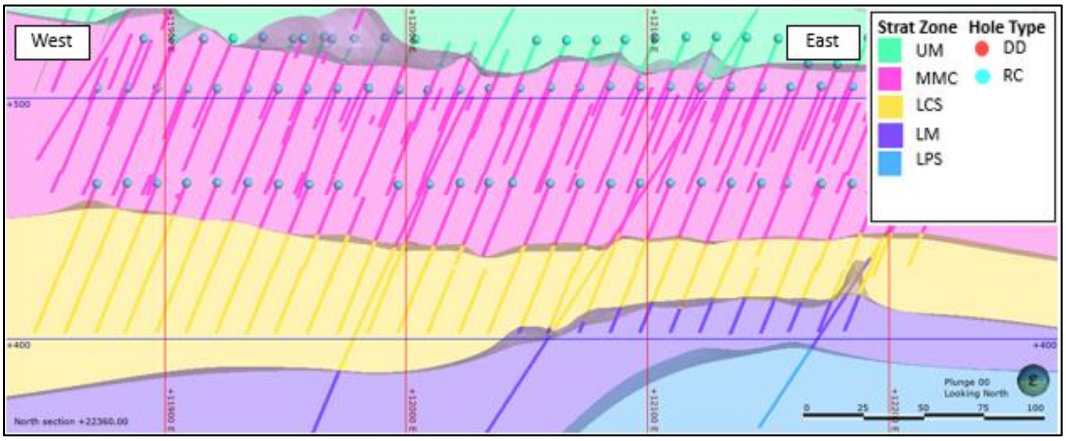


Figure 4.2: Section view – modelled tectonostratigraphic layers looking north

4.3.2 Vein model

Consistent with observations and the literature describing the geology of the mine, the mineralised quartz-carbonate veins were observed to be present predominantly in the MMC unit. Using the Vein tool, two dominant trends were identified, with one set of veins striking in a north-northwest orientation and a second set striking in an east-west orientation. The process involved flagging onto the drillhole logs individual veins based on the continuity of these structures along strike and down-dip. In total, 36 individual veins were identified and modelled, ranging from less than one metre up to 15 metres in thickness. Continuity along strike varied between a few tens of metres up to 400 metres along strike.

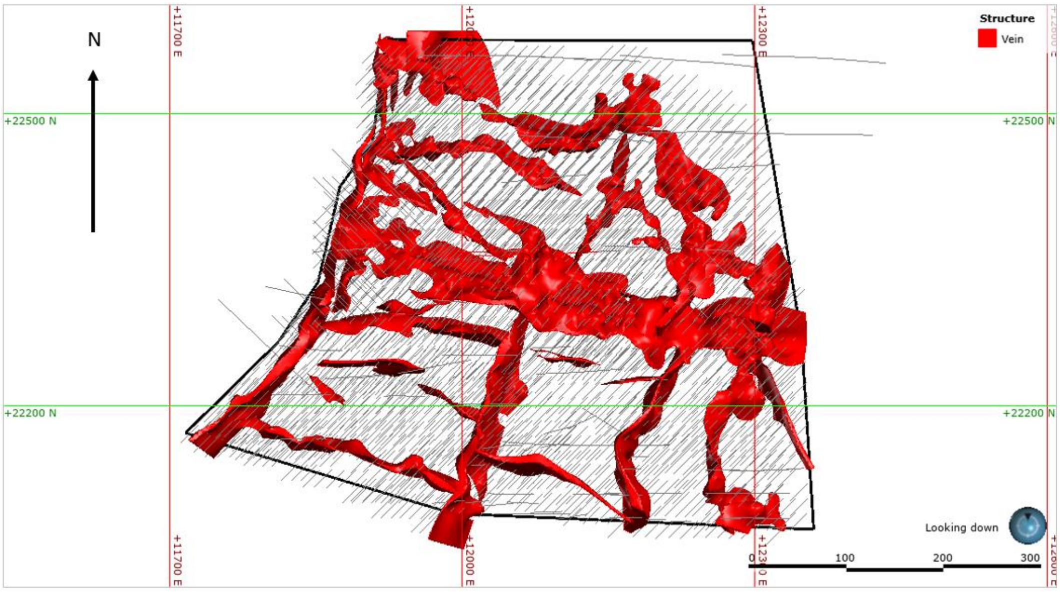


Figure 4.3: Plan view: modelled veins (red) and drillhole traces (grey) inside the area of interest

Due to the large number of veins modelled (Figure 4.3), a single vein was selected for the focus of this research project, namely Vein 16 (see Figure 4.4).

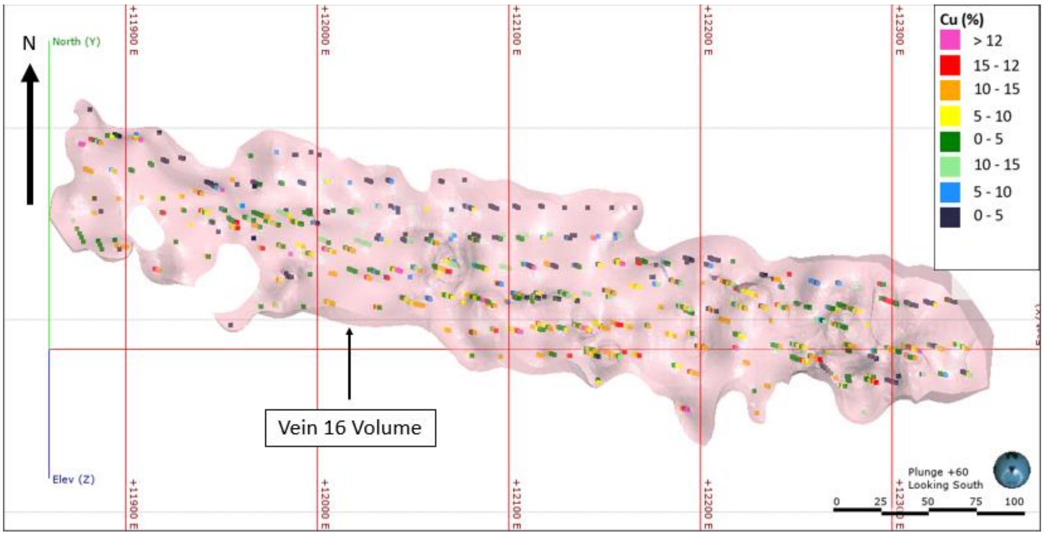


Figure 4.4: Modelled volume of Vein 16 looking along 60° plunge towards the south

The criteria for selecting Vein 16 were based on it being the longest vein modelled, thus containing the largest number of samples to be used for estimation purposes. Vein 16 strikes approximately east-west, crossing the entire length of the project area. It dips between 60° and 70° north and it has a variable thickness of a few metres up to 20 metres in some locations. While Figure 4.4 shows Vein 16 in plan view, Figure 4.5 presents Vein 16 in an orthogonal view so as to illustrate this variable thickness more clearly.

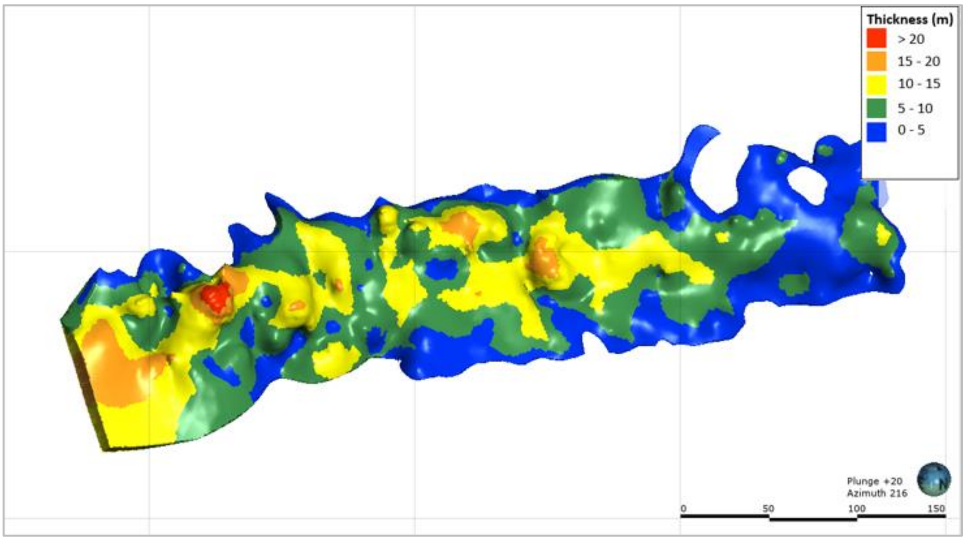


Figure 4.5: Orthogonal view of Vein 16 coloured on thickness (metres) – view is looking southwest at an azimuth of 216° and plunge of 20°

4.4 Exploratory Data Analysis of Samples

EDA was conducted for total copper assays selected inside the vein. EDA entails the application of statistical analysis to the sample dataset as a means of characterising the different data populations and identifying any mineralised subdomains that may be important for accurate resource estimation (Abzalov, 2016). This is an iterative process, often requiring the initial domain to be further subdivided to achieve a homogenous, statistical population that remains invariant under translation, in what is known as stationarity (Abzalov, 2016; Rossi & Deutsch, 2014). Strict stationarity is difficult to achieve and seldom found in nature. Therefore, second-order stationarity, where the mean and variance are constant, is used in practice (Abzalov, 2016). Summary statistics (Table 4.3) and a histogram (Figure 4.6) for the sample grade data inside Vein 16 show that the distribution of copper assays is positively skewed with a long tail (mode<median<mean).

Table 4.3: % Cu summary statistics

Domain	Vein 16
Number of samples	1 060
Minimum	0.001
Maximum	33.63
Range	33.63
Mean	2.44
Median	1.07
Mode	0.01
Variance	13.51
Standard Deviation	3.68
CV	1.51

The mean copper (Cu) grade of the samples is 2.44% Cu, and the dataset is highly variable, due to the high coefficient of variation (CV) value of 1.51.

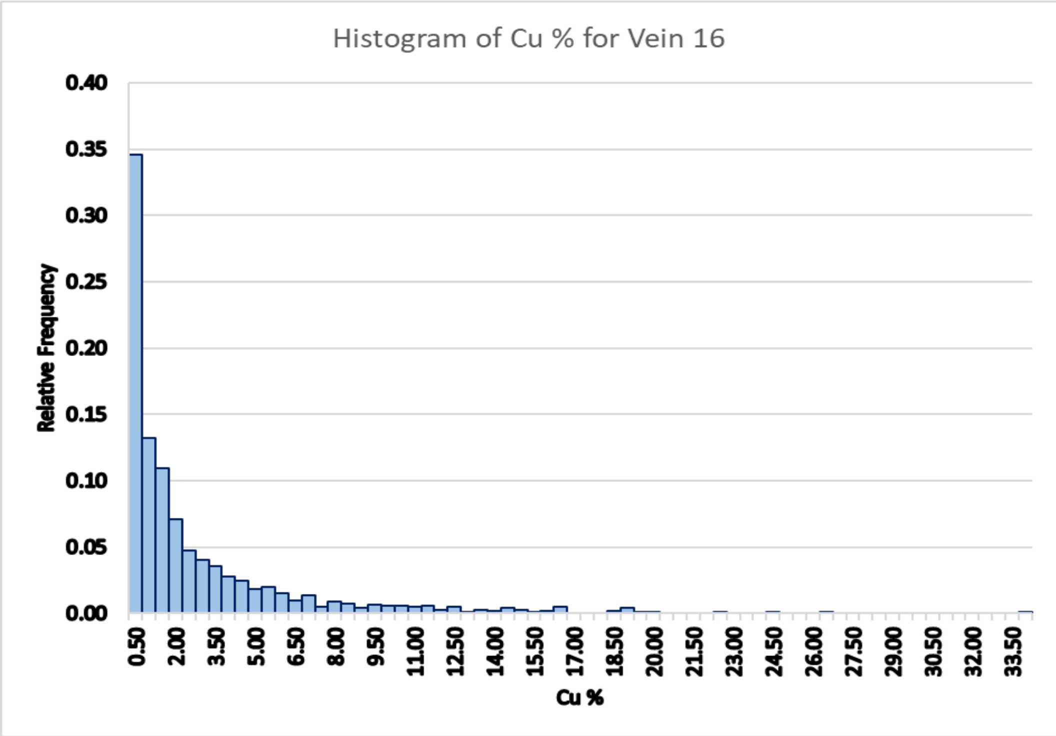


Figure 4.6: Histogram of the copper (% Cu) grades inside Vein 16

The histogram in Figure 4.6 suggests the presence of a lower-grade domain, with up to 35% of the sample population having a grade below 0.5% Cu. A log probability plot showing the distribution of copper values inside Vein 16 is presented in Figure 4.7. The mean of the dataset is highlighted in red, while the 0.5% Cu threshold identified in the histogram (refer to Figure 4.6) is indicated by a blue line on the graph in Figure 4.7. Additional subdomains are noted at higher cut-off values, as evidenced by inflection points along the plotted line, although these were not explored due to the paucity of the data.

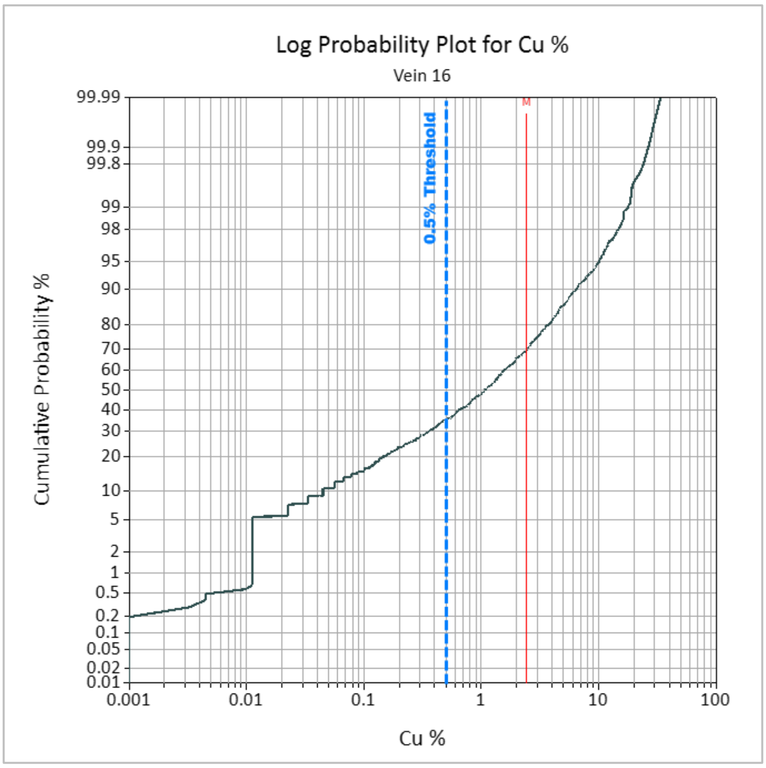


Figure 4.7: Log probability plot of copper (% Cu) for Vein 16

The copper grades were plotted against elevation in ten-metre bands (see Figure 4.8), showing that grades progressively drop with decreasing elevation. This indicates that a low-grade domain is predominantly present in the lower portions of the vein.

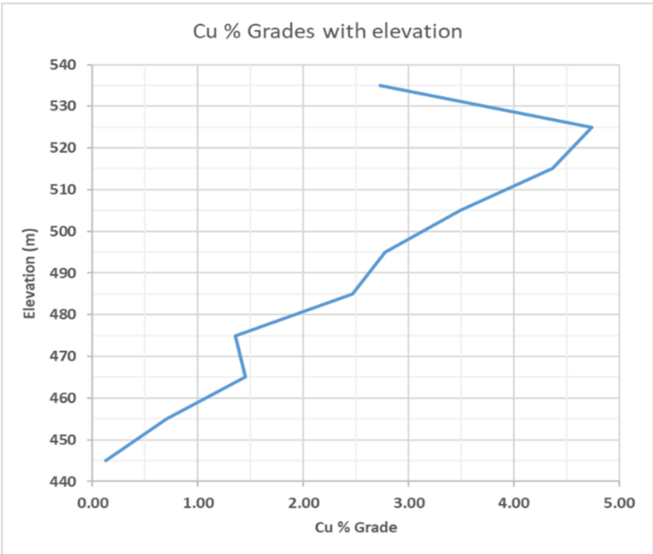


Figure 4.8: Plot of elevation (m) against copper grades (% Cu)

4.4.1 Domain definition

A cut-off of 0.5% Cu was used to separate the dataset, based on the observations from the sample histogram. Two contiguous subdomains were defined inside the vein – a high-grade (HG) domain and a low-grade (LG) domain – as depicted in Figure 4.9.

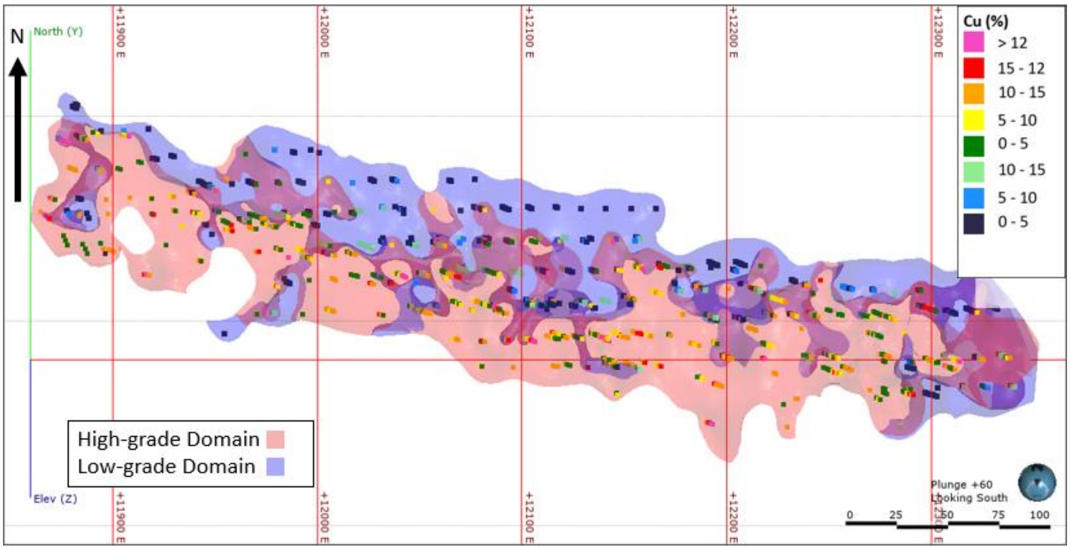


Figure 4.9: HG and LG domains modelled inside Vein 16 – view is orientated looking south along a plunge of 60°

Summary statistics for the two domains inside Vein 16 are shown in Table 4.4. The HG domain has a mean sample grade of 3.66% Cu, while the LG domain has an average sample grade of 0.16% Cu. The separation into two domains has lowered the CV, with the HG domain being more variable than the LG domain, as seen by the CV values of 1.11 and 0.87 respectively.

Table 4.4: % Cu summary statistics for the HG and LG domains

Domain	HG	LG
Number of samples	697	363
Minimum	0.02	0.001
Maximum	33.63	0.49
Range	33.61	0.489
Mean	3.66	0.16
Median	2.16	0.12
Mode	0.63	0.01
Variance	16.49	0.02
Standard Deviation	4.06	0.14
CV	1.11	0.87

Relative frequency histograms for the un-composited samples were plotted for the HG (Figure 4.10) and LG (Figure 4.11) domains. Both grade distributions were shown to be positively skewed, with some indications of possible further sub-domaining evident in the LG domain. However, it was determined that this would result in smaller, disconnected domains with too few samples for estimation purposes.

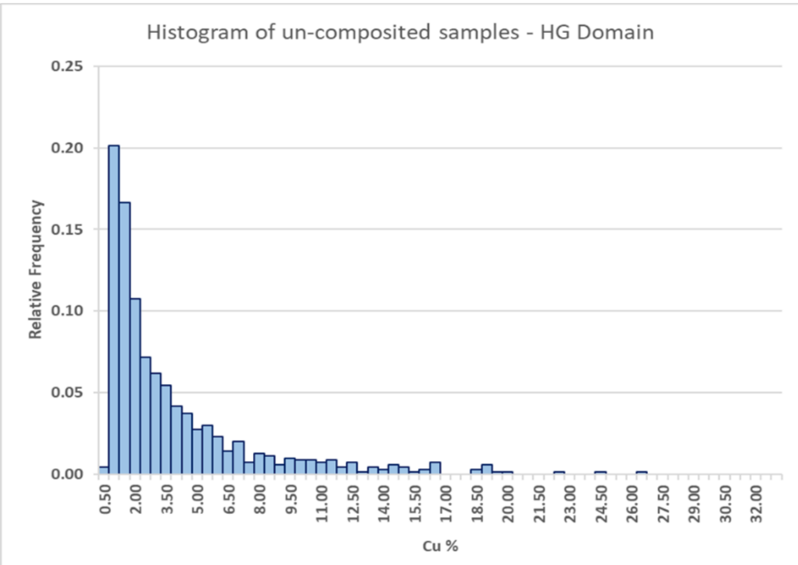


Figure 4.10: HG domain histogram for copper grades (% Cu)

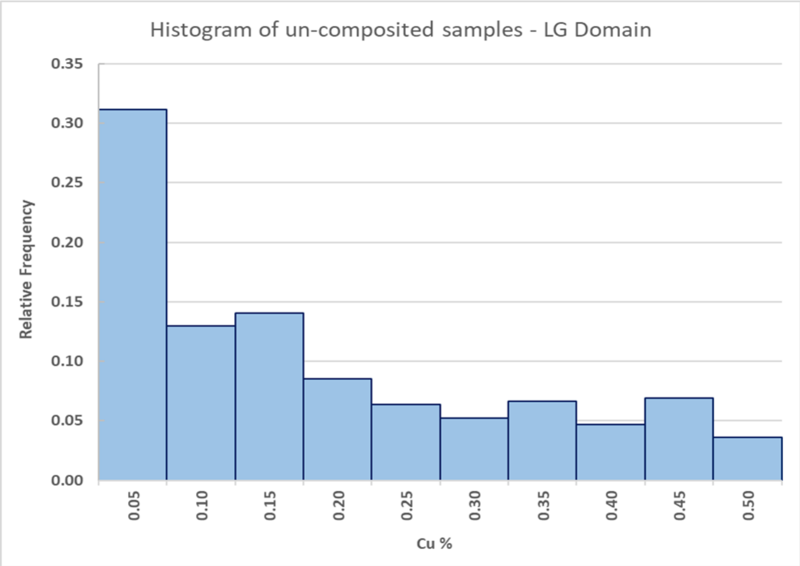


Figure 4.11: LG domain histogram for copper grades (% Cu)

4.4.2 Composite statistics

While samples tend to be drilled at specific lengths, the sampling procedure can result in variable lengths, as sampling techniques often follow geological and perceived mineralogical boundaries. In resource estimation, it is important to composite these as close as possible to the original interval, which ensures equal sample support and avoids introducing bias in the statistical analysis (Coombes, 2008).

