Resource Estimation with Multiple Data Types

by

Patrick Noel Donovan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Improving the accuracy of estimates is of great importance in the mining industry. These estimates can be improved with the use of available secondary data. This includes any less trusted data source containing bias and error such as production sampling, legacy drill holes, or cheaper lower quality measurements. Cokriging methods allow the application of this secondary data without transmitting bias and error through to the final estimates.

When secondary data is unequally sampled compared to the primary data source it is difficult to establish correlation between the data types. This is solved by modelling the cross covariance curve. A case study using exploration and production data shows that cokriging provides a more accurate result and a reasonable decrease in average estimation variance over ordinary kriging methods.

This application of cokriging can be extended to improve resource classification in terms of data spacing. A decrease in average estimation variance using cokriging methods can be shown to relate to a decreased 'effective' data spacing for classification purposes. This spacing reflects the improvement in the estimates with the use of secondary data. A repeatable power model relationship with correlation is defined to relate average estimation variance improvement to data spacing. A case study shows an application with real exploration and production data. The result is a scaling of combined data spacing with reference to the information content of the secondary data type.

To my wife, Lindsey.

Acknowledgements

Thank you to Dr. Clayton Deutsch for his guidance and support throughout my time at the University of Alberta.

I would like to thank my colleagues at the Centre for Computational Geostatistics. Help they have provided is greatly appreciated. Through discussion we have learned together and solved many problems.

Thank you to my parents, Tom and Ruth, who have been there for me as sounding boards and personal boosters.

Finally, I would like to thank my wife. Without Lindsey this would not have been possible. I appreciate every sacrifice she has made to facilitate my return to education.

Table of Contents

List of Figures

List of Tables

List of Symbols

Symbol	Definition	First Use
μ	Mean	17
$C(0)$	Collocated covariance	19
ρ	Coefficient of correlation	19
λ	Kriging weight	20
σ^2	Variance	21
Γ	Variogram structure	23
f	Modification factor for 'effective' data spacing	43
ds_{eff}	'Effective' data spacing	43
ds_{prim}	Primary data spacing	43
$ds_{prim+sec}$	Combined data spacing	43
$\bar{\sigma}^2$	Average variance	48

List of Abbreviations

Abbrv.	Definition	First Use
RC	Reverse circulation	3
SMU	Selective mining unit	3
DDH	Diamond drill hole	6
LMC	Linear model of coregionalization	9
QP	Qualified person	10
NI43-101	National Instrument 43-101	11
NE	Nugget effect	13
ANN	Artificial Neural Network	14
SK	Simple kriging	20
OK	Ordinary kriging	21
SCK	Simple cokriging	26
CIM	Canadian Institute of Mining, Metallurgy, and Petroleum	37
JORC	Joint Ore Reserves Committee	39
SAMREC	South African code for the reporting of exploration results, mineral resources, and mineral reserves	39
PERC	Pan-European reporting code	39
BH	Blast holes	50

Chapter 1

Introduction

In modern geostatistical studies the practitioner is often presented with multiple data sources of varying quality that provide varying information for resource estimation. This presents a challenge to employ all available data to greatest effect.

1.1 Problem of multiple data types

Each study, whether it is in the domain of mining, petroleum, or environmental relies on a variety of data to characterize the geological site in question. The amount of data has increased due to advances in data capture, resolution, and alternative data collection techniques. High quality data in terms of accuracy and precision is often expensive while lesser quality data could be gathered at a reduced cost. Characterization of the geological site under investigation would be improved if all data were considered. This requires the quantification of the relationships between the different types of data and the unbiased integration of all data into the final resource model.

1.1.1 Differing quality of data

At least one type of data should be considered high quality, that is, providing reliable measurements of the property under consideration. High quality data are easily replicated and regularly show good results when compared with known values and blanks (prepared samples known to contain zero value in assay of the variable of interest). In reference to multiple data types this would

be considered a primary or hard data type. In mining this is often assays from diamond drilling core samples while petroleum studies often refer to core and well logs. The primary data type is the best data available for use in the study. It represents the reference that all other data types can be measured against.

There are often lesser quality data available. There are many possible reasons for data types to be of lesser quality. These samples may not be complete homogeneous representations of the volume sampled. The sampling process may be biased by physical or operational design. Quality assurance and quality control measures may not enforce sampling repeatability. Historical records may be lost. Also, the measurement may be inexpensive yet not completely reliable. For these reasons, and others, certain data are considered secondary data types as they contain some amount of bias and error in comparison to the primary data type. A good example of secondary data would be chip sampling as done in many mining operations. The location and volume of chip samples are inherently non-repeatable. Though the assays resulting from this sampling type may show a similar outcome when compared to local primary data (e.g. diamond drilling assays), they often have increased error and possible bias.