An analysis of the sample lengths in both domains was done. It was found that most of the samples were collected in three-metre intervals – 87% for the HG domain and 91% for the LG domain.

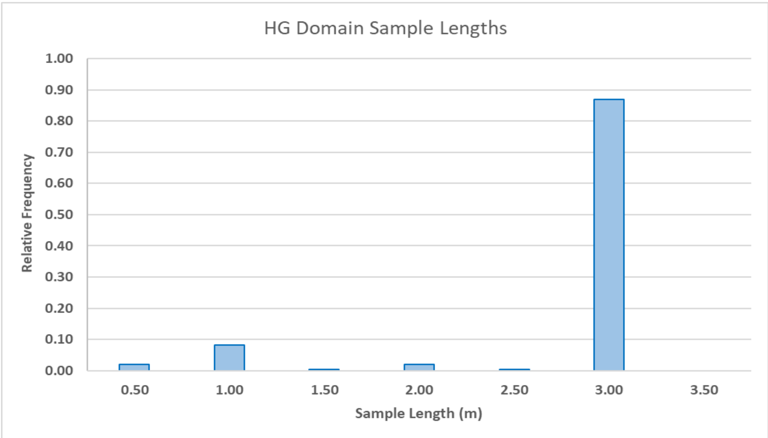


Figure 4.12: HG domain histogram of sample lengths (m)

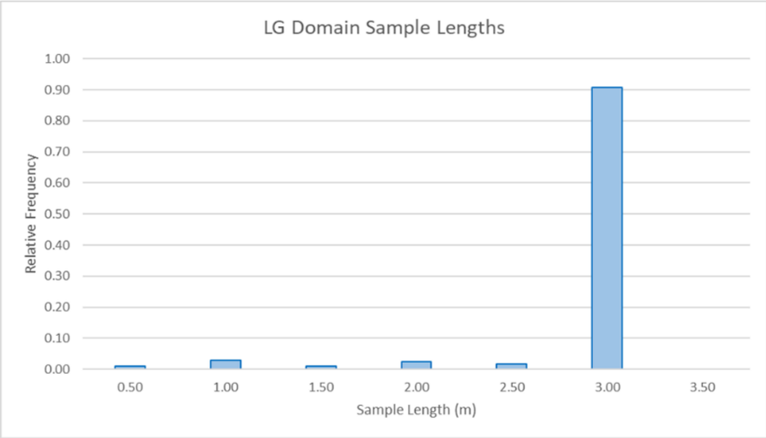


Figure 4.13: LG domain histogram of sample lengths (m)

Based on the sample length analysis, a three-metre composite length was used for both domains. When compositing, often some residual lengths remain at the bottom of the drillhole. Whenever present, these residual lengths were added to the last composite in a drillhole, resulting in some samples exceeding the three-metre compositing rule.

A statistical comparison between un-composited samples and composite outputs (

Table 4.5) shows that the compositing has had minimal impact on the mean sample grades for both domains, with a difference of 0.08% and 0.17% for the HG and LG domains respectively. Due to the change of support, the CV value of the composites is lower than their respective inputs, with the HG domain showing a more noticeable difference.

Table 4.5: Summary statistics of the un-composited and composite sample grades

% Cu	Un-composited samples (<3 m)		Composite (3 m)	
Domain	**HG**	**LG**	**HG**	**LG**
Number of samples	697	363	637	344
Total length (m)	1 911	1 027	1 911	1 024
Minimum	0.02	0.001	0.13	0.001
Maximum	33.63	0.49	27.23	0.49
Range	33.61	0.49	27.11	0.49
Mean	3.66	0.16	3.66	0.16
Variance	16.49	0.02	15.51	0.02
Standard Deviation	4.06	0.14	3.94	0.14
CV	1.11	0.88	1.08	0.88

Cumulative histograms comparing the distributions for the un-composited and composite samples illustrate that no changes have been introduced to either the HG domain (Figure 4.14) or the LG domain (Figure 4.15) due to compositing.

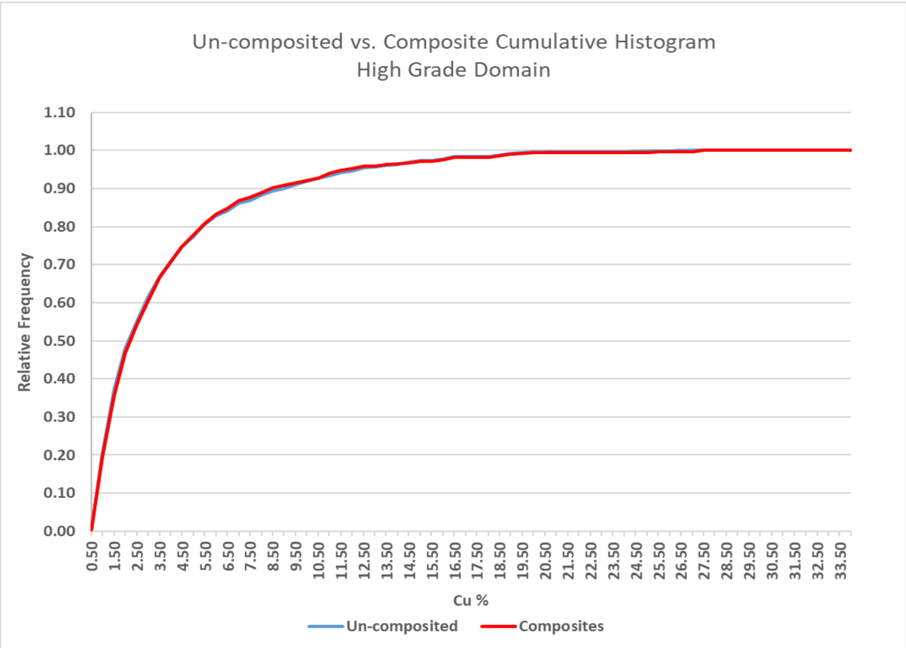


Figure 4.14: HG domain – Cumulative relative frequency distributions of un-composited and 3 m composite (% Cu) sample grades

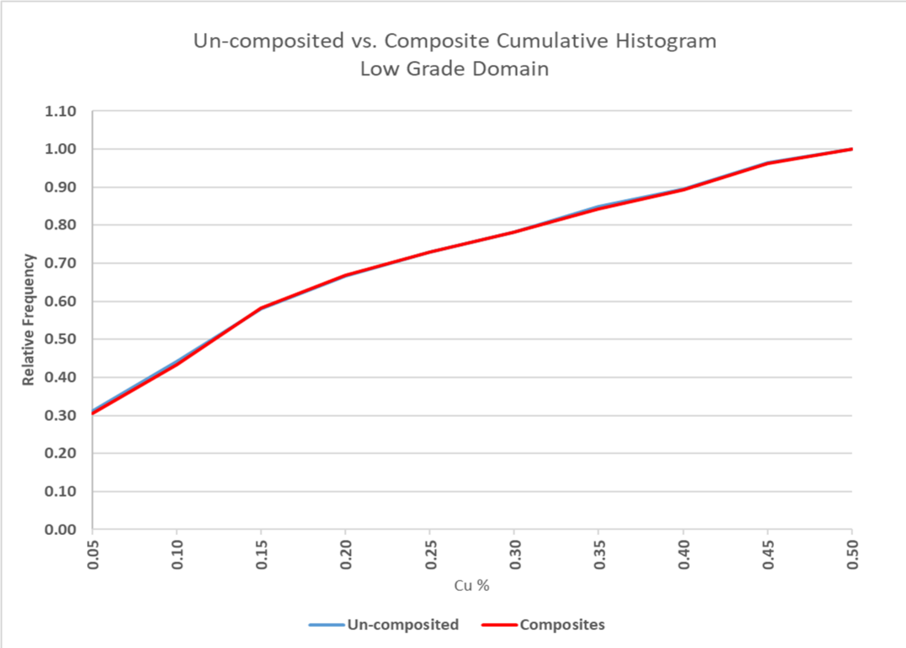


Figure 4.15: LG domain – Cumulative relative frequency distributions of un-composited and 3 m composite (% Cu) sample grades

4.5 Sequential Gaussian Simulation

The usefulness of a model is gauged by how close it is to predicting reality. In the mining industry, a model representing the orebody is often constructed from data representing a small subset of the total volume of the deposit. This model is then reconciled against the grade and tonnages reported by the processing plant, which tend to be the reality upon which the resource model is measured against.

Production and processing data are considered sensitive information, and thus were not made available for this research. Consequently, a representation of reality was needed that replicated the distribution of the samples, honouring the heterogeneity of the drillhole data available, while maintaining its spatial variability. For this purpose, SGS was used to create an artificial representation of a reality upon which estimates can be compared against. SGS is a stochastic simulation technique that reproduces the histogram of the input data, closely replicating the maximum and minimum values, the mean, the variance, and the variogram model (Rossi & Deutsch, 2014).

4.5.1 Normal score transformation

The first step in a simulation is to transform the input data into a Gaussian distribution via a normal score transformation. The result is a dataset with a mean of zero and a variance value of one. Normal score transformation is applied when data distribution is strongly skewed, which tends to produce noisy, unstable semivariograms (Abzalov, 2016). A normal score transformation was done for the data in each domain – descriptive statistics for the transformed data are shown in Table 4.6 and their respective histograms are depicted in Figure 4.16.

Table 4.6: Summary statistics of the normal score transformed data for each domain

Domain	HG	LG
Number of samples	637	344
Minimum	-3.16	-2.98
Maximum	3.16	2.98
Range	6.32	5.95
Mean	0.00	0.00
Median	0.00	0.00
Variance	1.00	1.00
Standard Deviation	1.00	1.00

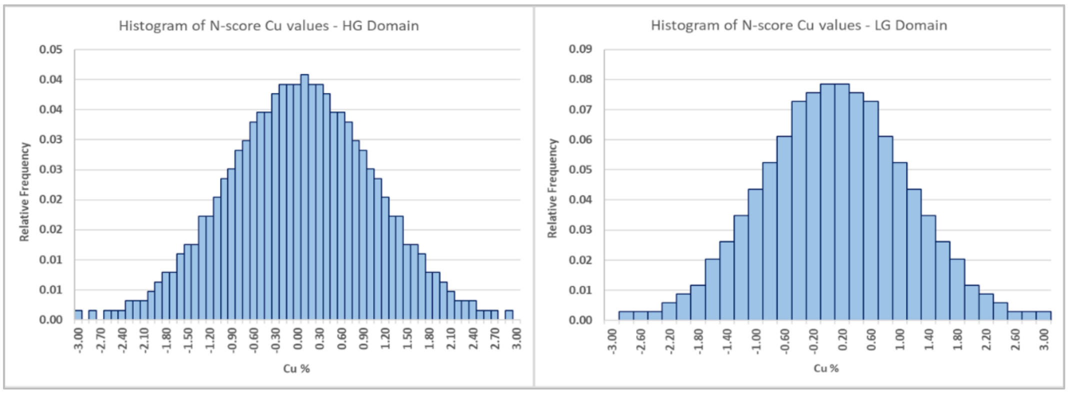


Figure 4.16: Normal score relative frequency distributions of the HG domain (left) and the LG domain (right)

4.5.2 Normal score variograms

Experimental semivariograms were calculated using the transformed data along the three major axis of continuity, aligned along the strike and dip of the vein, with a plunge component identified using a variogram map (referred to as a Radial Plot in EDGE®). The nugget effect was fixed by downhole variograms for each domain and double structure, nested models were fitted to the experimental points (Figure 4.17 and Figure 4.18 for the HG and LG domains respectively) with the resultant variogram model parameters shown in Table 4.7.

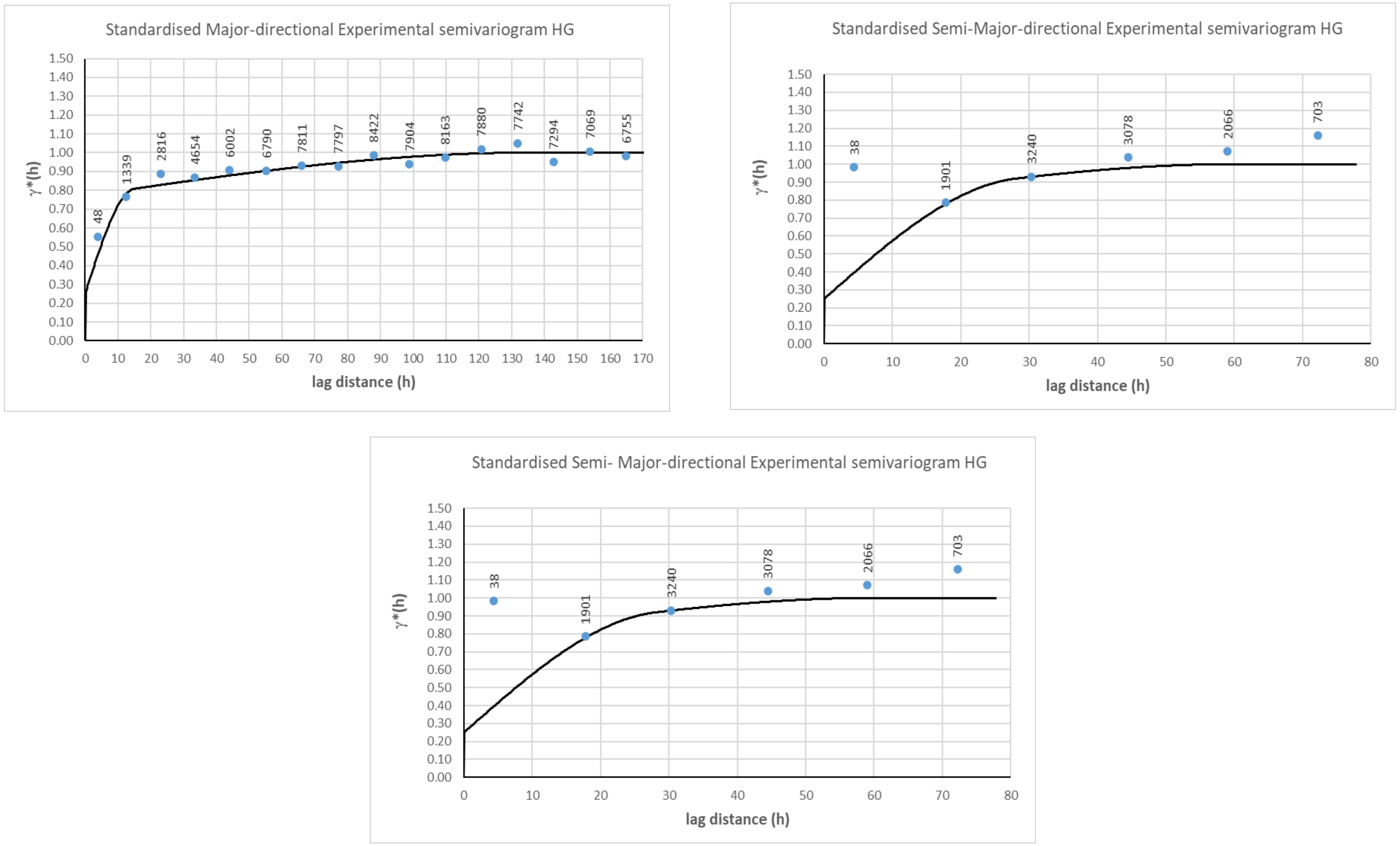


Figure 4.17: HG domain – normal score experimental and model semivariograms for the three major directions

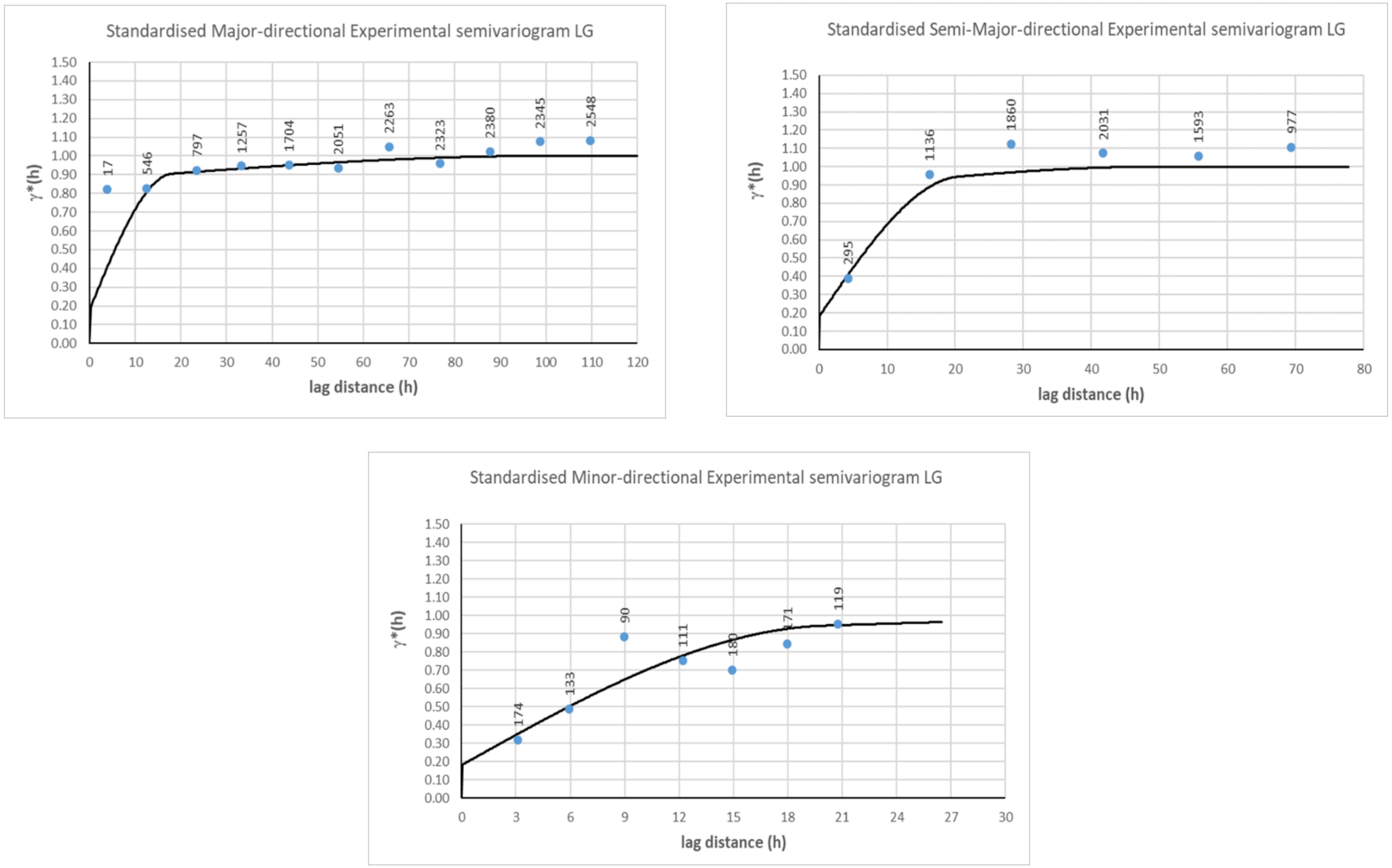


Figure 4.18: LG domain – normal score experimental and model semivariograms for the three major directions

Table 4.7: Normal score semivariogram model parameters for the HG and LG domains

Domain	Nugget Effect	Structure 1				Structure 2			
		Sill 1	Major	Semi-major	Minor	Sill 2	Major	Semi-major	Minor
HG	0.25	0.52	15	28	9	0.23	135	60	30
LG	0.18	0.69	18	20.5	20	0.13	100	50	50

All simulation work was done using Datamine Studio RM®. Datamine® is geological modelling and resource estimation software widely used in the mining industry. The SGS algorithm in Datamine® is based on the GSLIB code and the "SGSIM" process (Deutsch & Journel, 1997), which requires a set of input parameters, including variogram and search parameter files.

The data for each domain underwent QKNA using SoR, KV, and KE as metrics for optimisation. QKNA was utilised to determine the optimal search ranges, minimum and maximum number of samples, and maximum number of samples per drillhole to be used in the search neighbourhood. Vann et al. (2003) underlined the importance of QKNA as a preceding step for any kriging or conditional simulation to minimise conditional bias in estimates. Conditional bias is inherent to OK (Rossi & Deutsch, 2014) and is often observed as an overestimation in LG areas and an underestimation in HG areas (Glacken & Snowden, 2001).

QKNA tests were conducted by testing blocks in optimal and sub-optimal positions to assess the impact of data coverage inside the search neighbourhood on the estimation outputs. The results of the QKNA exercise are shown in Table 4.8 and Table 4.9.

Table 4.8: Search parameters used for the simulation of the HG domain

Search distances			Datamine search angles			Number of samples	
X	Y	Z	Z	X	Y	Minimum	Maximum
210	90	45	0.43	63.25	-13.07	4	30

Table 4.9: Search parameters used for the simulation of the LG domain

Search distances			Datamine search angles			Number of samples	
X	Y	Z	Z	X	Y	Minimum	Maximum
200	100	100	0.65	54.14	-11.72	4	30

These parameters were used to set up a macro automating the SGS process in Datamine®. Due to differences in the rotation conventions, the Dip, Dip Azimuth, and Pitch rotation settings used in Leapfrog® had to be converted to the Datamine® ZXY convention.

4.5.3 Simulation

The methodology employed in the SGS was guided by step-by-step procedures found in the work of Nowak and Verly (2005) and Rossi and Deutsch (2014). A grid of nodes was defined, spaced 5 m x 5 m in the X and Y coordinates and 1.5 m in the Z coordinate, ensuring that the vein was covered fully by nodes. The grid configuration used for both domains is shown in Table 4.10.

Table 4.10: Configuration of nodes used for simulation

Minimum coordinates			Node spacing			Number of nodes		
X	Y	Z	X	Y	Z	X	Y	Z
11 840	22 180	410	5	5	1.5	112	64	100

Special attention was given to how the tails of the distributions were back-transformed in the simulated outputs. This is because stochastic simulations are data expansive, generating many more data values than input points, hence the maximum resolution for the cumulative distribution function is considered insufficient for the tails (Deutsch & Journel, 1997).

Guided by Deutsch and Journel (1997), several parameters were tested and it was concluded that a power model for the upper tail of the HG domain was applicable. No back-transformations for the lower tail of the HG domain nor either tails of the LG domain were deemed necessary. The summary of the simulation inputs is tabulated in Table A.1 of APPENDIX A.

Each simulation is considered an independent realisation, achieved by defining a random path, where each node is visited by the SGS algorithm. Using the input data within the search neighbourhood, an estimate through SK is done on a node-by-node basis that follows this random path. The kriging estimate and KV of the transformed data are utilised to define a normal distribution upon which a Monte Carlo simulation is used to randomly sample the cumulative distribution and draw a simulated value. This value is then placed at the node, becoming part of the conditioning data. The process then travels along the random path to the next node and is repeated until all nodes are estimated (Coombes, 2008).

Rossi and Deutsch (2014) opined that 20–50 simulations should be done to characterise the range of possible values. For this study, 100 equiprobable models were generated. A single simulation was selected randomly by drawing a number between one and 100, with the selected number representing the simulation used to represent the reality upon which all subsequent work will be carried out and results compared against.

4.5.4 Validation of outputs

Validation of the simulated outputs were done by selecting the nodes inside the domain wireframes and comparing these to their respective sample composite inputs. Figure 4.19 shows the simulated 5 m x 5 m x 1.5 m nodes selected within the HG volume.

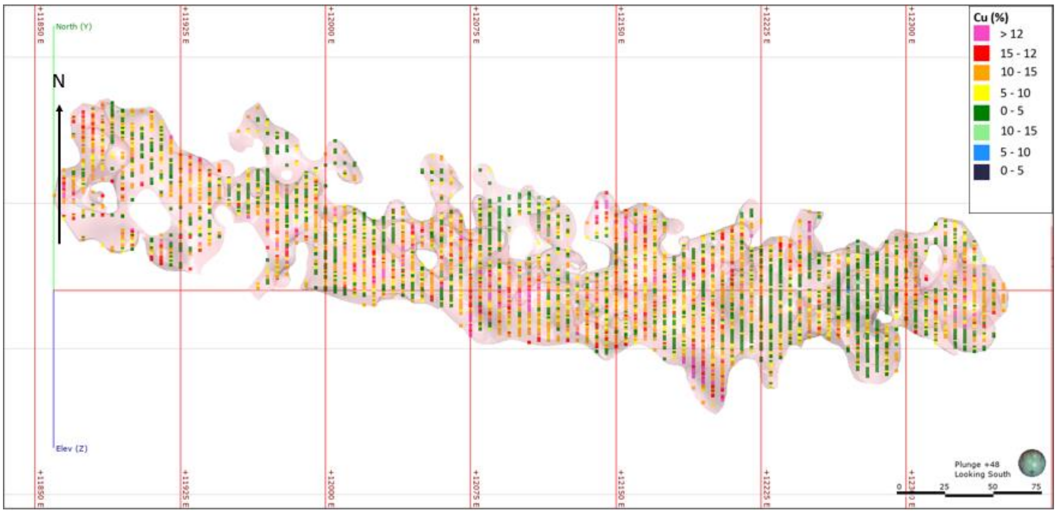


Figure 4.19: HG domain – Simulated nodes (5 m x 5 m x 1.5 m)

The same procedure was used to select the simulated nodes inside the LG domain (refer to Figure 4.20).

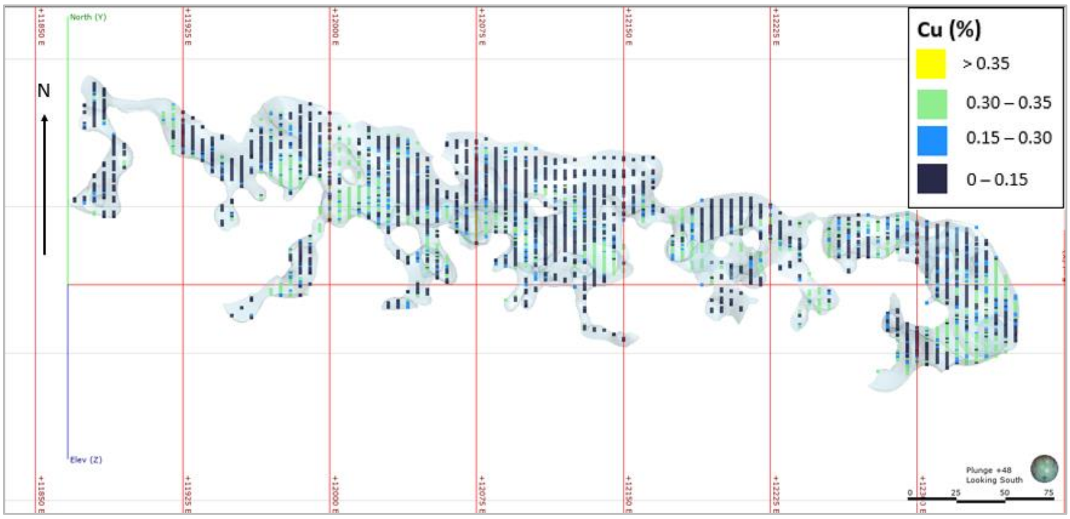


Figure 4.20: LG domain – Simulated nodes (5 m x 5 m x 1.5 m)

Summary statistics for the simulated nodes of both domains are shown in Table 4.11.

Table 4.11: Summary statistics (% Cu) of the simulated nodes

Domain	HG	LG
Number of nodes	6 734	3 935
Minimum	0.07	0.00
Maximum	27.59	0.5
Range	27.52	0.5
Mean	3.63	0.15
Median	2.28	0.11
Variance	14.97	0.02
Standard Deviation	3.87	0.14
CV	1.07	0.92

Statistically, the simulated outputs compare favourably with the sample composite inputs. The simulated mean for the HG domain shows a small relative difference of 0.86% compared against the composites, while the CV values are identical. For the LG domain, the relative percentage difference between means is larger at 5.19%, although this equates to a difference of 0.01% Cu, with a slight difference being noted in the CV values.

The cumulative frequency distributions suggest that the simulation has succeeded in replicating the heterogeneity of the composite data (Figure 4.21). Smaller deviations are noted for the LG domain, with the SGS outputs slightly overestimating the lower grades (Figure 4.22).

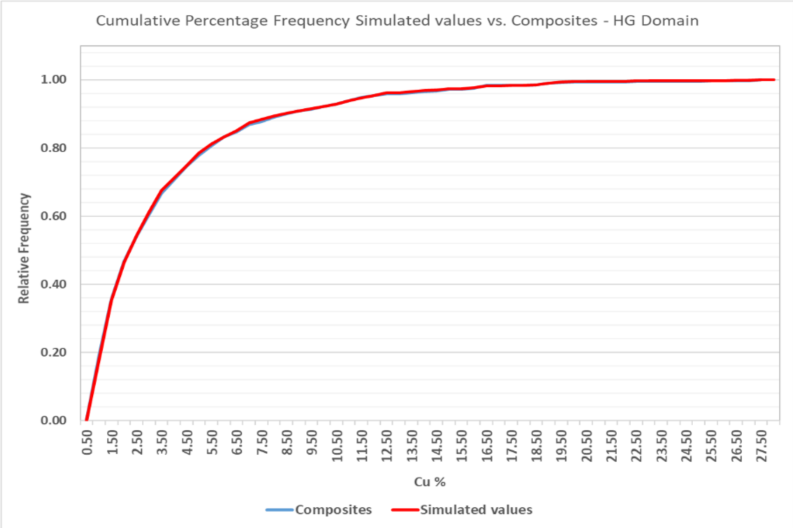


Figure 4.21: HG domain – % Cu cumulative relative frequency distributions of simulated and conditioning composite sample grades

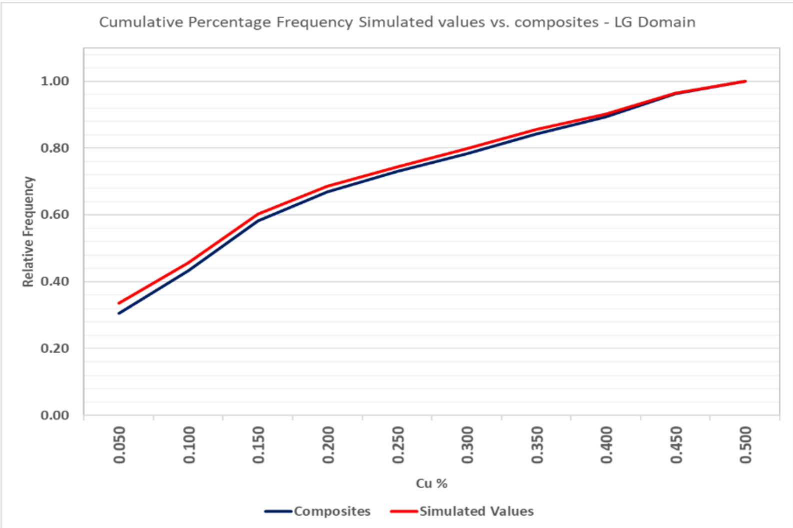


Figure 4.22: LG domain – % Cu cumulative relative frequency distributions of simulated and conditioning composite sample grades

4.6 Sampling Strategy

The research question for this project is based on the premise that sampling data are often sporadic and irregular. Therefore, a sampling strategy was devised on the simulated outputs that would replicate the reality of sparse data availability. The SGS outputs for each domain were sampled by manually selecting nodes inside their respective wireframe volumes, with the selection process beginning in the west and ending in the east. The resultant sampling grid resembles a realistic drilling grid, with samples spaced 15 metres apart in the X and Y directions and 1.5 metres in Z (refer to Figure 4.23 and Figure 4.24 for the HG and LG domains respectively).

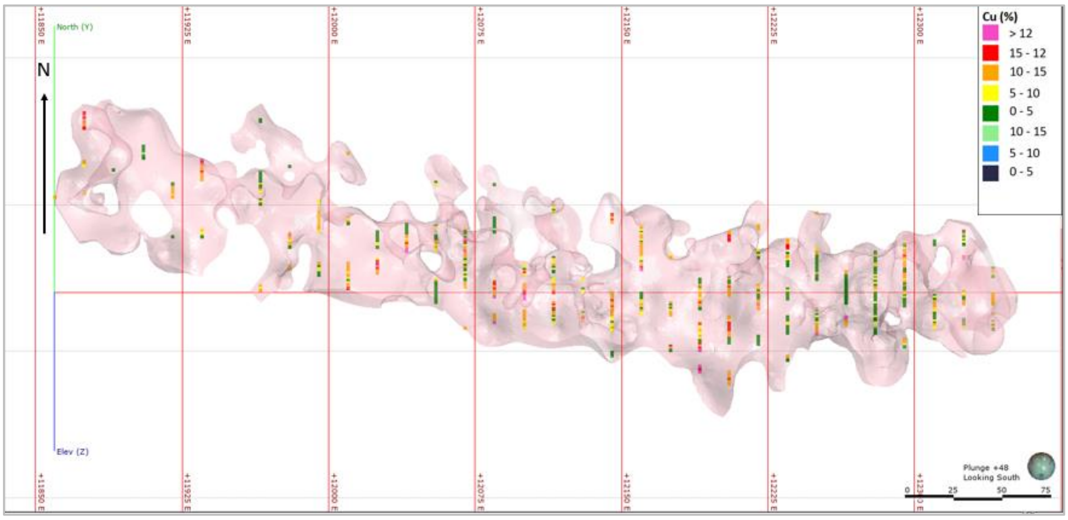


Figure 4.23: HG domain – the 15 m x 15 m x 1.5 m sample grid

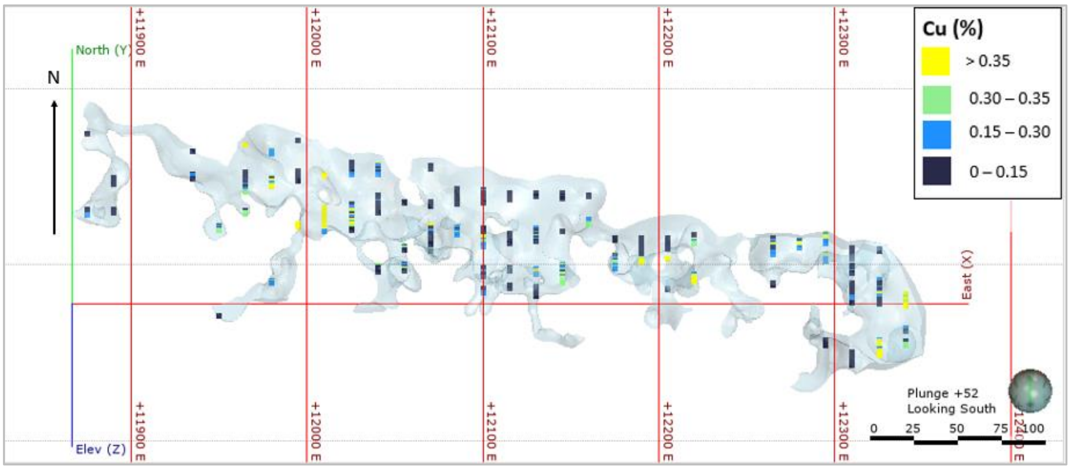


Figure 4.24: LG domain – the 15 m x 15 m x 1.5 m sample grid

This sampling strategy resulted in 704 samples inside the HG domain and 445 samples inside the LG domain, which were used for estimation purposes. The original 5 m x 5 m x 1.5 m grid will be the reality upon which results will be compared against.

4.7 Exploratory Data Analysis of Simulated Samples

As conducted previously for the conditional simulation, EDA was done on the data inputs, but this time on the samples from the simulation. Summary statistics are shown in Table 4.12, while the histograms are depicted in Figure 4.25 and Figure 4.26.

Table 4.12: % Cu summary statistics for samples for each domain

Domain	HG	LG
Number of samples	704	445
Minimum	0.382	0.001
Maximum	27.22	0.48
Range	26.84	0.48
Mean	3.66	0.16
Median	3.57	0.16
Variance	2.35	0.11
Standard Deviation	12.67	0.02
CV	3.56	0.15

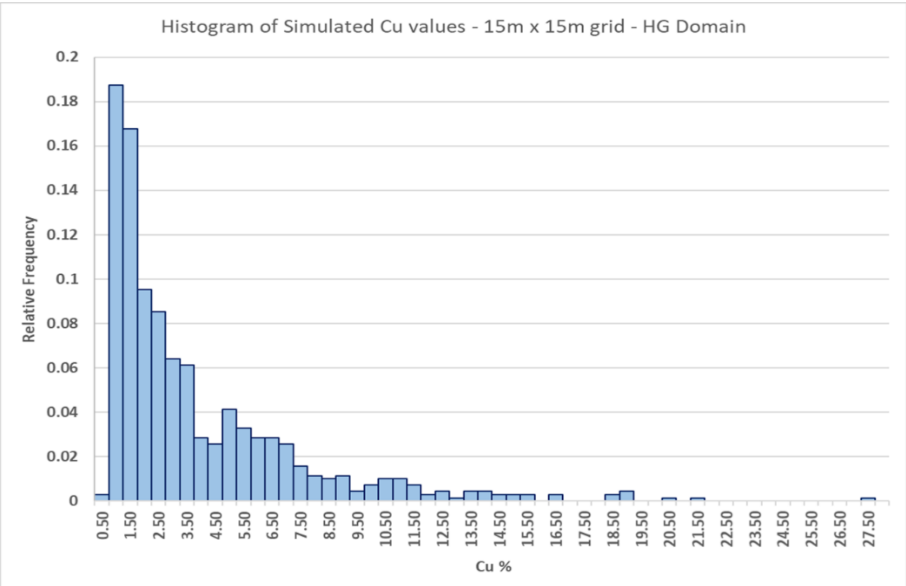


Figure 4.25: Histogram of the simulated samples in the HG domain

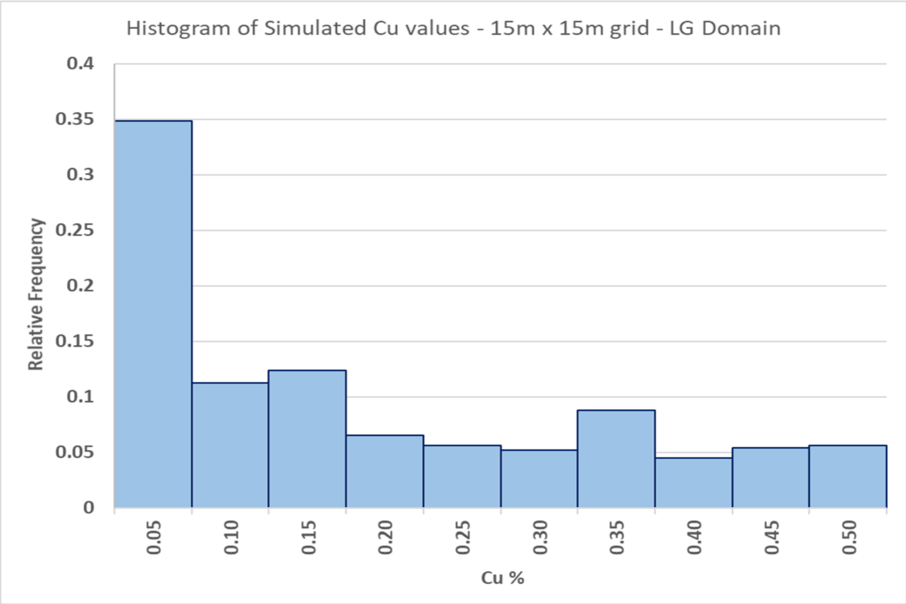


Figure 4.26: Histogram of the simulated samples in the LG domain

The statistical distributions for the simulated samples were observed to be similar to the sample composites. This indicates that the sampling strategy on the simulated nodes managed to capture the statistical nature of the original samples.

4.7.1 Variography

Experimental semivariograms were calculated for each domain. For simplicity purposes, variogram models were standardised, where the total population variance is rescaled to a value of one. The experimental semivariograms were calculated using Leapfrog EDGE® and the experimental data were then exported into Microsoft Excel and plotted. The models were fitted to these experimental points.

Downhole semivariograms are used to quantify the short-scale variability of the data, the cause of which can be due to a natural phenomenon occurring at a range shorter than the sample support (true nugget effect) or due to measurement errors from sampling and assaying (human nugget effect) (Carrasco, 2010). The nugget effect is modelled by extrapolating from the first two experimental points. The downhole semivariograms for the HG and

LG domains are shown in Figure 4.27 and Figure 4.28 respectively, the number of sample pairs is indicated next to experimental variogram points.