If samples of the different data types are available at the same location they are considered homotopic; otherwise, they are considered heterotopic. When there is some collocation in samples this is partial heterotopy. In practice, multiple data types are almost always heterotopic.

The data collection process has a great effect on how data are treated. It is important to maintain high standards in quality assurance and quality control in terms of the processes involved in sampling. Poor data collection provides poor inputs for project definition and a risk that bias and/or error may be passed on in results. It is always better to avoid the problem in the first place than to have to deal with this problem later in the process.

A high quality data type is required as a reference to measure the quality of secondary data. In the absence of a high quality primary data type there is little that can be done. Also, secondary data that are considered very poor quality, or with unknown local variations in quality, may be rejected.

Note that some practitioners may consider the reference to secondary data

relationship to be associated with data that is the result of some form of interpretation of a primary data type. An example is structural or depositional information that is interpreted from diamond drill data. This meaning is not applied here; secondary data represent variables for which there are lower quality measurements.

1.1.2 Data scale

An additional consideration for multiple data type integration is the scale of the data. While primary data types often consider relatively small sample volumes secondary data may consider a larger or smaller volume. For example, reverse circulation(RC) drilling may sample larger volumes than diamond drilling. The simultaneous use of multiple data types at different sample support will require either the explicit use of volume variance scaling laws or considering one of the data scales as a reference and others relative to it.

At the initial stages of domain characterization it may be possible to sidestep the varying support problem though appropriate compositing of data. Compositing consists of averaging values to a common length. The composite size is defined by structural and/or engineering requirements and is often restricted to the selective mining unit (SMU), size that is defined by the selective volume which can be extracted using current methods (Rossi and Deutsch, 2014).

1.1.3 Bias and error in data

Most modern geostatistical studies consider a number of data types. It is considered critical to manage data with significant bias and error.

In addressing bias and error it is first important to understand the various sources and their causes. Bias and error may be communicated from the sampling protocol. If a sampling scheme is poorly conceived it may introduce bias and/or error. An example would be chip sampling in a tunnelling setting. A geologist is expected to measure sample lengths and extract a representative sample. Bias and error can be introduced by the geologist in terms of differing sample division and inattention to protocol. As well, lab procedures may

contribute to increased bias and error. Problems should be addressed and corrected directly at the source if possible. Equipment and human error may also cause bias and error. Once these issues are noted there may be room for correction in terms of higher quality sampling.

The number of samples required and the configuration of these samples may be improved. Procedural bias and error can be identified and random errors can be reduced to a minimum. One may also be successful in identifying bias and error in the initial statistical assessment of the raw data.

Once bias and error are identified they must be dealt with. They can be considered separately. Firstly, bias due to preferential sampling may be removed through declustering of the data and understanding how the mean is different for each data type and location. The data values may be biased due to detection limits or by capped maximum values. This can create too many low or high values. In this case the bias could be removed through an ordered ranking of the values. Bias could also be removed in the application of cokriging as operations are performed on the data residuals. The estimates, therefore, only consider the primary data mean. Error in sampling can be identified using standard and blank samples. Sample standards are manufactured samples with known assay values designed as a check of the assay process. These provide a baseline to measure the sampling error and make corrections. When cokriging with error in the secondary data, the error is removed through the application of different mean values and use of the cross-covariance. This will be discussed in Chapter 3.

The calibration of bias and error can be difficult. Primary and secondary data types are not measured at the same location so there are no paired data for direct comparison. This will also be addressed in Chapter 3

1.2 Approaches in multiple data-type usage

When considering secondary data in estimation there are currently few options. The secondary data could be rejected completely or treated as the same quality as the primary data. There may be a user-defined weighting scheme applied

so as to reduce the effect of secondary data on a study. Finally, mathematical weighting systems are available for data integration and this is the focus of this thesis.

1.2.1 Reference usage

In some cases the secondary data is considered of such low quality, when compared to primary data, that it is applied only as a reference or a check to confirm knowledge gained from the primary data. An example of this is the collection of channel samples in development mining. Due to the nature of this sampling practice it is often highly biased. However, the data are used to confirm the occurrence of ore and general diamond drilling assay values and locations. This is a very limited application given the cost to collect the secondary data.

1.2.2 User-weighted systems

In cases where the secondary data is considered of good quality it may be applied to an estimate in a user-defined weighting scheme. The key idea is to weight the secondary data at a fraction of the primary. As an example from the authors experience, given primary and secondary data sets that have a correlation coefficient of 0.5 then a weight of 0.5 may be applied to all the secondary data to account for the interpreted relationship. Obviously, it does not take into account the spatial relationship of individual points and blankets the entire secondary data to a weight of half that of the primary. As well, it does not consider the often complex relationship between primary and secondary data points.

This simplistic weighting system may transfer bias and error to the resultant estimate. There is no accounting for mean value difference and no consideration of primary versus secondary average variance values. The result is an estimate that may contain remnant bias and unknown error.

1.2.3 Cokriging

The most comprehensive method for integrating primary and secondary data in estimation is cokriging. The cokriging system of equations accounts for the particular spatial continuity of the primary and secondary data points. The system of equations chosen for estimation will require primary-primary covariances, secondary-secondary covariances, and primary-secondary cross-covariances. These models can be quite time consuming for the user to develop and this contributes to the general lack of application of these methods in industry. Firstly, the direct covariance equation needs be established for the primary data as would be required for kriging. The covariance function must then be established for each secondary data set that is to be used in assisting the estimate. Finally, the cross-covariance equations must be developed between all data sets. Note that these covariances are inferred by calculating and modelling the corresponding variogram.

To illustrate this one could consider the simple example of a primary variable with one secondary variable set with a single location to estimate as illustrated in Figure 1.1. The point can be estimated as a function of only primary data, as expressed in Equation 1.1. With the addition of secondary data, as expressed in Equation 1.2, there should be improvement over the primary alone. This is the application of cokriging.

$$Z_{\square}* = f_x(Z_{x_1}, ..., Z_{x_{n_x}}) \tag{1.1}$$

$$Z_{\square} ** = f_{xo}(Z_{x_1}, ..., Z_{x_{n_x}}, Z_{o_1}, ..., Z_{o_{n_o}}) \tag{1.2}$$

where $Z_{\square} **$ is an estimate using primary data only, $Z_{\square} **$ is an estimate using primary and secondary data, and x and o are primary and secondary data respectively.

In order to apply the secondary data, direct and cross spatial relationships must be defined. A matrix of these relationships is illustrated in Figure 1.2. An estimate made with primary and secondary data uses such matrices populated

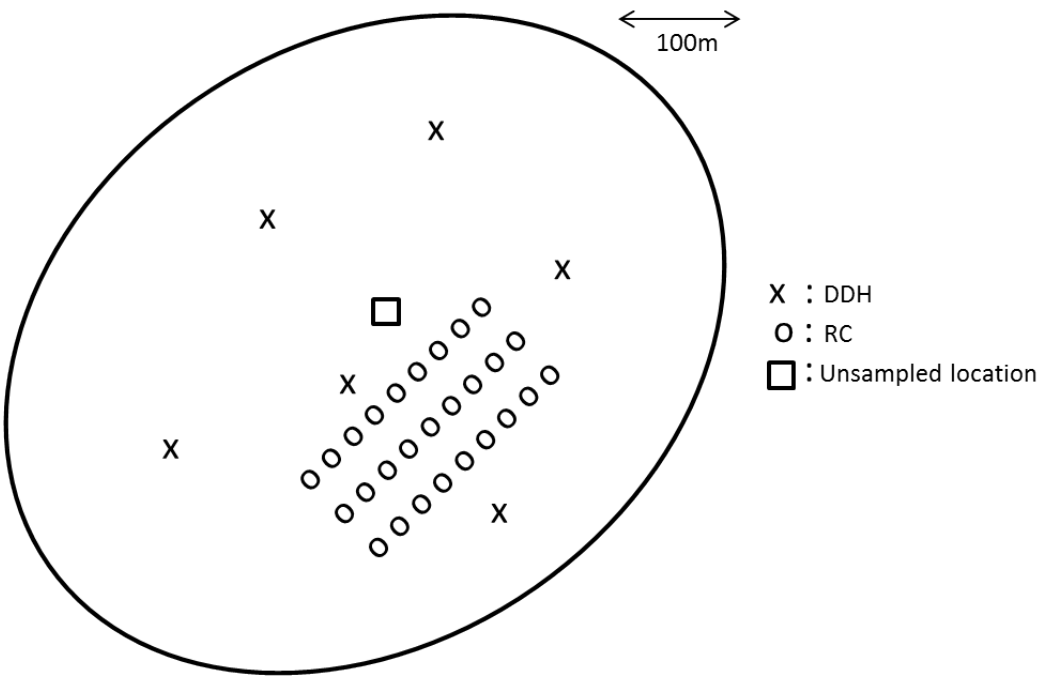

Figure 1.1: An example of primary and secondary data configuration and a considered point estimate. Diamond drill holes (DDH) represent primary data and reverse circulation (RC) drilling represent secondary data