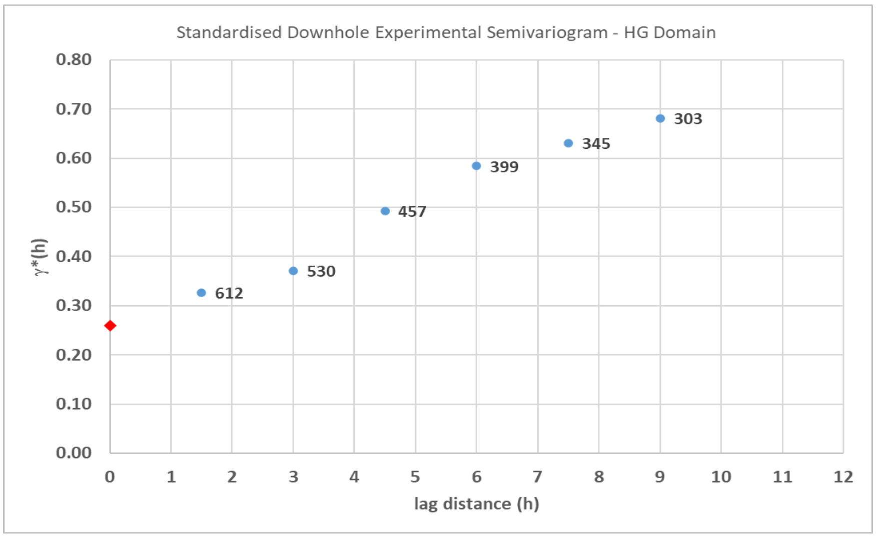


Figure 4.27: HG domain simulated (% Cu) samples – downhole experimental variogram and estimated nugget effect

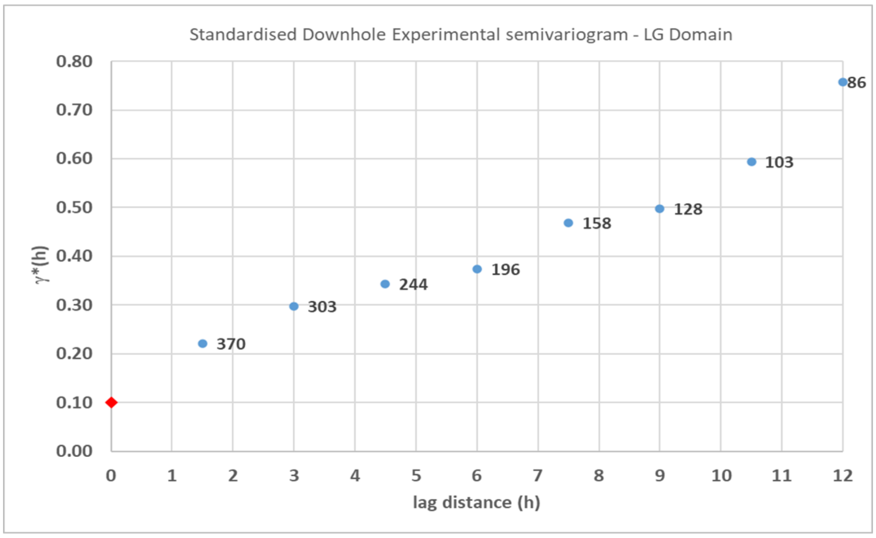


Figure 4.28: LG domain simulated (% Cu) samples – downhole experimental variogram and estimated nugget effect

Single-structured, directional semivariogram models were fitted to the experimental data using the parameters shown in Table 4.13, with directions given as plunge and trend. Structurally, the two variograms were similar, with the HG domain having a longer range along the major axis, while the semi-major and minor axes have similar ranges. The nugget values set the two apart, with the LG domain having a proportionally smaller modelled nugget.

Table 4.13: Semivariogram model parameters for both domains

Domain	Nugget Effect	Spherical model				Directions		
		Sill	Major	Semi-major	Minor	Major	Semi-major	Minor
HG	0.27	0.73	35	23	22	13→098	60→344	26→195
LG	0.10	0.90	23	23	23	15→296	51→045	35→195

The fitted models are illustrated in Figure 4.29 and Figure 4.30 for the HG and LG domains respectively. The LG domain proved challenging to model due to its high variability, as evidenced by the short range. This could be because of the presence of a smaller subdomain mentioned previously, which could not be modelled separately.

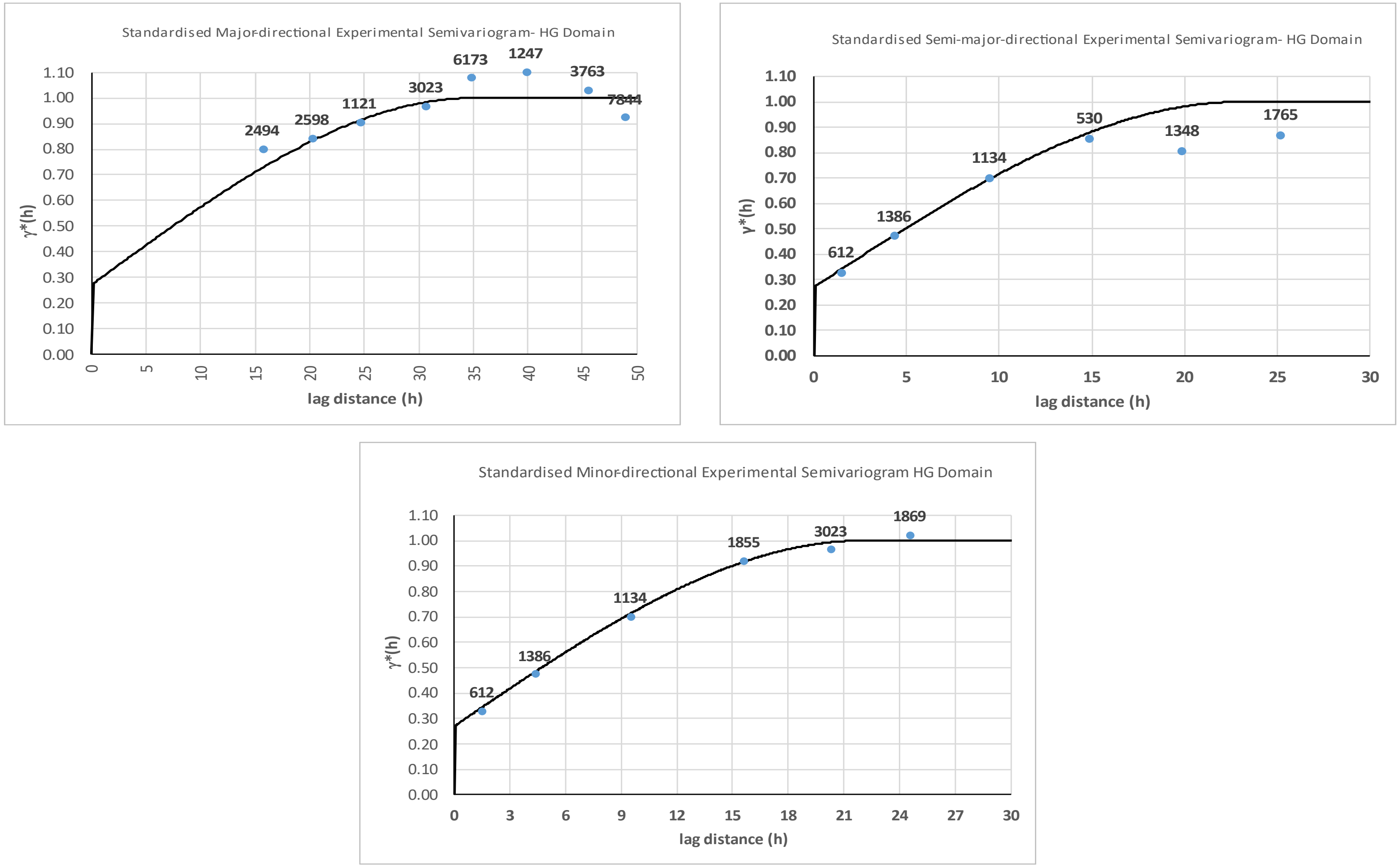


Figure 4.29: HG domain simulated samples – experimental and model semivariograms for the three major directions

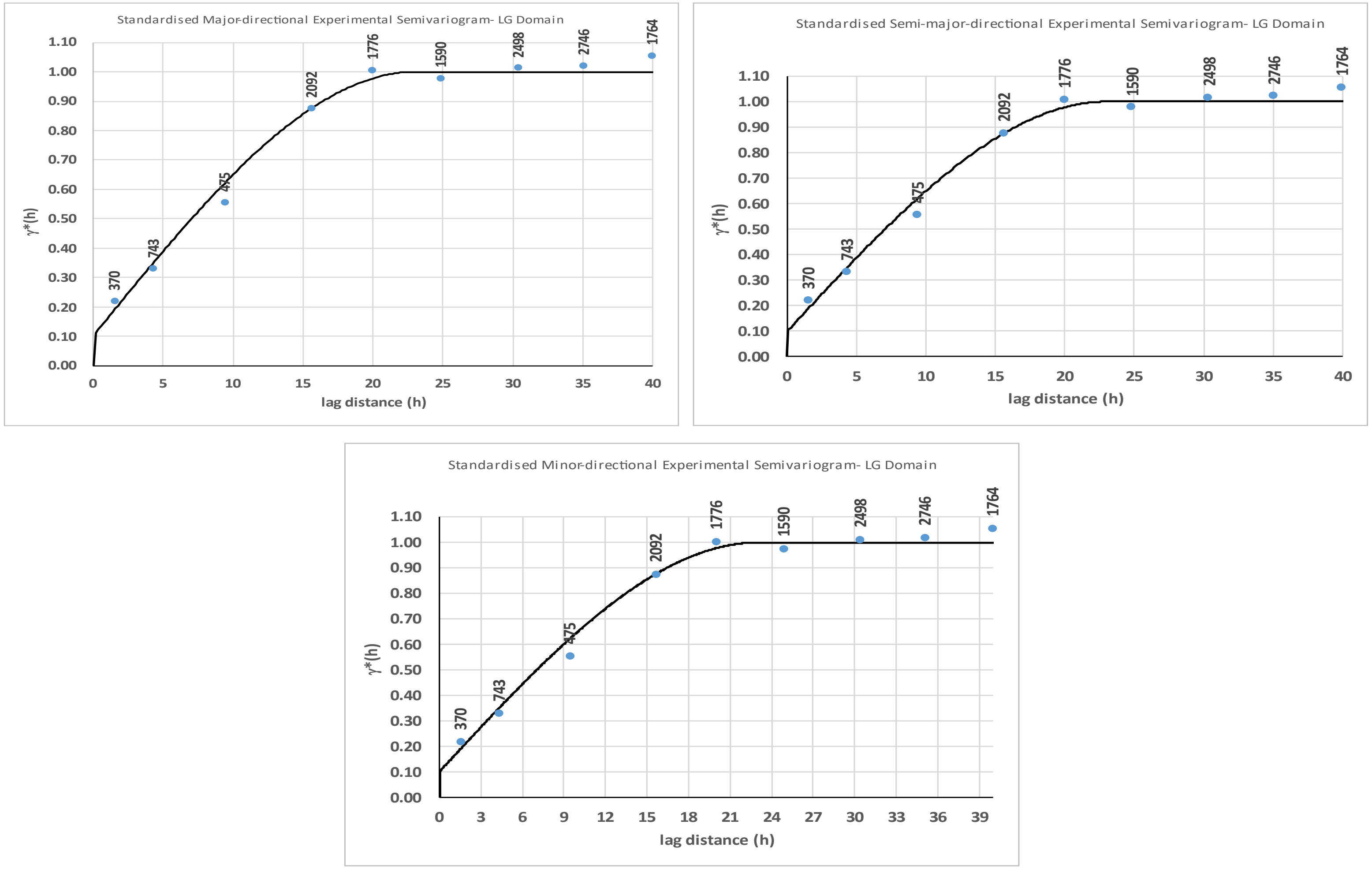


Figure 4.30: LG domain simulated samples – experimental and model semivariograms for the three major directions

4.7.2 Quantitative kriging neighbourhood analysis

A QKNA was done to optimise the search parameters as well as the minimum and maximum number of samples. A block size of 7.5 m x 7.5 m x 3 m in X, Y, and Z was chosen as half the sample node spacing in the XY-direction. QKNA tests were run on an optimal block position, where a sample is located on the block centroid, and a sub-optimal position, where no samples fall in the block.

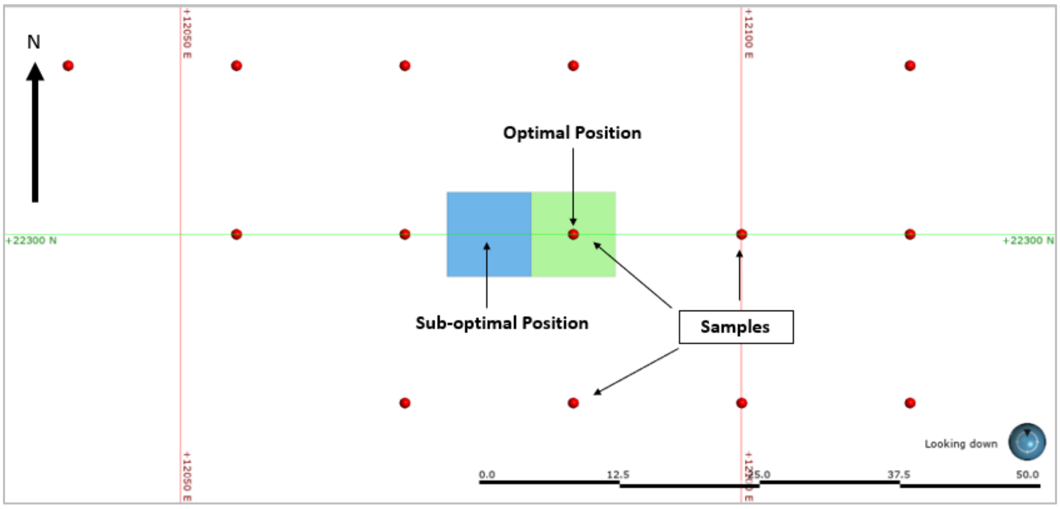


Figure 4.31: An example of an optimal position (green) and sub-optimal position (blue) block used in QKNA testing

Examples of the QKNA results are shown in the figures that follow, where the KV, KE, and SoR metrics of the block being kriged were assessed by incrementally increasing the search ranges, aligned to the directions of continuity (Figure 4.32), the maximum number of samples (Figure 4.33), and the maximum number of samples per drillhole (Figure 4.34). Since simulated nodes were used as inputs, a "drillhole" in this case is considered a single vertical string of nodes. This analysis is done with the objective of selecting the optimal kriging neighbourhood parameters that would reduce the conditional bias inherent in kriging – i.e., minimise the KV, and aim for KE and SoR values close to one. The point where no significant impact on the metrics is observed with changing the neighbourhood parameters is accepted as the optimal parameter. The chosen optimum parameter is indicated by a dashed line in the following figures. To fill the domains with estimated blocks, a search volume of double the variogram range was selected for each domain.

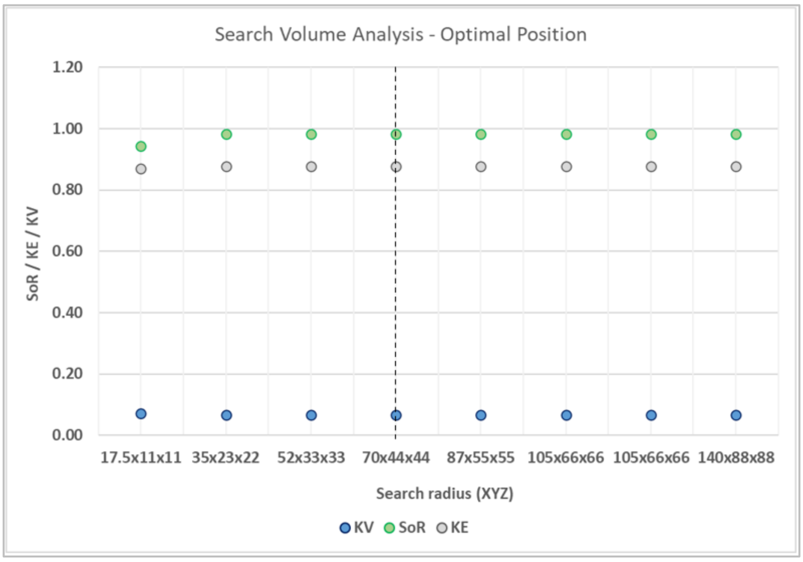


Figure 4.32: HG QKNA – SoR, KE, and KV results for search volume optimisation on an optimally positioned block

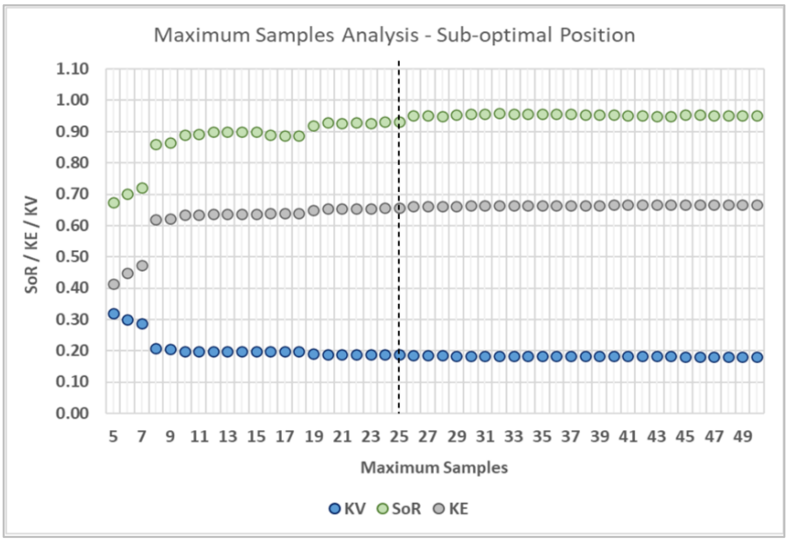


Figure 4.33: HG QKNA – SoR, KE, and KV results for optimising the maximum number of samples to be considered in the kriging neighbourhood on a sub-optimally positioned block

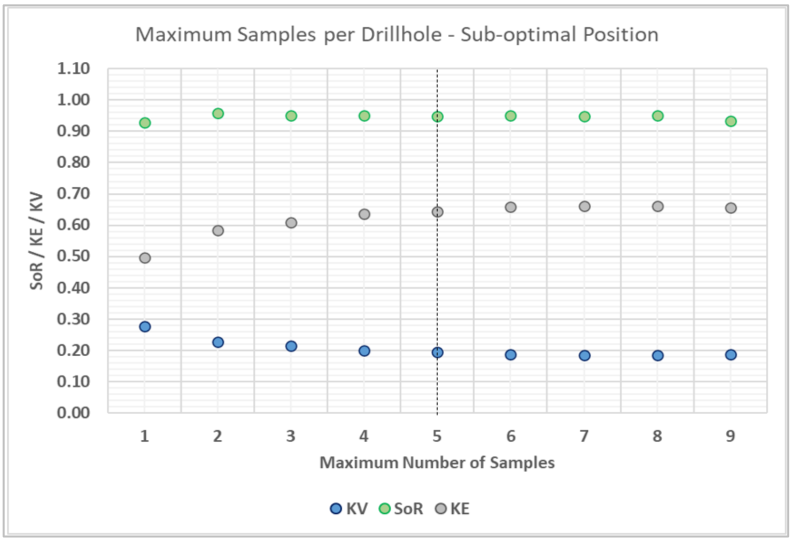


Figure 4.34: HG QKNA – SoR, KE, and KV results for the optimisation of the maximum number of samples per "drillhole" to be considered in the kriging neighbourhood on a sub-optimally positioned block

The optimised search parameters for both domains are shown in Table 4.14.

Table 4.14: Optimised search parameters

Domain	Ellipsoid ranges			Directions		
	Maximum	Intermediate	Minimum	Dip	Dip Azimuth	Pitch
HG	70	44	44	64	15	165
LG	46	46	46	55	15	18

The optimised parameters for the kriging neighbourhoods of the respective domains, minimum and maximum samples, maximum number of samples per "drillhole", and discretisation points are shown in Table 4.15.

Table 4.15: Optimised sample parameters and discretisation

Discretisation			Samples		Drillhole limit
X	Y	Z	Minimum	Maximum	
3	3	3	4	25	5
3	3	3	4	25	5

4.8 Grade Estimation

The grade estimations were done in Leapfrog EDGE®. The choice of estimators available are OK, SK, IDW, nearest-neighbour interpolation, and RBF.

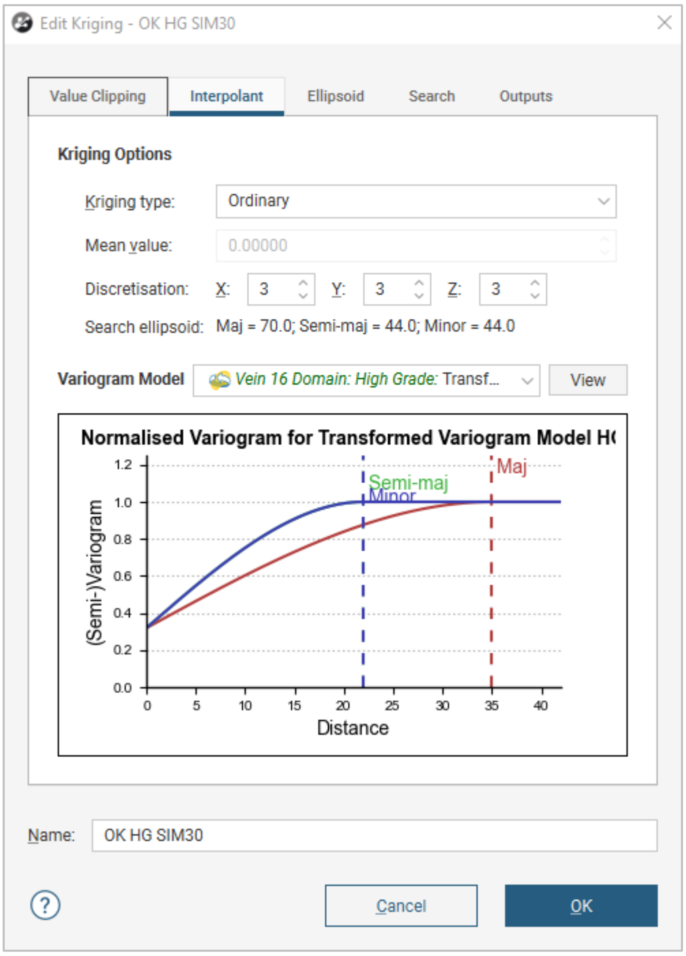


Figure 4.35: Example of a kriging interpolator setup in Leapfrog EDGE®

A block model was defined using a parent cell with the 7.5 m x 7.5 m x 3 m dimensions in the X, Y, and Z directions. Sub-celling was used to optimally fill the wireframes with blocks. Three sub-blocks in the X, Y, and Z directions were used, resulting in a minimum sub-block size of 2.5 m x 2.5 m x 1 m. Sub-blocking in EDGE® is achieved by activating triggers, which can be volumes or surfaces. Sub-blocking was triggered by adding the domain wireframes and a weathering model. Estimators were added to the block model using a drag-and-drop user interface. Thereafter, these estimators were evaluated within the parent cell.