with the relationships for each data point found within the search range of the point being estimated.

1.3 Modelling the data relationship

In order to apply cokriging, a practitioner needs to define the relation between the data types being used. A full understanding may require a geologic history of physical processes that may not be entirely understood under current informational circumstances. However, using available knowledge and spatial correlation techniques this relationship may be modelled.

It is important in any geostatistical analysis that the physical processes that caused the development of the distribution under analysis be explored. These processes, if defined correctly, can provide a great deal of information that may help define the way in which geostatistical modelling proceeds. If, for example, the depositional setting of a deposit is understood a practitioner may be able to get some information from an appropriate analog.

Establishing the correlation between the data types is a main requirement.

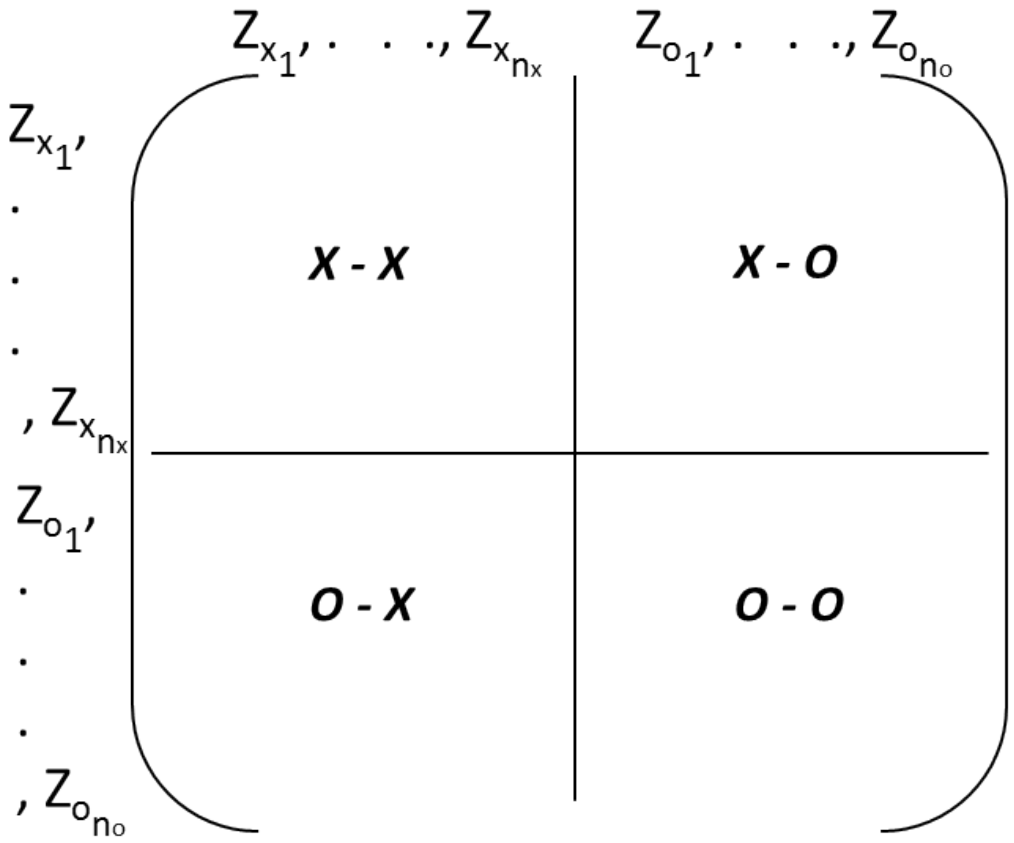

Figure 1.2: An example matrix of contributions from primary and secondary data for estimation where primary(X), secondary(O), and cross-relationships are defined for each data point considered.

This correlation must be above minimum threshold, say 0.2 as suggested by Cuba, Babak, and Leuangthong (2009), to allow the development of a relationship that would ultimately inform the estimates.

A lack of correlation suggests that the secondary data may not be suitable for application in this process. When no relationship exists the estimate reverts to kriging with the primary data only.

One may also consider the rank correlation coefficient to be more robust to outlying data points. Additionally, information on general trends and outliers may be observed from scatterplots of primary and secondary data.

The correlation coefficient is an important summary of the relationship between primary and secondary data. In some cases, namely consideration of

multiple variables from the same sample assays, all data are collocated and the correlation is easily calculated. However, the interesting case of complete non-collocation (heterotopic data) is common with different data types. The relationships must be inferred from the relationships of values separated by some vector distances. This process will be discussed further in Chapter 3.

The relationship between primary and secondary data must then be modelled by an acceptable mathematical function for application in cokriging. This requires variogram or covariance models for both direct and cross relationships. The only acceptable model is the Linear Model of Coregionalization (LMC). The LMC will be reviewed in Chapter 2.

1.4 Cokriging with unequally sampled data

To apply the relationship between primary and secondary data to estimation we use the cokriging formalism. Though there are alternatives, this approach allows the application of all data using the covariances described by the modelled relationships discussed above.

There are several options in cokriging. Each option differs based on the number of secondary data considered and/or the extent to which the primary-secondary data relationship is modelled. Collocated cokriging considers only the secondary data at the point being estimated. Intrinsic collocated cokriging also considers the secondary data at the primary variable locations and applies a scaled model of the same relationship for direct primary, direct secondary, and cross relationships. Full simple cokriging applies a complete model of the data relationship to all data within the search range of the estimate. Ordinary cokriging as well as standardized ordinary cokriging may also be applied, however, due to fundamental weighting restrictions these methods are not recommended.

For the purposes of estimation with unequally sampled data the cokriging method that provides the greatest advantages is full simple cokriging with an LMC. This allows the most accurate portrayal of the data type relationship through the LMC. There is no restriction on the weights attributed to the

secondary data. Bias and error in the secondary is taken into account through the proper application of the cokriging equations and, therefore, not communicated to the estimate of the primary. This will be discussed further in Chapter 3 and demonstrated in the Chapter 5 case study.

Simple cokriging, as with simple kriging, is a a minimum variance estimator. The application of secondary data allows for increased accuracy in point estimates as well as reduced estimation variance.

1.5 Data integration applied to resource classification

When secondary data is applicable in estimation the question arises as to how this additional information assists beyond the simple improvement of the estimates. In situations of collocation there is minimal information gained. However, when data is completely heterotopic then the combination of data influences the data spacing which is important if the data spacing is being used for resource classification.

Resource classification consists of grouping particular portions of a deposit in terms of quality of knowledge about the grade and extent of the resource. In many countries it is a legal requirement for publicly listed companies. The methods for going about this classification are often left to a qualified person(QP) to decide and/or develop. The methods available to perform this classification will be considered in Chapter 4 but many of the regularly practised methods rely heavily on data spacing.

The data spacing of the primary data type is often designed considering the optimal resource definition and economics. The economics of the project considered may require that the overall data spacing is relatively large for definition of the deposit. This may lead to classification of resources both into lower or higher categories. These problems can occur locally and influence mine planning.

The secondary data may fill in the gaps of widely spaced primary data. This information will better inform the resultant estimate. Of course, the utility of the secondary data is contingent on a thoroughly checked cokriged

estimate.

The relationship between a combined primary-secondary data spacing and the average estimation variance will be used to help the classification. The relationship is shown to mimic a reduction in overall data spacing with a power relation to the correlation between data types. This will be discussed in Chapter 4.

1.6 Previous works in unequally sampled data integration for prediction

As mentioned previously, the problem of secondary data type usage is not commonly considered. No National Instrument 43-101(NI43-101) report in the years of 2012 and 2013 considers cokriging (or any multivariate methods) for resource estimation and classification. This may mean that many practitioners avoid the use of secondary data in resource estimation for reporting. The majority of works directly related to cokriging with completely heterotopic data for improved estimates have been produced by members of the Centre for Computational Geostatistics at the University of Alberta-Edmonton and consist of Wawruch et al (2002) Babak and Deutsch (2007), and Minnit and Deutsch (2014). Beyond this, earth science-related studies and extensions to the field of traffic monitoring (Zhao et al., 2006) and bio-medical concerns (Huang, 2014) produce a very few studies applying artificial neural networks (Besaw and Rizzo, 2007), kernel estimation (Sanchez-Borrego et al., 2014), and cokriging. Most apply methods to homotopic and partially heterotopic data.