4.8.1 Ordinary kriging

An estimator for OK was set up in EDGE® for the HG and LG domains using the semivariogram parameters of Table 4.13 and the search parameters from Table 4.14 and Table 4.15. The KE, KV, SoR, and the sum of negative weights were selected as output attributes to be added to the model. Using these attributes offers a means to assess the confidence and quality in an OK estimate. The OK estimators for both domains were evaluated within the block model parent cells.

4.8.2 Radial basis function

The RBF interpolator in EDGE® requires a spatial model to be defined, which can be a semivariogram model or a structural trend. When utilising a semivariogram, the software only allows the use of single-structure models – in this case, the same models used in the OK estimate were applicable. Alternatively, a spheroidal model can be applied but, for comparative purposes, the same model was retained for the OK and RBF. The structural trend method applies a user-defined mesh to define the anisotropic directions of the domain. This is used in conjunction with a model, where a nugget value, range and total sill parameters must be defined along with a drift component, which is required for both methods. The structural trend method was not utilised in this study.

Due to the use of a global drift where coefficients and sample weights are calculated for all samples in a domain, the RBF does not require a search neighbourhood to be optimised. Besides the choice of model, the only other parameter needed is the definition of a drift component, which informs how the estimates trend when data become sparse. The drift options available were: “Specified”, which allows a mean value to be defined; and “Automatic”, where the declustered mean is used instead. The Automatic parameter was used for this study. The RBF interpolant object was evaluated onto the same block model with the OK estimates.

4.8.3 Inverse distance weighting

The IDW estimator was set up using the same search neighbourhood as the OK. Moreover, the same minimum and maximum sample criteria and limits per drillhole were applied.

EDGE® offers various power models for the IDW method. The higher the power model, the more weighting is allocated to the nearest sample, eventually approximating a nearest-neighbour interpolation. Conversely, as a power approaches zero, the weights become more equally distributed to the samples in the neighbourhood and the estimate equates to the global mean. For this project, inverse distance squared (power value of 2, IDW2) was used.

No discernible difference in run time was noted for either of the three estimation methods, as the estimation domains and the number of samples within them are relatively small.

4.9 Validation

The validation of the estimates was achieved by comparing estimated outputs against their respective inputs using statistics, histograms, cumulative frequency plots, and swath plots.

4.10 Comparisons

The estimates for the three methods were compared against the simulated reality – the 5 m x 5 m x 1.5 m grid of simulated nodes – using descriptive statistics, histograms, and cumulative frequency distributions. Quantile-quantile (Q-Q) plots were used to assess how comparable the three estimates were. Swath plots were plotted to spatially identify the differences in grades between the three methods and to assess how the number of samples impacted the estimated outputs.

Unlike OK, the RBF method offers no means to measure confidence in the estimates. Consequently, comparisons were done against the OK estimates by looking at how close to the 99% confidence interval (CI) the RBF came to for each swath plot. The formula for calculating a lower and upper confidence limit (CL) is given by the formula below:

$$\mathrm{CL} = \bar{\mathrm{x}} \pm \mathrm{z}\ (s/\sqrt{n})$$

Where $\bar{x}$ is the sample mean; $z_\alpha = 2.576$ for a 99% confidence interval; s is the standard deviation; and n is the number of samples.

The CLs were calculated on the mean copper grades of the kriging estimates for each swath in the X, Y, and Z directions. A block estimate is always done on a parent cell, but when sub-celling is used, this estimated grade is partitioned into each daughter sub-cell inside the parent block. As a result, it was found that comparisons using the sub-celled model resulted in very narrow 99% CL ranges. The narrowness of these ranges was due to the variance of each swath being reduced artificially by the presence of many small sub-blocks with the same grades.

To make meaningful comparisons between the three estimation methods, the sub-celled blocks were filtered out. This allows for more optimal, relevant comparisons. Additionally, for representativity, any swath with less than five parent cells were excluded.

The 99% CLs were plotted on the swath plots against the three estimates. Observations considered cases where the RBF estimate fell inside these limits, relative to the OK estimates and its relation to the sample numbers available.

5 RESULTS

This chapter summarises the results of the estimates done on the two domains, using the methodology outlined in Chapter 4. Only results pertaining to the estimation of the two domains and comparisons between them are presented, with the discussion of the results reserved for Chapter 6.

5.1 Statistical Validation

The estimated outputs for the three methods were validated against the samples by comparing descriptive statistics, frequency distributions, and swath plots for each domain. Descriptive statistics for the HG and LG domain comparisons are shown in Table 5.1 and Table 5.2 respectively. The values in the tables are reported at three decimal figures for comparative purposes.

Table 5.1: HG domain % Cu summary statistics for samples, OK, RBF, and IDW2 estimates

Estimate	Min	Max	Range	Mean	Median	Variance	Std Dev	CV
Samples	0.382	27.220	26.837	3.572	2.353	12.675	3.558	0.996
OK	0.589	15.013	14.424	3.639	3.240	3.752	1.937	0.532
RBF	0.306	14.831	14.525	3.640	3.279	3.505	1.872	0.514
IDW2	0.382	15.231	14.848	3.599	3.176	3.502	1.871	0.520

Table 5.2: LG domain % Cu summary statistics for samples, OK, RBF, and IDW2 estimates

Estimate	Min	Max	Range	Mean	Median	Variance	Std Dev	CV
Samples	0.001	0.482	0.481	0.163	0.112	0.022	0.149	0.918
OK	0.013	0.449	0.437	0.153	0.144	0.009	0.094	0.612
RBF	0.009	0.438	0.430	0.153	0.141	0.007	0.086	0.564
IDW2	0.016	0.453	0.438	0.156	0.146	0.009	0.094	0.606

Cumulative frequency distributions comparing the samples to their respective estimates were plotted to identify how the estimates differ at increasing cut-offs from respective inputs (HG in Figure 5.1 and LG in Figure 5.2).

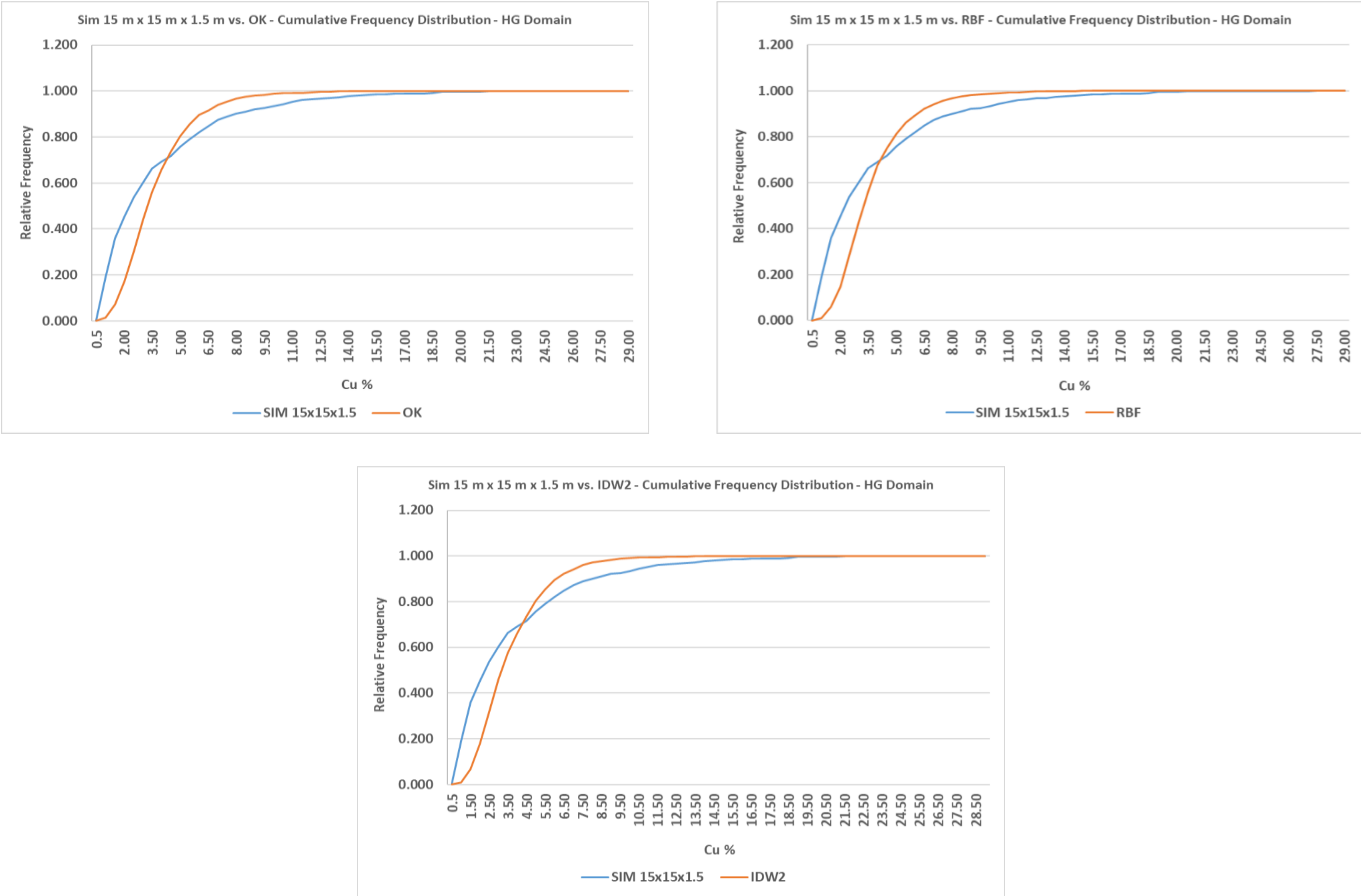


Figure 5.1: HG domain % Cu cumulative relative frequency distributions comparing the samples and the OK, RBF, and IDW2 estimates

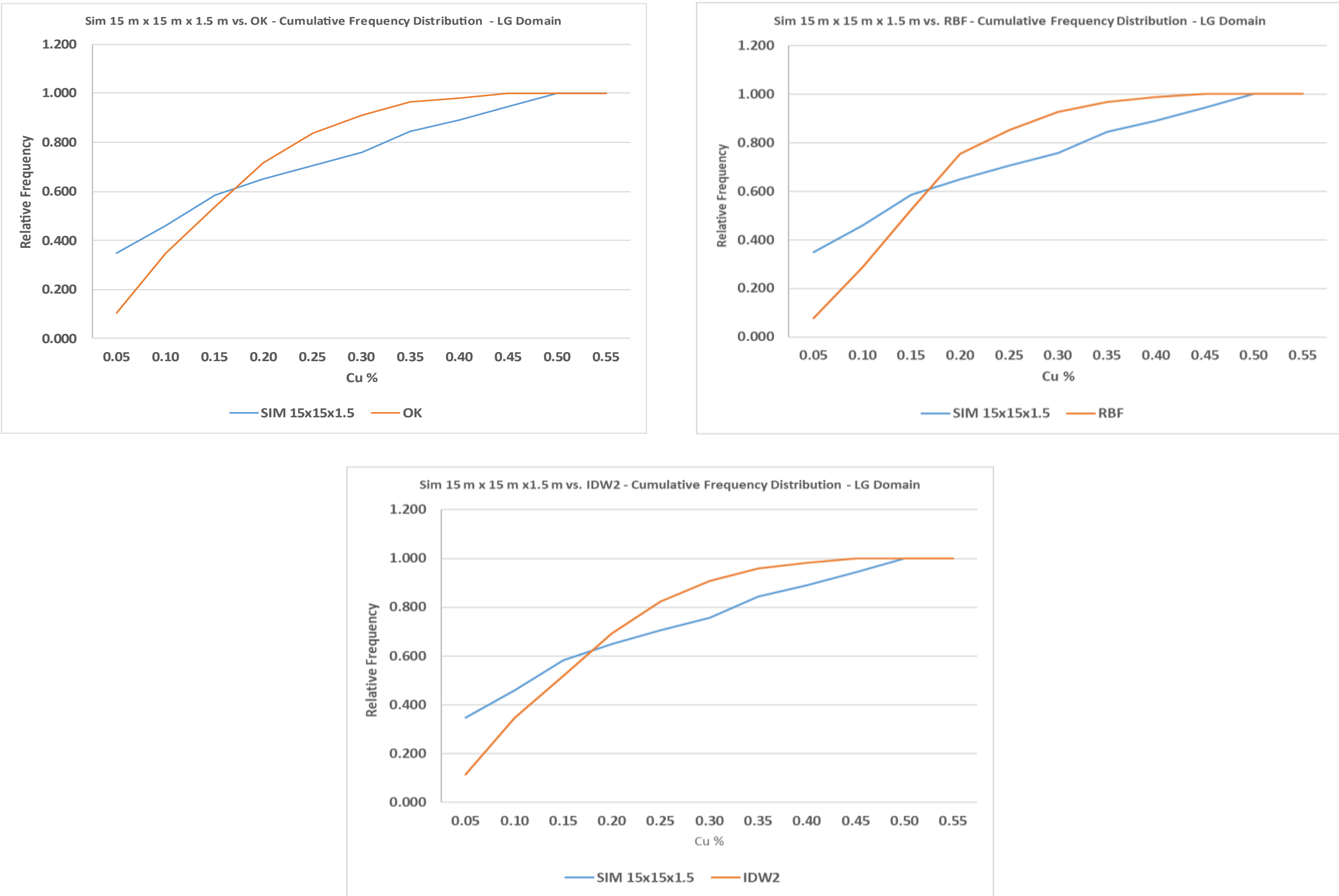


Figure 5.2: LG domain % Cu cumulative frequency distribution comparing the samples and the OK, RBF, and IDW2 estimates

5.1.1 Comparisons between methods

The cumulative frequency distributions for the three methods were plotted against one another for the HG (Figure 5.3) and LG (Figure 5.4) domains.

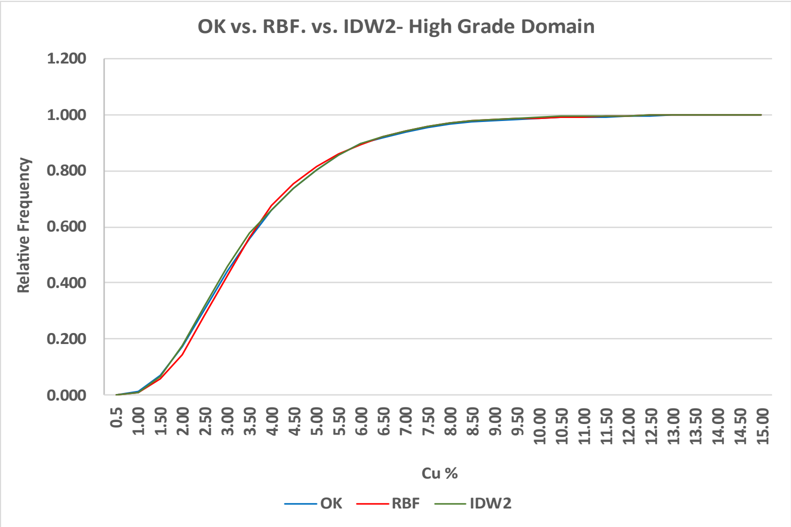


Figure 5.3: HG domain block estimate comparison, cumulative relative frequency distributions for OK (blue), RBF (red), and IDW2 (green)

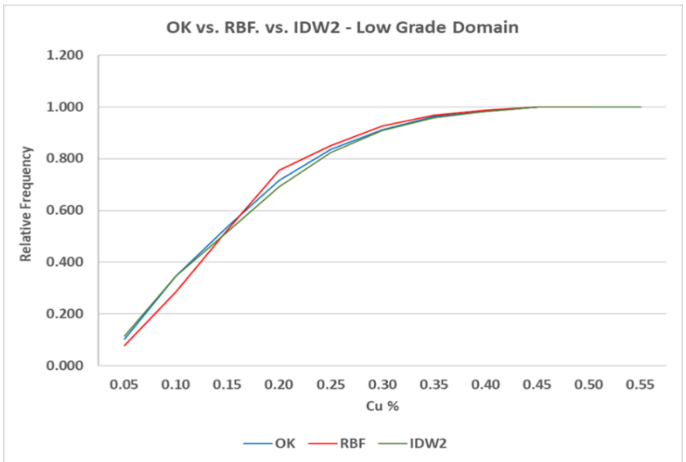


Figure 5.4: LG domain block estimate comparison, cumulative relative frequency distributions for OK (blue), RBF (red), and IDW2 (green)

5.1.2 Q-Q plots

Q-Q plots compare the probability distribution functions of two variables by plotting their respective quantiles against one another. The resultant plot is a line, where each quantile for the first dataset, plotted on the X-axis, corresponds to the same quantile for the second dataset, plotted on the Y-axis. If the resultant line plots on the X=Y line, known as the 45° line, then the distributions are considered the same and further departure from this line is evidence that the two distributions are not equal. The RBF estimates were plotted against the OK estimates in a Q-Q plot (Figure 5.5).

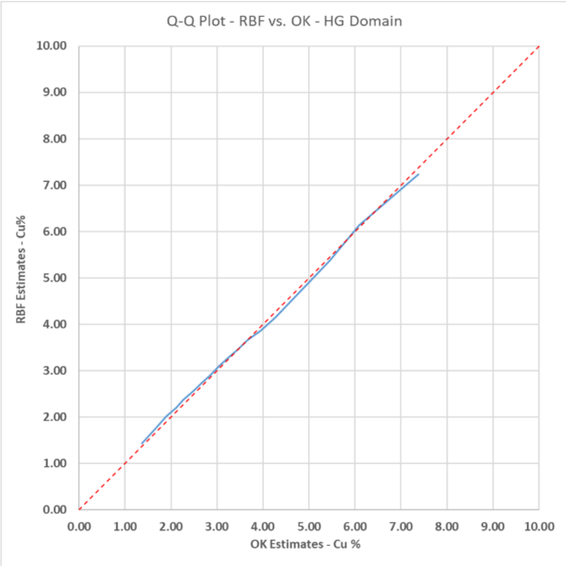


Figure 5.5: HG domain Q-Q plot comparing RBF against OK estimates

Similarly, Q-Q plots comparing the RBF and IDW2 estimates, and the OK and IDW2 estimates are shown in Figure 5.6.

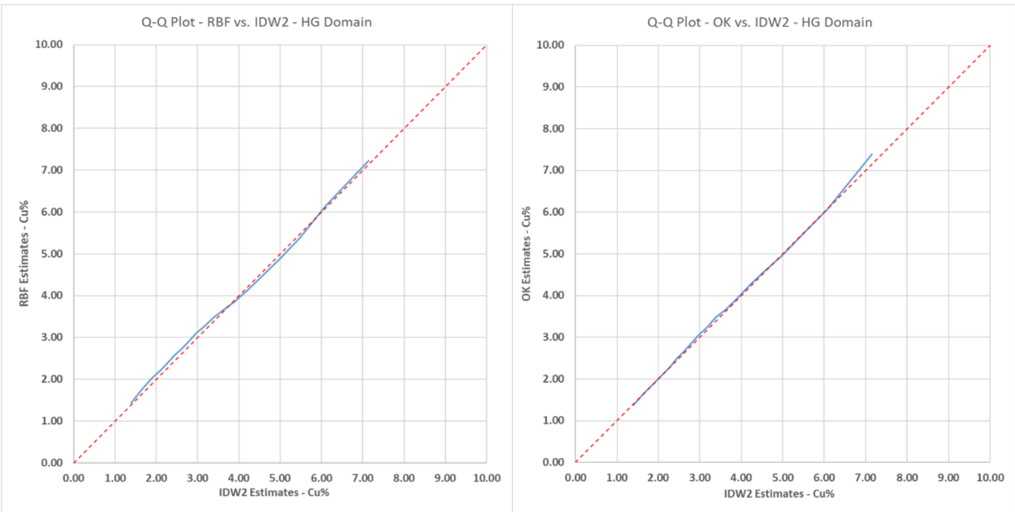


Figure 5.6: HG domain Q-Q plots comparing RBF (left) and OK (right) against IDW2 estimates

Moreover, Q-Q plots for the LG domain were generated. For the LG domain, a Q-Q plot showing RBF versus OK can be seen in Figure 5.7, while the Q-Q plots comparing the RBF and OK estimates with the IDW2 estimates are depicted in Figure 5.8.

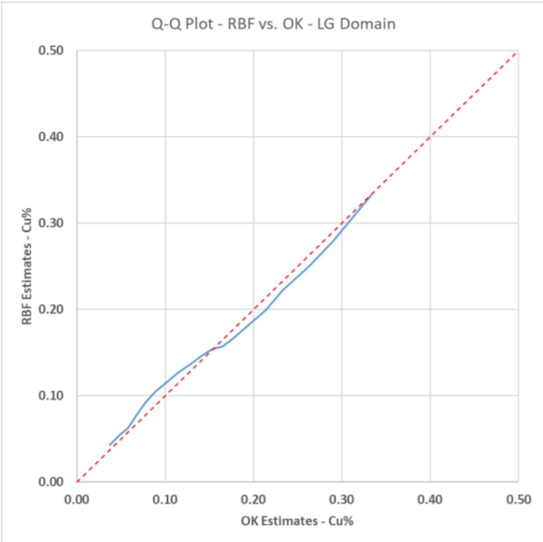


Figure 5.7: LG domain Q-Q plot comparing RBF against OK estimates

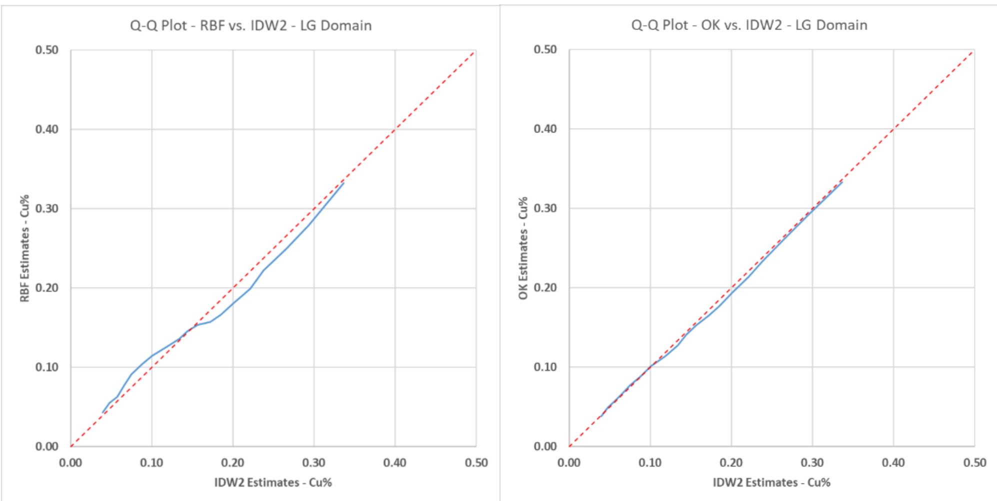


Figure 5.8: LG domain Q-Q plots, comparing RBF and OK against IDW2 estimates

5.1.3 Swath plot validation

Swath plot validations were generated for both domains (Figure 5.9 and Figure 5.10). Each graph plots the estimated grades for the three methods against the grades from the input samples. Sample numbers are displayed graphically as bar charts. The X-axis of the plots displays the swath numbers, which correspond to a band of a determined thickness, 15 m in the X-direction, 7.5 m in the Y-direction, and 3 m in the Z-direction. The extents of the swaths in the X-direction are 11 840–12 365E, in the Y-direction 22 232.5–22 405N, and in the Z-direction 442–560 elevation.

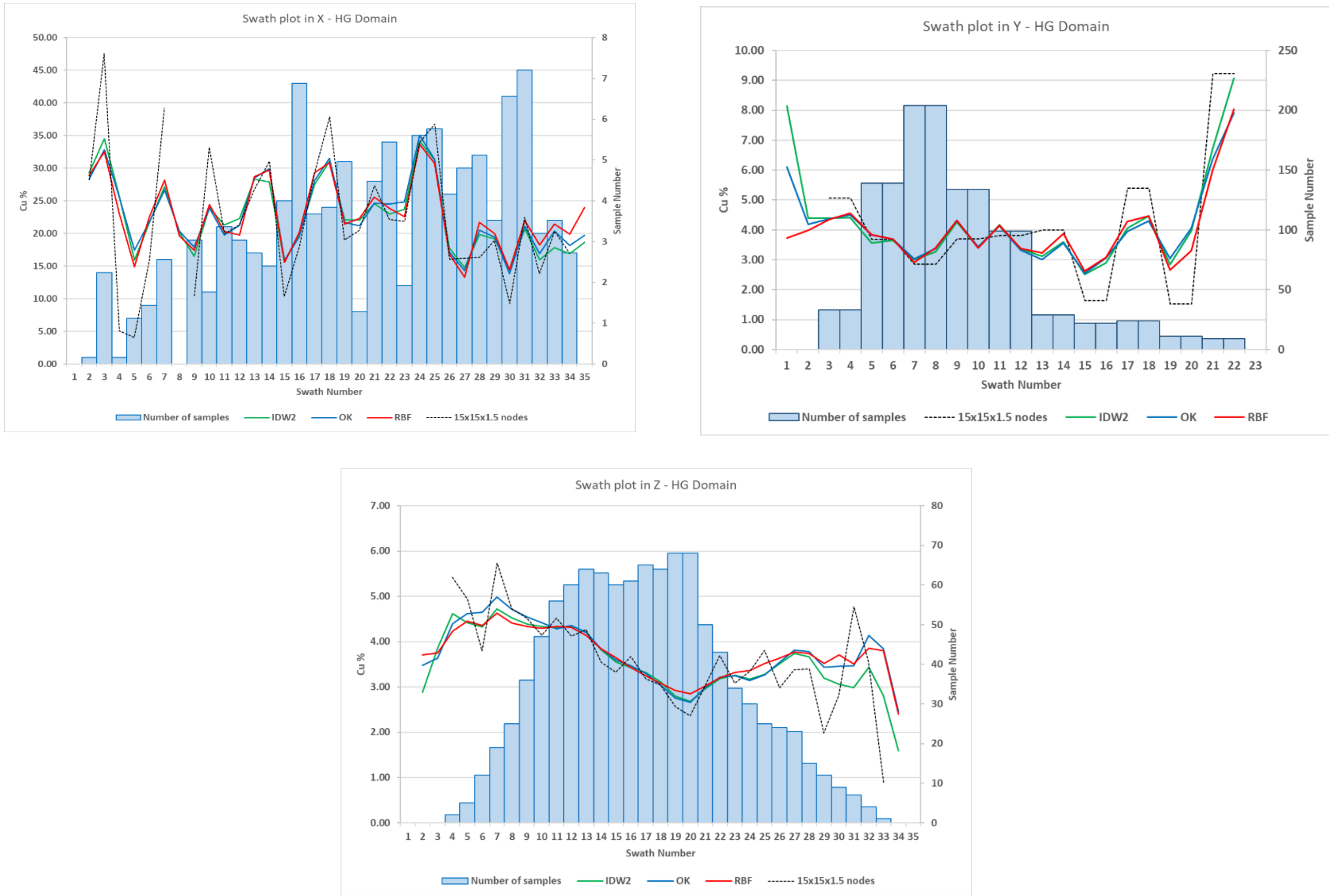


Figure 5.9: HG domain % Cu grade swath plots of the averages of simulated samples, IDW2, OK and RBF block estimates in the X, Y, and Z directions

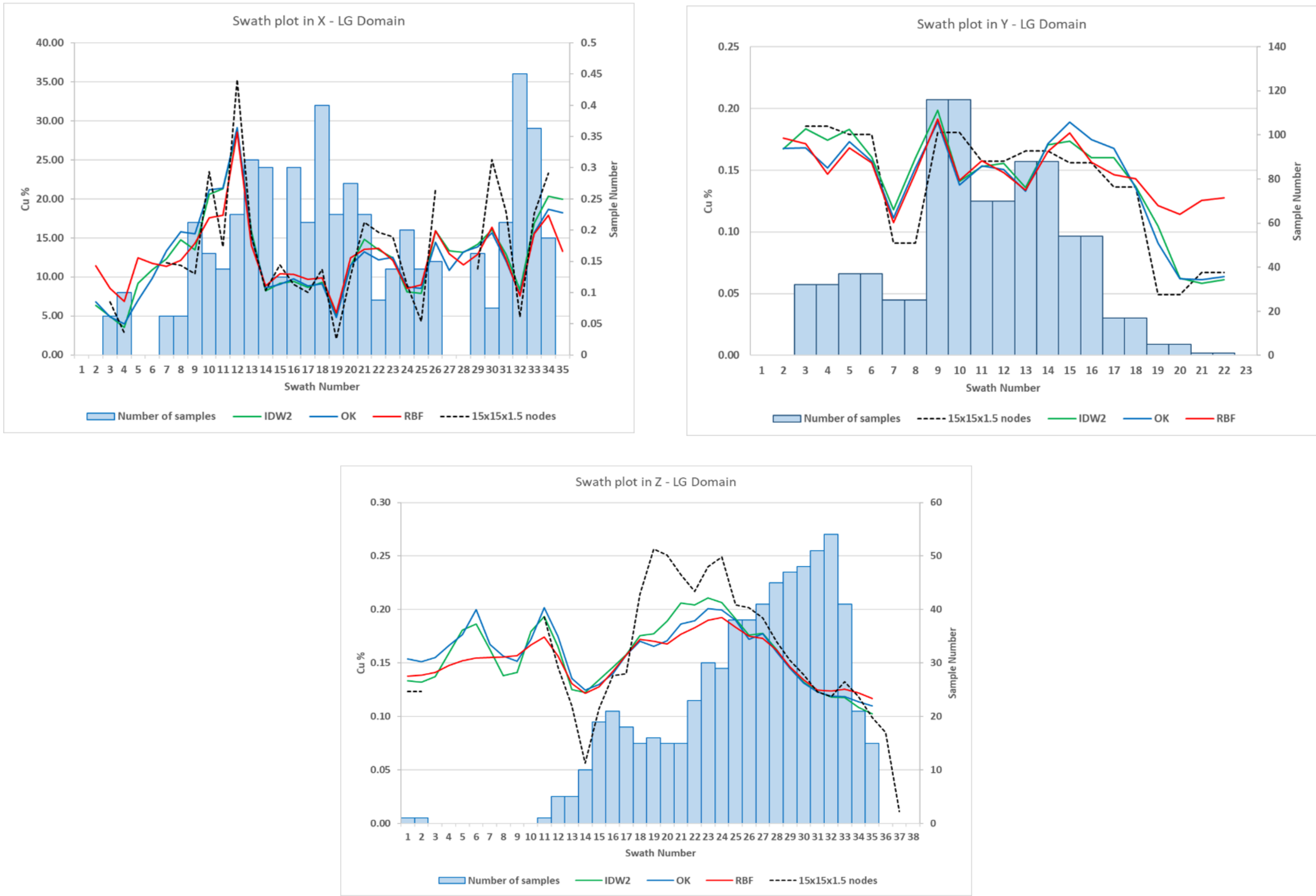


Figure 5.10: LG domain % Cu grade swath plots of the averages of simulated samples, IDW2, OK and RBF block estimates in the X, Y, and Z directions

5.1.4 Visual validations

Visual validations of the estimation outputs were done in 3D, in cross-section, and in plan-view. Three cross-sections are presented below, with each section displaying the RBF, OK, and IDW2 block estimates. The locations of these sections in plan-view are shown in Figure 5.11.

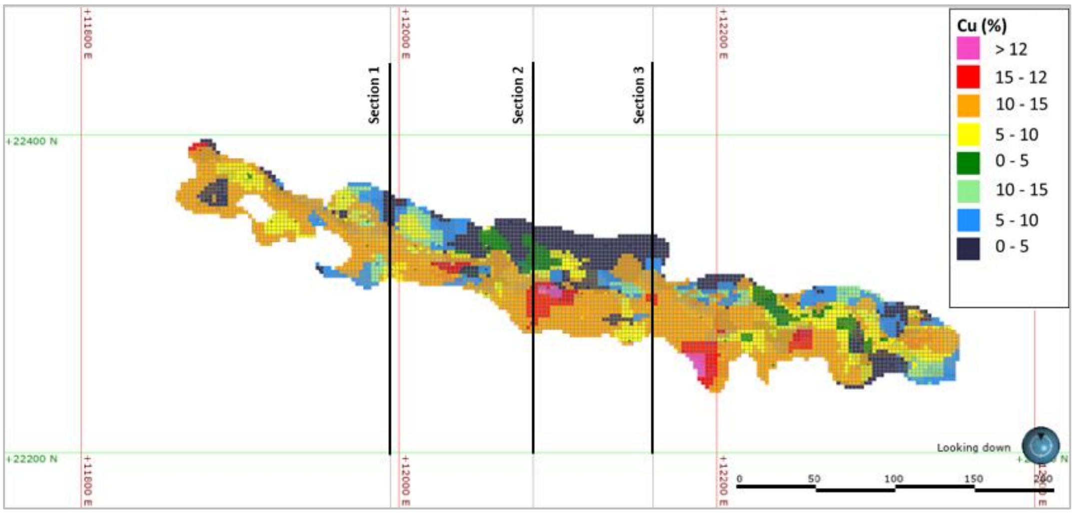


Figure 5.11: Plan-view of estimated block model for Vein 16 with three north-south cross-sections identified (sections 1–3)

The HG and LG domains are illustrated by the wireframe outlines, the samples are displayed as points.

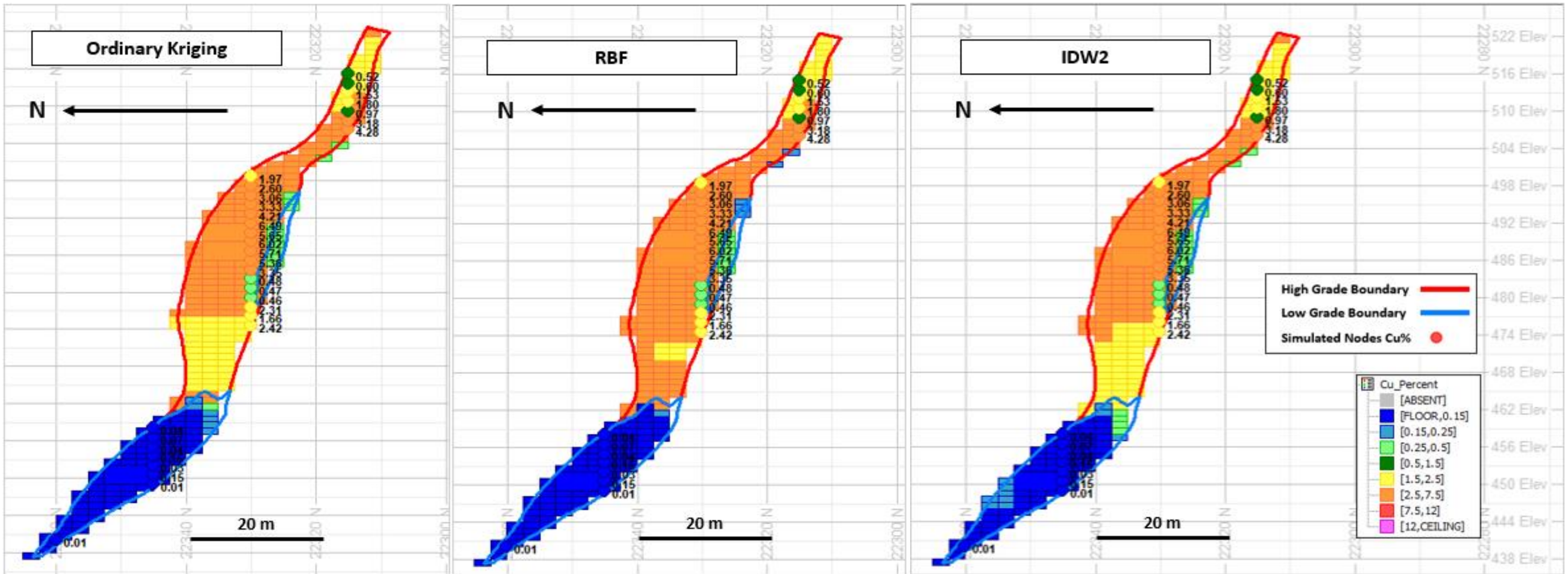


Figure 5.12: Cross-section 1 – comparison of estimation results for OK, RBF, and IDW2

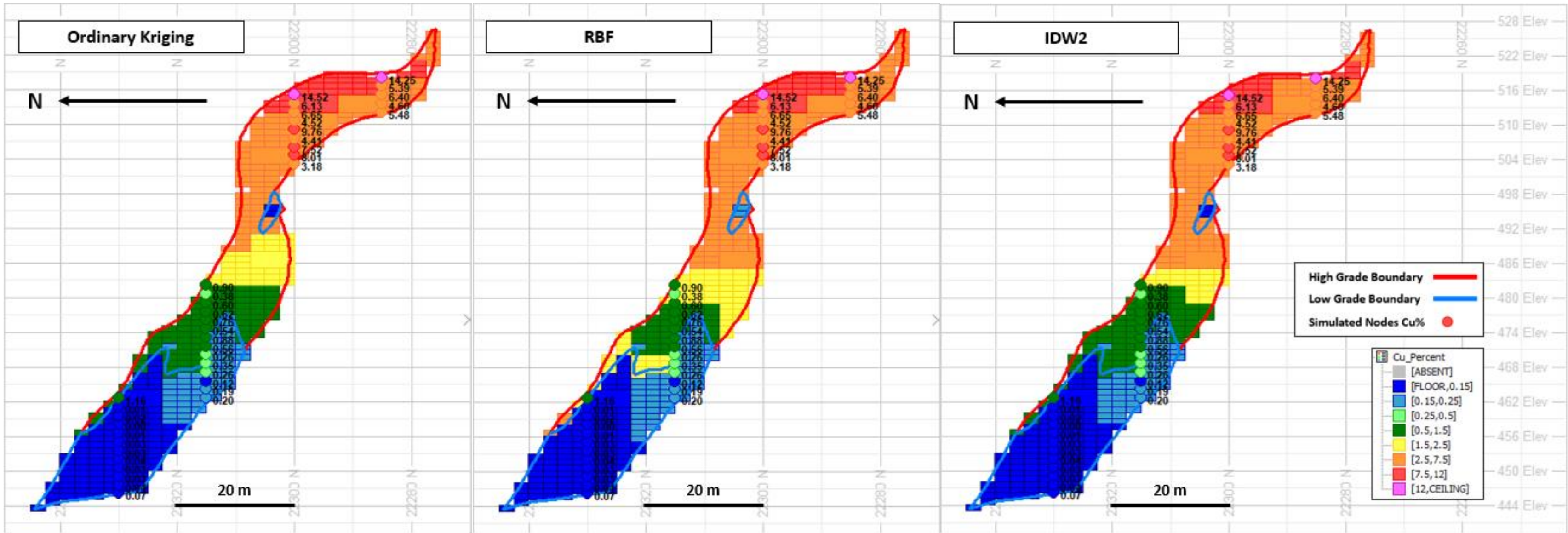


Figure 5.13: Cross-section 2 – comparison of estimation results for OK, RBF, and IDW2

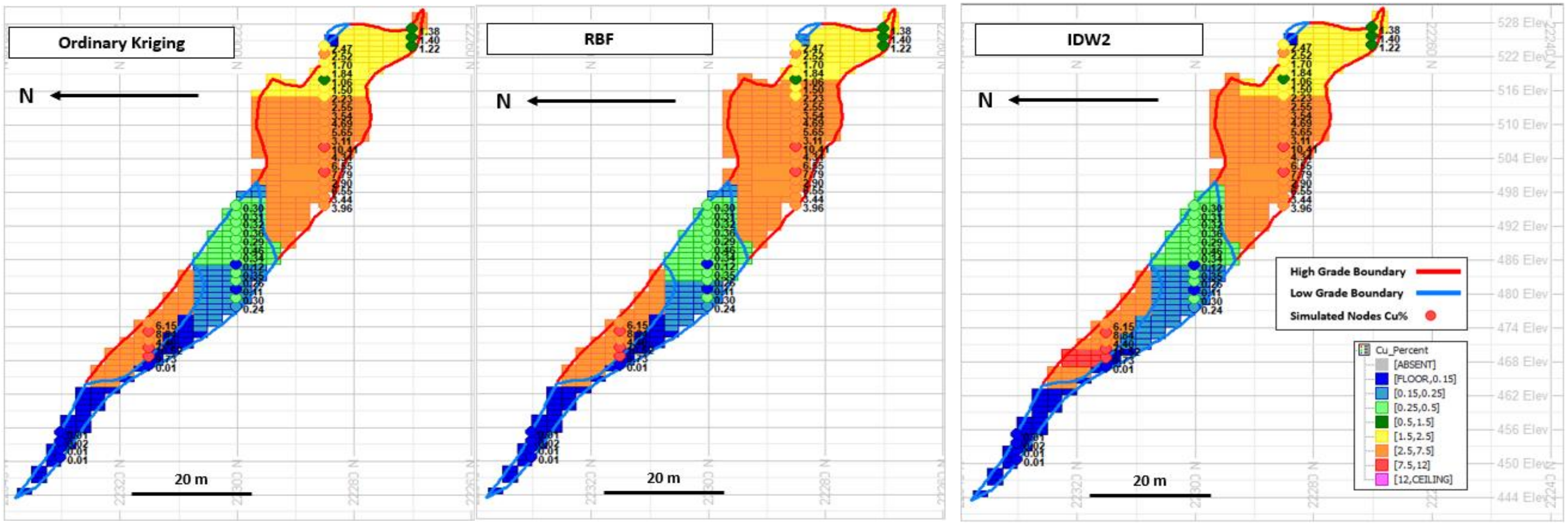


Figure 5.14: Cross-section 3 – comparison of estimation results for OK, RBF, and IDW2

5.2 Comparison against Simulated Reality

The descriptive statistics for the % Cu grade in the HG and LG domains appear in Table 5.3 and Table 5.4 respectively. Statistics for the simulation and the three block estimation methods are shown for comparison.