Kriging methods, as related to spatial interpolation and extrapolation were outlined by Gandin (1963) in reference to meteorological estimation. The methods were defined as optimal extrapolation. These equations and the development mimic those put forward by Matheron (1963). Matheron also developed these equations with the cokriging extension to the multivariate. The estimator was defined in a general form where each estimate is based on weighting multiple data values.

Details of the cokriging methods are shown by Journel and Huijbregts (1978) where, along with well-defined matrix development, the proposition is put forward that the use of cokriging methods are best employed when one variable is under-sampled with reference to a second variable. This idea is easily understood in consideration of the volume of work required to develop the covariance matrix. Myers (1984) displays a method for equation reduction related to the matrix of equations designed to reduce computational cost. This, however, is no longer of practical concern due to exponential increase in cheaply available computing power.

The extended requirements of the cokriging system over the single-variable kriging system was shown by Deutsch and Journel (1992) where two points were made clear. First, the extension to cokriging requires the definition of many covariance functions. Secondly, matrix notation is required. As well, these authors discuss the subject of the unbiasedness conditions of cokriging. Matheron (1971), Journel and Huijbregts (1978), Gandin (1963) and others reference ordinary cokriging and simple cokriging. Ordinary cokriging (the type most often referenced in previous works) forces the primary data weights to sum to one while secondary data weights sum to zero. Standardized ordinary cokriging, as put forward by Isaaks and Srivastava (1989) refers to the standardization of all variables and constraining the resultant cokriging weights to sum to one. These restrictions, however, limit the impact of the secondary data. Simple cokriging does not restrict the weighting and the optimal weights are extracted from the primary-secondary spatial relationship.

Overviews of cokriging can be found in Gotway and Hartford (1996), Goovaerts (1997), Deutsch and Journel (1992), and more completely in Rossi and Deutsch (2014). For a more stringent definition with derivations one may look to Journel and Huijbregts (1978).

The focus of this work is in the context of applying simple cokriging for estimation with unequally sampled data types. Previously Wawruch, Deutsch, and McLennan (2002) proposed methods to approach the assessment of the relationship between non-isotopic data types. In their work a completely isotopic data set was divided into two unequally sampled data sets that occupied

the same domain. It was known that the correlation between the two data sets should be 1.0. From this the methods put forward could be compared with the real values left aside. Primarily, the authors approached the problem through extrapolation. This required the construction of the direct variograms for the primary and secondary data types as well as construction of the cross-covariance at minimum sample spacing. The cross-covariance for distances less than the minimum sample spacing was extrapolated with consideration to the known cross-covariance shape and with the practitioners knowledge of the underlying geological structure. Nugget effect(NE) values were calculated as an average from the direct variograms. These methods were shown, with consideration to uncertainty in the various portions of extrapolation, to have very good correlation reproduction. Continuing in this line, Minnitt and Deutsch (2014) applied similar processes in two artificial studies. In this case, the studies were applied in validation of the simple cokriging approach and so comparisons were made between the original underlying distribution, ordinary kriged estimates, and simple cokriged estimates at varying block sizes. A notable difference here is the assertion that, in extrapolation of the cross covariance, the nugget effect plays no role in final cokriging estimation and can be set to zero for the cross covariance modelling. The result was a notable advantage in application of cokriging over ordinary kriging. This was displayed, not so much in average tonnage and grade prediction but in reference to 'ore, waste, ore-loss, ore-gained' cross plots. An important point was that these studies displayed the high quality of estimates produced with carefully constructed cross-covariance values.

Wackernagel (2003) provides background on heterotopic cokriging but suggests the impossibility of application in the face of completely heterotopic data sets. This is due to perceived inability to calculate cross variograms or cross covariances.

Prior to these studies there are many papers critical of a full Linear Model of Coregionalization for primary-secondary data co-estimation. Odeh, McBratney, and Chittleborough (1995) compared the utility of these methods to that of regression-kriging. The case considered soil property prediction using two

unequally sampled data sets. Regression-kriging was shown to have a slight predictive edge over cokriging. The limitation is that some collocated data are required. This eliminates regression-kriging as an option for heterotopic data sets. A similar soil property predictive study was performed by Knotters, Brus, and Oude-Voshaar (1995) with the same overall result.

Ver Hoef and Barry (1998) also showed the advantage of the cokriging approach over ordinary kriging. Here the authors suggested poor development of cross variograms could be overcome by fitting moving average functions and applying the weighted least squares method for fitting to data. A limitation exists in the prediction of variogram parameters. As well, this method also requires some equally sampled data. The same applies to Chica-Olmo (2007) who applied cokriging as a predictor of housing location price for real estate concerns. Partial equal sampling was required in order to establish cross-variograms.

Beyond direct comparisons using cokriging there are other methods applied for data integration. For example Zhao et al (2006) apply the Markov theorem in a Gaussian framework in order to integrate two types of traffic flow information of differing quality. Though the authors suggest the methods work well in practice, this method, as the methods above, requires some equal sampling of the spatial data.

Besaw and Rizzo (2007) approach data integration through application of an Artificial Neural Network(ANN). While the approach put forward allows a novel option for integration of both continuous and categorical data without a model of coregionalization the ANN requires a known output with which to train the ANN. This is limiting when collocated correlation is not known.

1.7 Thesis outline

Chapter 2 will develop the theory required for the LMC and cokriging. Chapter 3 will review the process of data integration using cokriging and issues and assumptions therein. Bias and error as well as results checking are discussed. Chapter 4 considers resource classification in terms of guidelines, current prac-

tice and the calculation of data spacing in the presence of multiple data types. A case study will be presented in Chapter 5 that tests the application of the simple cokriging methods as well as the extension to 'effective' data spacing. Chapter 6 presents conclusions developed from the results of the associated case study and considers possible directions for future research and development. Updates to the GSLIB cokriging program are described in Appendix A while estimation parameters are shown in Appendix B.

Chapter 2

Theory of estimation with multiple data types

The theoretical foundations of this thesis can be divided into four sections. These include (1) the notion of covariance, correlation and the variogram, (2) the development of kriging for optimal local estimation, (3) modelling the relationship between coregionalized variables, and (4) the extension of optimal estimation to multiple variables (cokriging).

The random function is an important concept applied in geostatistics and throughout the following sections. This is the idea that multiple regionalized random variables ($Z_k(\mathbf{u})$) that are defined throughout a domain(A) is defined by:

$$Z_k(\mathbf{u}) = [Z_k(\mathbf{u}), \forall \mathbf{u} k = 1, ..., K, \mathbf{u} \in A] \tag{2.1}$$

where k defines a specific random variable and $\mathbf{u}$ is a data location.

The correlation between any two random variables depends on the variables and the vector($\mathbf{h}$) that separates the two locations, e.g. $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$ (Journel and Huijbregts, 1978).

2.1 Stationarity

Stationarity is the set of assumptions that allow the consideration of a probabilistic interpretation of regionalized variables as a random function.

One aspect of stationarity is the choice of mineralized units, or domains, that have similar geological controls. Another aspect of stationarity is that

the basic statistics within these domains, such as the mean, variance, and spatial correlation, are invariant of location. These assumptions are related to the degree of homogeneity of the domain under investigation (Journel and Huijbregts, 1978).

A stationary first moment or mean (μ) over a deemed stationary domain or study area, A, is represented by:

$$E[Z(\mathbf{u})] = \mu(\mathbf{u}) = \mu, \forall \mathbf{u} \in A \tag{2.2}$$

The second order moments of variance, covariance, and variogram are written as:

$$E[[Z(\mathbf{u}) - \mu(\mathbf{u})]^2] = \sigma^2(\mathbf{u}) = \sigma^2, \forall \mathbf{u} \in A \tag{2.3}$$

$$E[Z(\mathbf{u}+\mathbf{h})Z(\mathbf{u})] - [\mu]^2 = C(\mathbf{h}), \forall \mathbf{u}, \mathbf{u}+\mathbf{h} \in A \tag{2.4}$$

$$Var[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = 2\gamma(\mathbf{h}), \forall \mathbf{u}, \mathbf{u}+\mathbf{h} \in A \tag{2.5}$$

Practical studies attempt to define domains where the location independence of these statistical moments is reasonable.

2.2 Correlation

One of the primary requirements of a multiple data type estimate is the need to establish the nature of the relationship between data types.