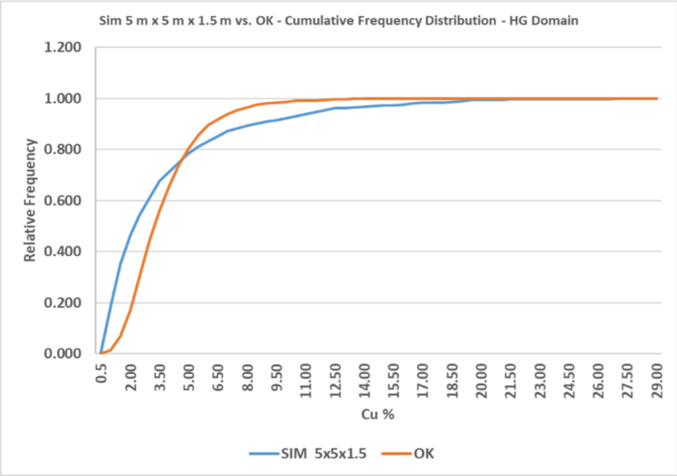
Table 5.3: HG domain % Cu summary statistics for the simulation and OK, RBF, and IDW2 block estimates

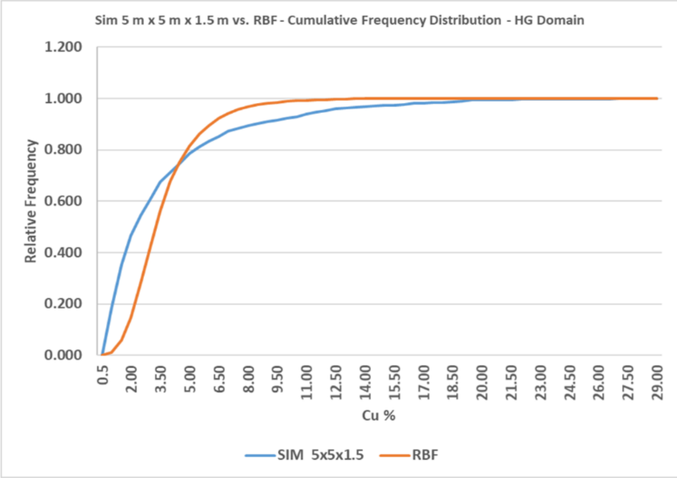
Estimate	Min	Max	Range	Mean	Median	Variance	Std Dev	CV
Simulation	0.074	28.991	28.918	3.630	2.280	14.990	3.872	1.065
OK	0.589	15.013	14.424	3.639	3.240	3.752	1.937	0.532
RBF	0.306	14.831	14.525	3.640	3.279	3.505	1.872	0.514
IDW2	0.382	15.231	14.848	3.599	3.176	3.502	1.871	0.520

Table 5.4: LG domain % Cu summary statistics for the simulation and OK, RBF, and IDW2 block estimates

Estimate	Min	Max	Range	Mean	Median	Variance	Std Dev	CV
Simulation	0.001	0.498	0.497	0.155	0.112	0.020	0.142	0.917
OK	0.013	0.449	0.437	0.153	0.144	0.009	0.094	0.612
RBF	0.009	0.438	0.430	0.153	0.141	0.007	0.086	0.564
IDW2	0.016	0.453	0.438	0.156	0.146	0.009	0.094	0.606

The comparative cumulative frequency distributions plots between the three estimates and the simulated reality within the HG and LG domains are shown in Figure 5.15 and Figure 5.16 respectively.





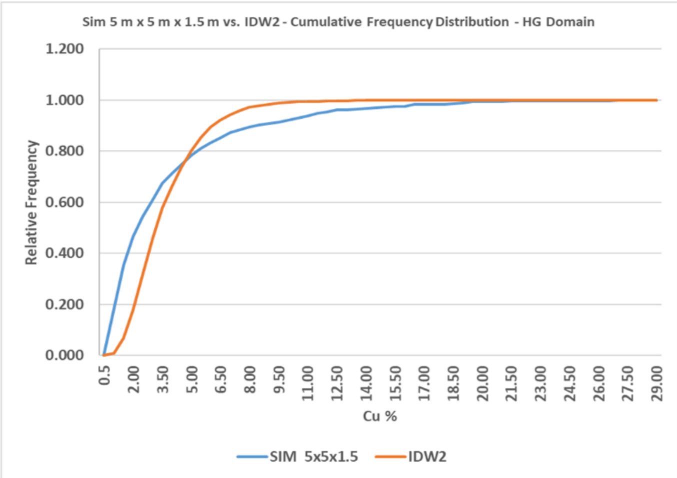


Figure 5.15: HG domain cumulative relative frequency distribution comparisons between the simulation and the OK, RBF, and IDW2 estimates

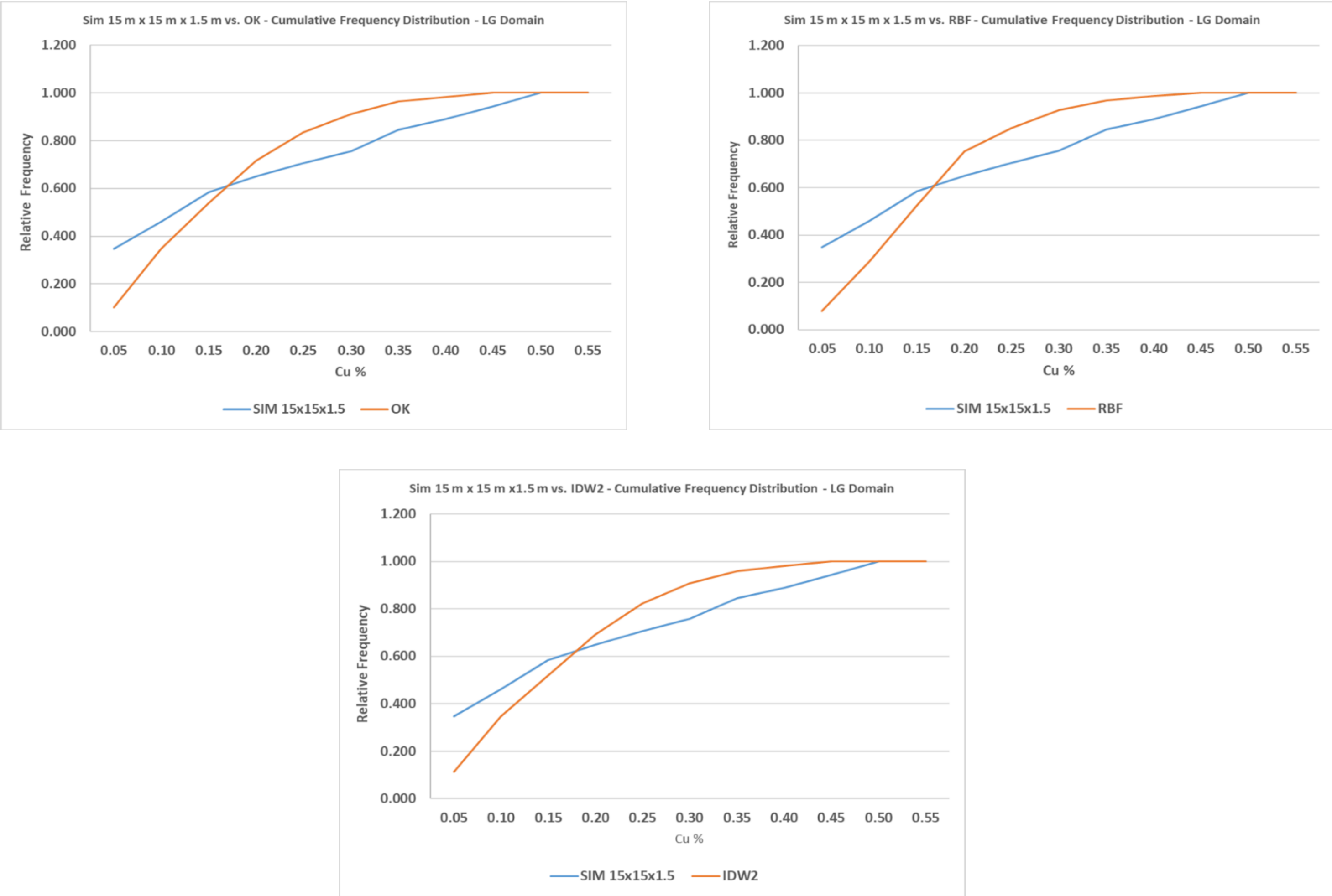


Figure 5.16: LG domain cumulative relative frequency distribution comparisons between the simulated reality and the OK, RBF, and IDW2 estimates

5.3 Measuring Confidence

As outlined in Chapter 4, sub-cells in the block model were filtered out and confidence limits were calculated for the OK estimates on the parent blocks only, which were graphed as swath plots. The plots illustrate the mean grades for the three estimates as individual lines and the number of parent cells are shown next to these lines. The 99% confidence limits were calculated on the mean OK grades and plotted as dotted lines. Only swaths with five blocks or more were considered in the comparison but, to create a continuous line for display purposes, some swaths with lower block counts were included.

5.3.1 High-grade domain

The figures below show the 99% confidence limits plotted with the three estimates for the HG domain in the X, Y, and Z directions.

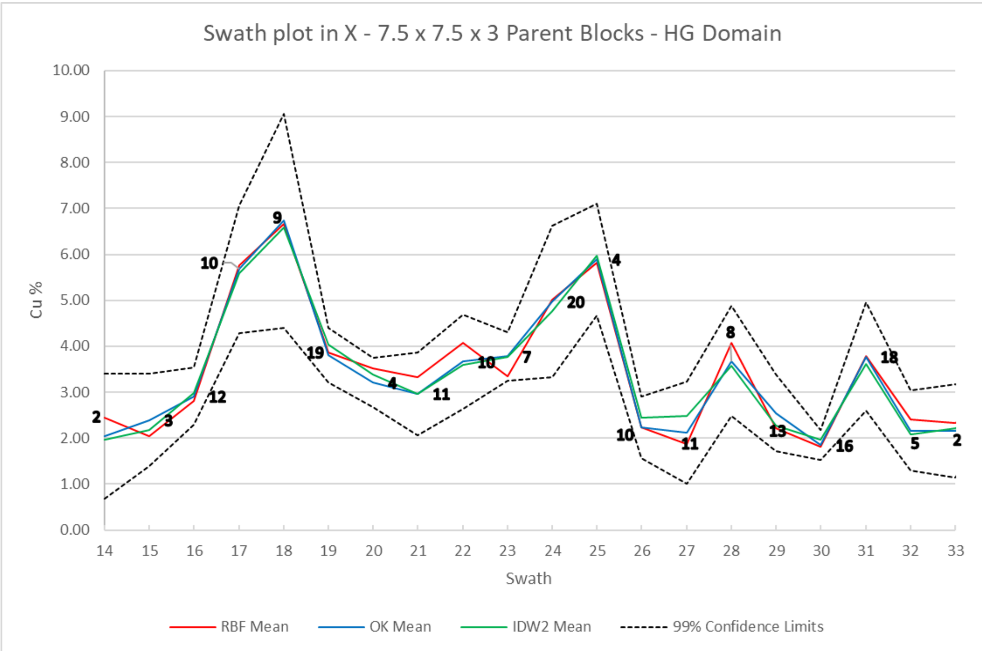


Figure 5.17: HG domain % Cu swath plot in X-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

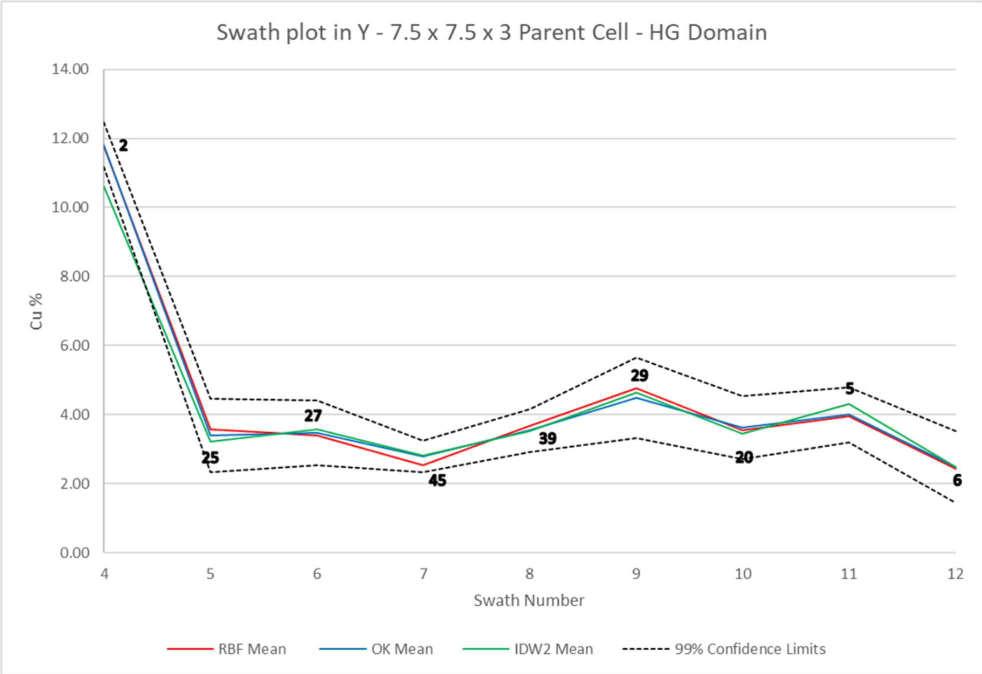


Figure 5.18: HG domain % Cu swath plot in Y-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

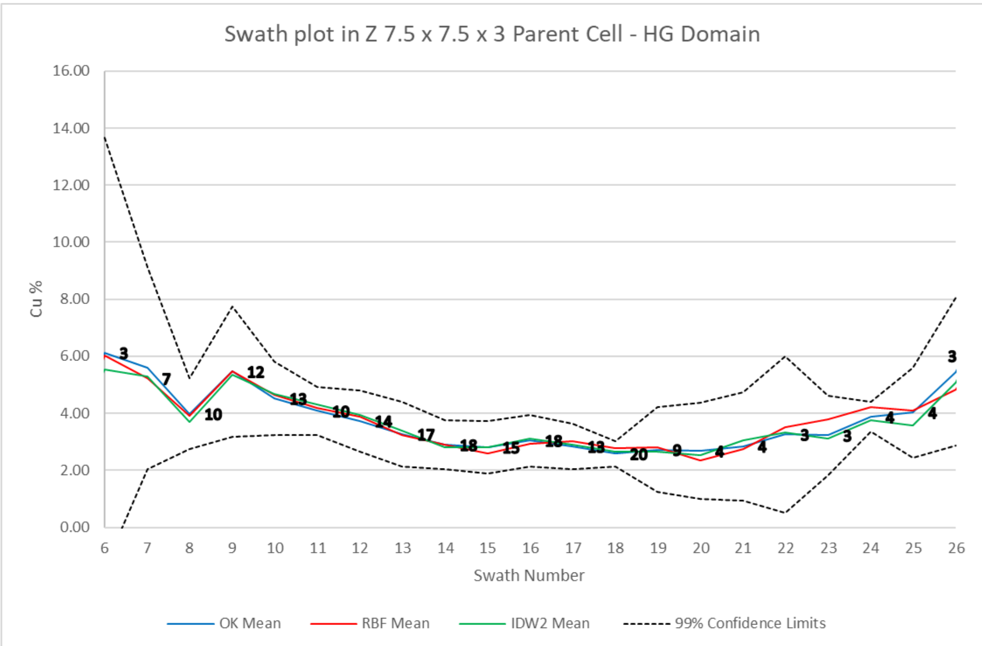


Figure 5.19: HG domain % Cu swath plot in Z-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

5.3.2 Low-grade domain

Swath plots containing the 99% confidence limits for the LG domain are shown in the graphs that follow. Due to the presence of fewer parent cells, some of these plots include swaths with parent cell numbers below the set limit.

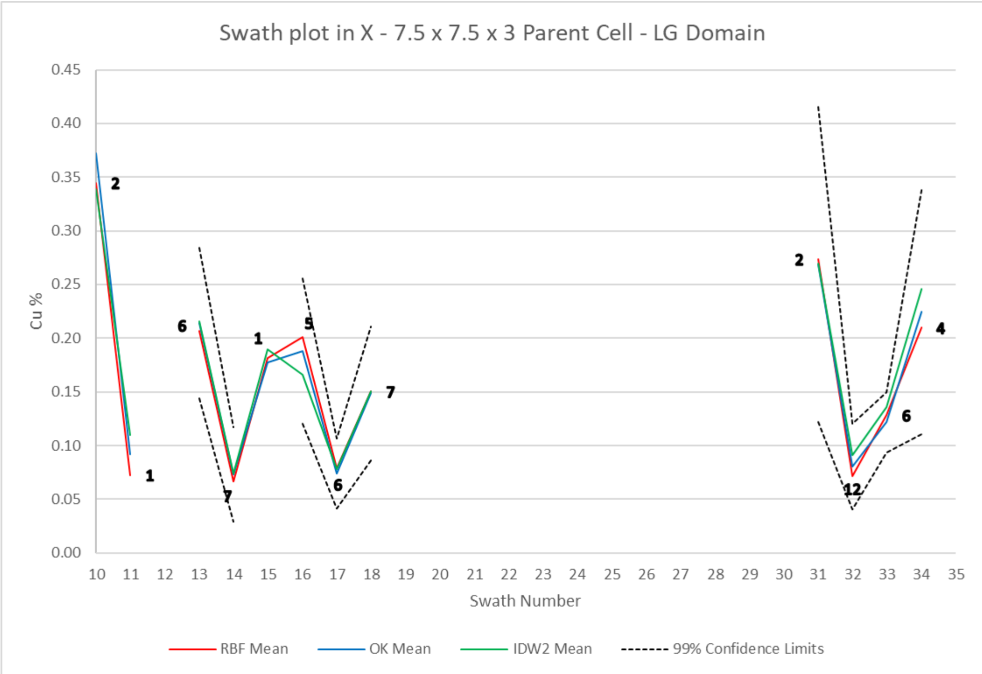


Figure 5.20: LG domain % Cu swath plot in X-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

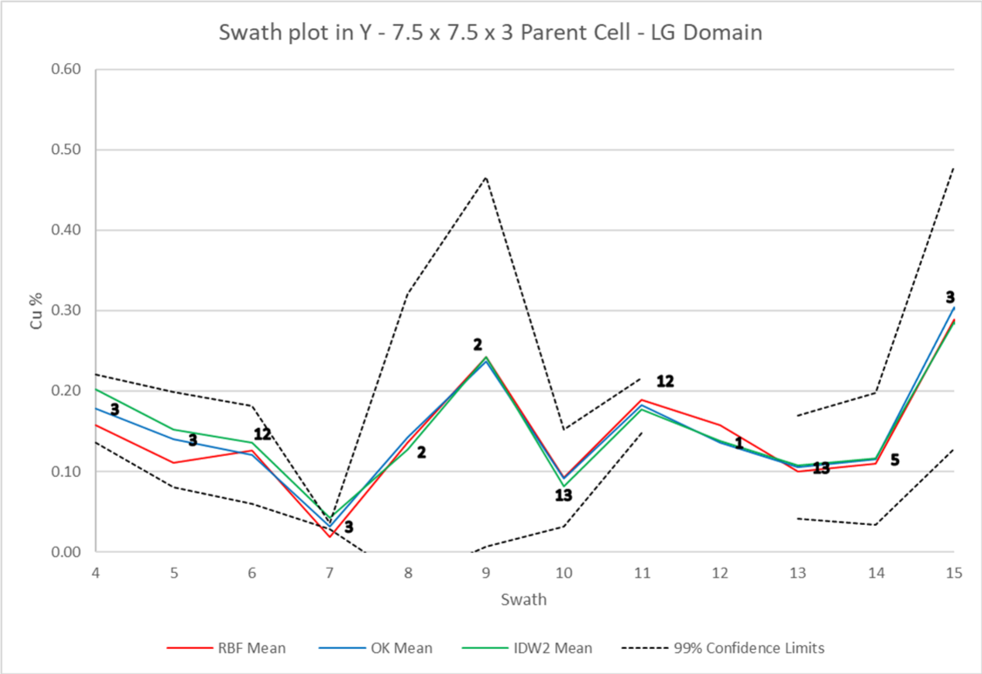


Figure 5.21: LG domain % Cu swath plot in Y-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

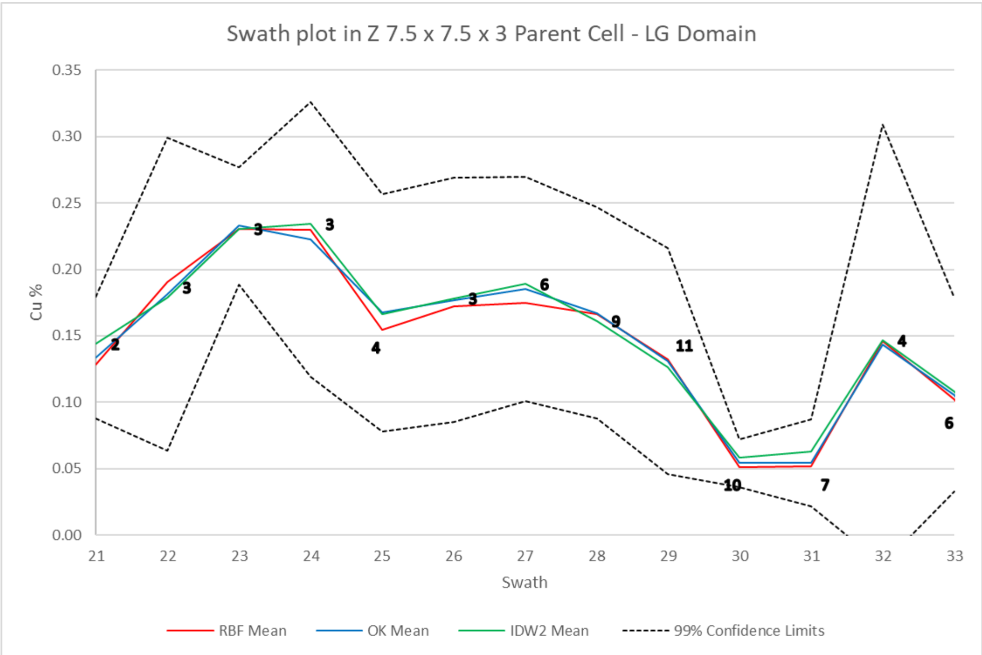


Figure 5.22: LG domain % Cu swath plot in Z-direction – RBF, OK, and IDW2 grade estimates and the 99% OK confidence limits

6 RESULTS ANALYSIS AND DISCUSSION

The purpose of this section is to discuss the results presented in the previous chapter. To contextualise the discussion, the research question is reiterated to provide a frame of reference for the reader. The discussion follows the same order as the results are presented in Chapter 5.

The aim of the research was to answer the following questions: Can the RBF method be used to estimate grades in situations where it is difficult to apply kriging methods where data coverage is poor and how do these estimates compare against more established methods? Since kriging is accepted as the best linear unbiased estimator, in which the error variance is minimised, it is used as a benchmark upon which to compare the RBF. As an alternative, IDW is often used when kriging methods cannot be applied easily or as a validation tool, and therefore it was included in the comparisons with the RBF. The questions to be answered pertain to the useability the RBF as an estimator and how comparable it is to more established techniques.

6.1 Statistical Validation

Comparing the summary statistics of the estimates against the model inputs offers a means to assess how comparable the three methods are to one another globally. Table 5.1 shows that, in the HG domain, the mean copper grades for all three methods are similar – the OK and RBF estimates are essentially identical, with a small difference noted with the IDW2 estimate. All three methods slightly overestimate the grades, with the mean grades for the OK and RBF estimates having a relative difference of 1.9%, while the IDW2 is closer to the average grade of the samples, with only a small relative difference of 0.8% noted. The CV for the three methods is similar, which suggests that the change of support occurs equally across all three methods. Table 5.2 shows that, for the LG domain, all three estimates are essentially the same, with a slight underestimation observed and relative differences in the mean close to 6% for the RBF and OK estimates and 4.4% for the IDW2 when compared against the average grades of the samples.

The CV values are similar, although for the RBF it is noted to be comparatively lower, indicating that, globally, these estimates have a lower variability, with a slightly greater degree of smoothing taking place for the RBF due to the estimates not being constrained by a search neighbourhood as the sample weights and drift coefficients are calculated globally.

The use of global statistics suggests that, on average, the three estimators are comparable for both domains. Cumulative frequency distributions for the HG and LG domains allow for the identification of how the statistical distributions of the estimation outputs differ to the input samples at an increasing cut-off.

For the HG domain (Figure 5.1), overestimation of the lower grades is observed to take place in a similar manner across all three methods. For both the block estimates and the input samples, approximately 70% of the grades are below 4.5% Cu. However, smoothing of grades is observed in the estimates, as the block grades tend to be overestimated – for example, at a cut-off of 1.5% Cu, less than 10% of the block estimates are below this grade, whereas this value is close to 40% for the samples due to the support effect – drillholes vs. block estimates. Conversely, above a cut-off of 4.5% Cu, the block grades tend to be underestimated in comparison to the samples. These observations are replicated for the LG domain (Figure 5.2), where the same change of support effect is noted.

6.2 Comparison between Methods

The previous discussion focused on the validation of the three methods against the samples, giving an assessment of how similar they were to the model inputs. Subsequently, the three estimates were plotted against one another in cumulative frequency plots to assess their similarity.

Cumulative frequency plots for the HG domain (Figure 5.3) show that there is little distinction between the three estimation methods. Cumulatively, a small difference is noted for the RBF, which shows a slight underestimation in the 1.5–2.5% Cu cut-off range and a slight overestimation in the 4–5% Cu cut-off. For the LG domain (Figure 5.4), the cumulative plots highlight more noticeable differences between estimates, with the RBF producing slightly higher-grade estimates below a cut-off of 0.15% Cu, while a slight underestimation between a cut-off of 0.15–0.30% Cu is noted.

Q-Q plots comparing the estimates confirm previous observations regarding the similarity of the estimates. For the HG domain, a comparison between the RBF and OK estimation shows a line that plots near the 45° line (Figure 5.5). Slight deviations along this line are noted, which correspond to the deviations seen in the cumulative plots, although the two methods are overall comparable. A similar observation can be seen for the RBF and IDW2 Q-Q plot, while the OK and IDW2 comparison is included for validation purposes (Figure 5.6), showing that no bias is introduced in the OK estimates.

For the LG domain, the observations made on the cumulative plots are emphasised in the Q-Q plot comparing the RBF and OK estimates (Figure 5.7). These differences are more accentuated when comparing the RBF and IDW2 estimates, while the OK and IDW2 estimates show a good correlation (Figure 5.8). These differences are important, because they illustrate how the search neighbourhood plays an important role for OK and IDW2. Due to the low sample numbers in the LG domain, the RBF is not constrained by a search neighbourhood; hence, samples far from the block will carry some weight in the local estimates. The presence of a smaller subdomain of a slightly higher grade was noted previously, and the influence of these samples is visible on the RBF, which highlights the importance of achieving stationarity in the domain definition phase.

6.3 Swath Plot Validation

The similarity between the three estimates is expanded upon in the swath plot validations by looking at their differences spatially along three defined directions, where the X-direction is along strike, Y-direction across strike, and the Z-direction which shows differences down the dip of the vein. For the HG domain (Figure 5.9), along the X-direction, the RBF, OK, and IDW estimates are comparable, showing the same degree of overestimation and underestimation in relation to the samples. The larger deviations from the input grades tend to occur in the west, between swaths 1 and 7, where fewer samples are present in the HG domain. From swaths 33 onwards, sample numbers become lower and extrapolation takes place along the edges of the domain, resulting in deviations between the three methods.

Along the Y-direction, where sample numbers are high, the three estimates are closely similar and display comparable differences to the input grades. Large deviations occur where sample numbers are lower or non-existent and for swaths 1 and 2, where extrapolation is taking place, the outputs vary substantially from one another.

Along the Z-direction, the same behaviour described above is observed, with the three estimates closely equating to each other and the samples, where abundant samples are present. At the upper and lower portions of the domain, where extrapolation takes place, the RBF and OK behave in a similar manner, while IDW2 tends to be comparatively underestimating.

For the LG domain (Figure 5.10), the X-direction swath plot shows a good correlation between the three methods when supported by sufficient data. The largest deviations coincide with lower sample numbers and, in comparison with the other two methods, the RBF estimate tends to be overestimated in some places where sample numbers are lower (swaths 1–7) and underestimated in others where extrapolation occurs (swaths 34 and 35).

In the Y-direction, the same strong relationship between grades and sample numbers is observed. The RBF shows a relative overestimation for swaths 19–23, where sample numbers begin to drop. The RBF estimate for the Z-direction swath in the LG domain tends to show a smoother profile, wherever extrapolation of grades take place (swaths 1–10) or when the copper grades for the samples tend to be more variable (swaths 17–26).

The swath plots for both domains support the observation that where sample numbers are higher, the RBF, OK, and IDW2 estimates tend to be more similar. The poor availability of samples affects the estimates differently, with the highest deviations detected at the boundaries of the domain, where a higher degree of extrapolation occurs. The observation that the IDW2 and OK estimates tend to follow similar trends is expected, as these two methods use the same search neighbourhood, optimised by QKNA. However, the RBF assigns weights to data that fall outside of this search neighbourhood. These differences can be attributed to how the different methods assign weights to the samples. IDW2 will always tend to give more weight to the nearer samples, while OK optimises the kriging weights by introducing the Lagrange multiplier to the kriging system of equations, which ensures the weights equal one and using the variogram model. Furthermore, due to the screening effect, clustered samples are accounted for with negative weights. However, due to the simulated drilling, grid samples were not clustered. The RBF uses the entire dataset, and hence when sample numbers become lower, more weight is given to samples further from the block, due to the global drift model inherent in the RBF.

6.4 Visual Validations

Visual validations are an important component of model validation. Observing the model in sections allows for small differences between the three estimation methods to be identified that otherwise would be impossible to discern from global statistics or swath plots. For all three sections (Figure 5.12, Figure 5.13, and Figure 5.14), differences in the block estimates tend to be small whenever these are well informed by nearby samples, with differences becoming more noticeable moving away from the data. As previously identified in the statistics, similar degrees of smoothing are observed across all three methods, with comparable replication of high grades and low grades in each domain.

In between sample lines, larger differences are noted, with shifting grade boundaries being observed across the three methods. The OK and IDW2 block estimates are visually similar, while the RBF block estimates are often quite different (Figure 5.12). In other sections, these differences are not as apparent.

6.5 Comparison against Simulated Reality

Up to this point, validations have shown that the three estimation methods successfully captured the relationship between grades for the samples. Additionally, validations suggest that the estimates are equally comparable amongst themselves and no single method can be singled out as superior, as they all have similar smoothing effects.

Comparing the global statistics of the estimated blocks against the simulated reality (i.e., the 5 m x 5 m x 1.5 m nodes), the three methods for both the HG (Table 5.3) and LG (Table 5.4) domains are similar. The change of support occurs in a similar way for all methods, as indicated by the similar CV values. However, as expected, neither of these methods can replicate the heterogeneity of the simulated grades. Thus, a certain amount of smoothing is always to be expected due to the change of support effect and the Central Limit Theorem, as the support increases the variability reduces.

This smoothing effect can be observed in the cumulative frequency distribution comparisons for the HG (Figure 5.15) and LG (Figure 5.16) domains. Since neither of the methods perfectly fits the distribution of the simulated grades, it cannot be stated categorically which method is superior. Nevertheless, it can be affirmed that the RBF performs well as an estimator and is comparable to the OK and IDW2 estimates.

6.6 Measuring Confidence

The swath plots (see Figure 5.17 to Figure 5.22) showing the 99% confidence limits allow the researcher to spatially assess the probability that the CI contains the true mean and whether the RBF falls within these limits. An RBF estimate inside of the confidence limits implies good confidence in the estimates, while one that falls outside can be assumed to be more biased.

For the HG domain, the swath plot in the X-direction (Figure 5.17) shows that there is a strong correlation between the number of parent cells and how close the three estimates are to each other. Where parent cell numbers are fewer, the RBF tends to deviate from the OK and IDW2 estimates, although there is no discernible pattern of consistent over- or underestimation along this direction. Similarly, for the Y- and Z-direction swath plots (Figure 5.18 and Figure 5.19 respectively), the three estimates correlate well and no bias in the RBF estimates can be identified. Where parent block numbers become lower for the Z-direction plot, the confidence limits become wider and some deviation in the RBF is observed. Consequently, only swaths with five or more parent cells are considered relevant.

The comparison of the LG domain estimates against the 99% CIs was more challenging, due to fewer parent blocks being available in the domain. This is to be expected, as the vein structurally pinches out at depth, meaning the block model relies more on sub-celling to optimally fill the domain volume. As a result, wider CIs were observed across all swath plots.

For the X-direction (Figure 5.20), the swath plot has a noticeable gap (between swaths 19 and 31), limiting any observation for these areas. For the sections where parent cells exceed five, it can be observed that the RBF estimates fall well within the plotted limits. For the Y-direction swath plot (Figure 5.21), a similar observation can be made and, although the RBF falls outside the lower confidence limit at swath 7, it is noted that there are only three parent blocks available, a number that is insufficient for optimal comparisons. The Z-direction swath plot (Figure 5.22) similarly shows no evidence of biased estimates for the RBF method.

The RBF for both domains plot well within the confidence limits calculated on the OK estimates, implying with a 99% confidence that the RBF estimates are within the limits of the true mean (i.e., that of reality). These comparisons indicate that the RBF offers a robust estimate that is comparable to OK and IDW2, where supported by data, the RBF is closely similar to the other two methods. However, where data are sparse, observed differences are still well within the confidence limits.

7 CONCLUSION

The RBF has seen wide applications in the mining industry over the past two decades as an implicit modelling technique for generating wireframe surfaces and volumes. Resource estimation in the mining industry remains to be dominated by deterministic geometric methods, such as IDW, and geostatistical methods like kriging. This may lead to alternative methods being overlooked, even in situations where the application of more established techniques becomes challenging and ill-suited for a particular deposit or dataset. The RBF has seen a multitude of applications in emerging technologies in computational learning and estimation, but its application has been hindered by the need for background knowledge and experience in computer programming languages, such as R and Python, which are skills not commonly possessed by production or resource geologists working at a typical mining operation. With the introduction of the RBF to Leapfrog EDGE®, the method is now more easily accessible to a wider range of users, but this does not preclude the need to know how the method works.

From a practical standpoint, this research study corroborates the conclusion made in similar studies that the RBF and OK give similar results when supported by sufficient data. Therefore, their application is well-suited for grade control model updates for production purposes, where data is more abundant. Furthermore, as the RBF is capable of handling data clustering in a similar way to kriging, it can be considered an alternative to IDW, which assigns higher weightings to samples nearer to a block without accounting for data clusters.

This study shows that in settings where data are sparse or where modelling experimental variograms is challenging, the RBF is comparable to OK and IDW, with estimates for the three methods being visually similar. The RBF tends to differ slightly in areas that are poorly informed – in the HG domain, estimates are closest to the OK outputs; while in the LG domain, differences are more noticeable, with the IDW2 and OK outputs being more similar.

These differences can be explained easily, as OK utilises a search neighbourhood, with search ranges, sample maximum and minimums, and contributions by individual drillholes all optimised by QKNA. In areas of sparse information, kriging estimates can become biased due to fewer samples being available introducing edge effects. Consequently, it is often common practice in the industry to expand search volumes to fill all blocks with estimates. The RBF does not have this limitation as it is a global estimator in terms of the drift function, and with the application of a drift component, estimates along domain boundaries will always be assigned a grade estimate.

A benefit of applying the RBF is its ease of use, as semivariogram modelling is not a prerequisite and stationarity is not necessary. Nevertheless, well-defined estimation domains that conform to the geology and structure of a deposit will result in a more robust estimate. Therefore, as with kriging, EDA of all available data backed up by a good understanding of the deposit is as important in applying the RBF.

The principal shortfall in using the RBF is the lack of comparable metrics to kriging that allow quantifying the confidence in an estimate. Even though this study showed that the RBF falls well within the 99% confidence limits for an OK estimate, the absence of measures of confidence means that the estimate can only be validated by visually checking the block grades against the input samples, through statistical checks and swath plot validations. In this sense, RBF is unlikely to displace kriging as a method in Mineral Resource estimations but can rather be seen as playing a role in grade control where data are more abundant and often not clustered. It is expected that with time the RBF will perform better as more grade control data at closer distances are incorporated in the definition of the RBF model.

The use of the RBF in computational learning techniques and its application in resource estimation presents an opportunity for further investigation. Future investigation into the performance of this method in different geological settings, particularly structurally complex deposits, may offer insights in estimating resources where applying traditional methods can be challenging.

REFERENCES

Abzalov, M. (2016). *Applied mining geolog*y. Modern approaches in solid earth sciences, Vol. 12. doi:10.1007/978-3-319-39264-6.

Afonja, T. (2017). *Kernel functions*. Available at: https://towardsdatascience.com/kernel-function-6f1d2be6091

Beatson, R.K., Cherrie, J.B. and Mouat, C.T. (1999). Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration. *Advances in Computational Mathematics*, 11(2-3): 253-270. doi:10.1023/A:1018932227617.

Broughton, D.W. (2014). *Geology and ore deposits of the Central African Copperbelt* (Doctoral thesis). Golden, CO: Colorado School of Mines.

Carrasco, P.C. (2010). Nugget effect, artificial or natural? *Journal of the Southern African Institute of Mining and Metallurgy*, 110(6): 299-305.

Chilès, J.-P. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty*. 2nd ed. Hoboken, NJ: John Wiley & Sons. pp. 186-187

Chinyuku, D.T. (2013). *The Kansanshi Cu-Au deposit, Domes Region, Zambia: Geology, mineralisation and alteration characteristics in the main pit* (Master's thesis). Grahamstown: Rhodes University.

Coombes, J. (2008). *The art and science of resource estimation: A practical guide for geologists and engineers*. Perth: Coombes Capability.

Costa, J.-P., Pronzato, L. and Thierry, E. (1999). “A comparison between kriging and radial basis function networks for nonlinear prediction”, in *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99), Antalya, Turkey, June 20-23, 1999* (pp. 670-674). Available at: https://www.eurasip.org/Proceedings/Ext/NSIP99/Nsip99/papers/155.pdf

Cowan, E.J., Beatson, R.K., Ross, H.J., Fright, W.R., McLennan, T.J., Evans, T.R. et al. (2003). “Practical implicit geological modelling”, in

S. Dominy (ed.), *5th International Mining Geology Conference* (pp. 89-99). Victoria: The Australasian Institute of Mining and Metallurgy. Available at: http://hyperfun.org/FHF_Log/Cowan_PIGM03.pdf.

Cowan, E.J., Beatson, R.K., Fright, W.R., McLennan, T.J. and Mitchell, T.J. (2002). “Rapid geological modelling” [Extended abstract], in S. Vearncombe (ed.), *Applied Structural Geology for Mineral Exploration and Mining International Symposium: Abstract Volume* (36, pp. 1-9). Sydney: Australian Institute of Geoscientists.

Deutsch, C.V. and Journel, A.G. (1997). *GSLIB: Geostatistical software library and user’s guide*. 2nd ed. New York, NY: Oxford University Press.

Dos Santos, T.C. and Yamamoto, J.K. (2019). Ore resource estimation based on radial based functions – case study on União Luiz and Morro do Carrapato gold deposits (Alta Floresta Gold Province). *Revista Escola de Minas*, 72(3): 493-499. doi:10.1590/0370-44672018720154.

Ghezelbash, R., Maghsoudi, A. and Carranza, E.J.M. (2019). Performance evaluation of RBF- and SVM-based machine learning algorithms for predictive mineral prospectivity modeling: Integration of S-A multifractal model and mineralization controls. *Earth Science Informatics*, 12(3): 277-293. doi:10.1007/s12145-018-00377-6.

Glacken, I.M. and Snowden, D.V. (2001). “Mineral Resource Estimation”, in A.C. Edwards (ed.), *Mineral Resource and Ore Reserve Estimation – The AusIMM Guide to Good practice* (pp. 189-198). Victoria: The Australasian Institute of Mining and Metallurgy.

Guttorp, P. and Lindgren, G. (2018). “Matérn, Bertil”, in N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggieri and J.L. Teugels (eds.), *Wiley StatsRef: Statistics Reference Online*. doi:10.1002/9781118445112.stat08143.

Hardy, R.L. (1971). Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research*, 76(8): 1905-1915. doi:10.1029/JB076i008p01905.

Kaplan, U.E. and Topal, E. (2020). A New Ore Grade Estimation Using Combine Machine Learning Algorithms. *Minerals 2020*, 10(10): 847 (pp. 1-17). doi:10.3390/min10100847.

Kentwell, D. (2014). Radial basis functions and kriging – a gold case study. *AusIMM Bulletin*, 6: 2-9.

Nowak, M. and Verly, G. (2004). “The practice of sequential Gaussian simulation”, in O. Leuangthong and C.V. Deutsch (eds.), *Geostatistics Banff 2004*. Quantitative Geology and Statistics (Vol. 14, pp. 387-398). doi:10.1007/978-1-4020-3610-1_39.

Pillay, N. (2021). Copper continues to climb. *LinkedIn*, 22 February. Available at: https://www.linkedin.com/pulse/copper-continues-climb-nevan-pillay/?trackingId=kxPJZWHrQQsP61P6es4ydA%3D%3D

Rossi, M.E. and Deutsch, C.V. (2014). *Mineral resource estimation*. doi:10.1007/978-1-4020-5717-5.

Salomon, D. (2006). “Spline Interpolation”, in *Curves and Surfaces for Computer Graphics* (pp. 141-173). New York, NY: Springer.

Samanta, B. (2010). Radial basis function network for ore grade estimation. *Natural Resources Research*, 19(2): 91-102. doi:10.1007/s11053-010-9115-z.

Stewart, M. (2019). The spheroidal family of variograms explained. *Seequent* [Blog post], 11 June. Available at: https://www.seequent.com/the-spheroidal-family-of-variograms-explained/

Stewart, M., De Lacey, J., Hodkiewicz, P.F. and Lane, R. (2014). “Grade Estimation from Radial Basis Functions – How Does it Compare with Conventional Geostatistical Estimation?”, in *Ninth International Mining Geology Conference 2014* (pp. 129-139). Victoria: The Australasian Institute of Mining and Metallurgy.

Theron, S.J. (2013). “The origin of the Central African Copperbelt: In a nutshell”, in *Base Metals Conference 2013* (pp. 21-36). Johannesburg: The Southern African Institute of Mining and

Metallurgy.

Torrealday, H.I., Hitzman, M.W., Stein, H.J., Markley, R.J., Armstrong, R. and Broughton, D. (2000). Re-Os and U-Pb dating of the vein-hosted mineralization at the Kansanshi copper deposit, Northern Zambia. *Economic Geology*, 95(5): 1165-1170. doi:10.2113/gsecongeo.95.5.1165.

Vann, J., Jackson, S. and Bertoli, O. (2003). “Quantitative kriging neighbourhood analysis for the mining geologist – a description of the method with worked case examples”, in S. Dominy (ed.), *5th International Mining Geology Conference* (pp. 215-223). Victoria: The Australasian Institute of Mining and Metallurgy.

Yamamoto, J.K. (2002). Ore Reserve Estimation Using Radial Basis Functions. *Revista do Instituto Geológico*, 23(1): pp. 25-38. doi:10.5935/0100-929X.20020003.

APPENDIX A

Table A.1: Parameters used for SGS for the HG and LG domains

Parameter	HG	LG
Minimum input grade value	0.13	0.00056
Maximum input grade value	27.23	0.49
Minimum simulated grade value	0	0
Maximum simulated grade value	30	0.49
Back-transformation method for lower tail (1-none, 2-power model)	1	1
Power parameter in back-transformation for lower tail	1	1
Back-transformation method for upper tail (1-none, 2-power model)	2	1
Power parameter in back-transformation for upper tail	0.45	1
Number of simulations	100	100
Minimum X coordinate	11 840	11 840
Minimum Y coordinate	22 180	22 180
Minimum Z coordinate	410	410
Spacing of nodes in X	5	5
Spacing of nodes in Y	5	5
Spacing of nodes in Z	1.5	1.5
Number of nodes in X	112	112
Number of nodes in Y	64	64
Number of nodes in Z	100	100
Minimum number of original data to be used	4	4
Maximum number of original data to be used	30	30
Maximum number of simulated nodes to be used	30	30
Search distance in X direction	210	200
Search distance in Y direction	90	100
Search distance in Z direction	45	100
First rotation angle for search ellipsoid (Z-axis)	0.431	0.646
Second rotation angle for search ellipsoid (X-axis)	63.255	54.142
Third rotation angle for search ellipsoid (Y-axis)	-13.067	-11.717