2.2.1 The variogram

The variogram is commonly used to quantify spatial variability (see Equation 2.5). The variogram is defined as:

$$2\gamma_{k,k'}(\mathbf{h}) = E[[Z_k(\mathbf{u}) - Z_k(\mathbf{u}+\mathbf{h})][Z_{k'}(\mathbf{u}) - Z_{k'}(\mathbf{u}+\mathbf{h})]], k, k' = 1, ..., K \tag{2.6}$$

where, by convention, z_k is the primary variable, z'_k represents a secondary variable, $\mathbf{h}$ is the lag separation vector, and $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$ are location vectors.

The difference between Equations 2.5 and 2.6 is that the former is for a direct variogram and the latter for direct ($k = k'$) and cross ($k \neq k'$) variograms between multiple variables.

It should be noted that practitioners use the terms variogram and semi-variogram interchangeably. The '2' is important for numerical correctness.

The variogram explains the variability observed within the domain as a function of increasing lag separation distance. Figure 2.1 shows that as lag distance increases the experimental variogram value often increases to a maximum variance, or sill. Such a well behaved variogram is not always the case and deviations may suggest a lack of data and/or underlying trends within the data. For in-depth discussion of variograms and construction details look to Rossi and Deutsch(2014, pg.99-110)

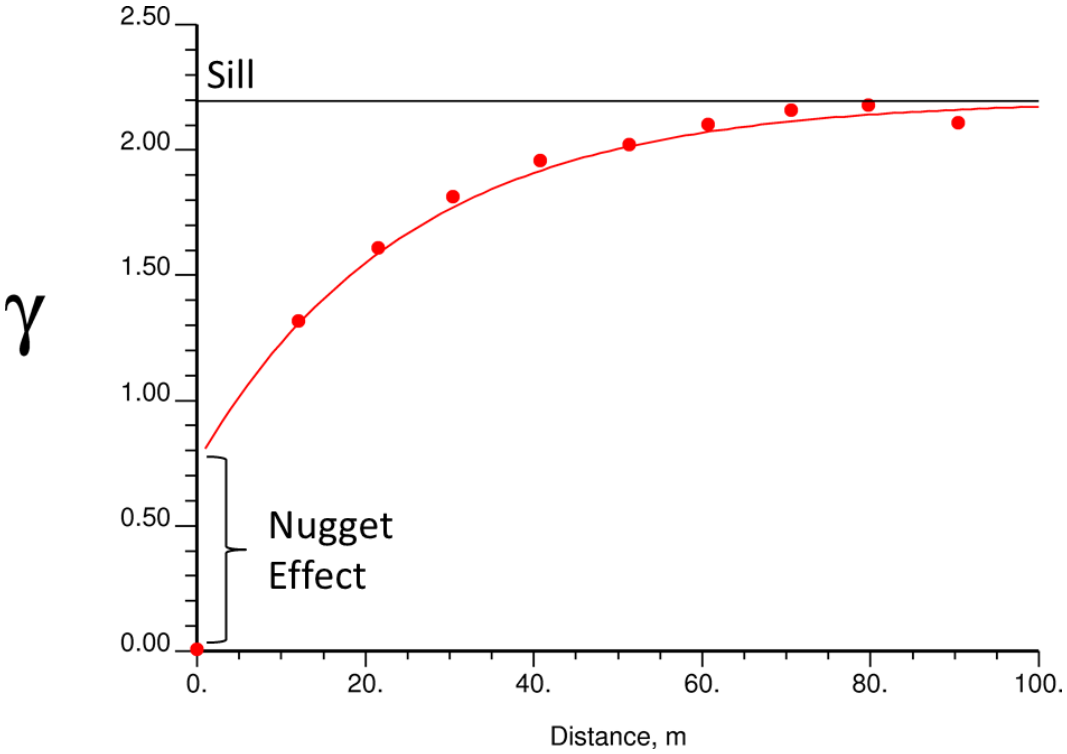

Figure 2.1: An example of a well-behaved experimental variogram plot (the dots) with overlaid model (the line)

By establishing the experimental points the relationship can then be modelled as a function or combination of functions. These functions, or structures, quantify each significant portion of the variance from nugget effect, to short

range, to long range.

2.2.2 The covariance relationship

A related measure of spatial relationship is the covariance. The covariance is widely used in kriging (Deutsch and Journel, 1992). The covariance is a measure of linear dependence between variables.

$$C_{k,k'}(\mathbf{h}) = E[Z_k(\mathbf{u})Z_{k'}(\mathbf{u}+\mathbf{h})] - \mu_{z_k}\mu_{z_{k'}}, k, k' = 1, ..., K \tag{2.7}$$

where k and k' signify the two variables under consideration. In cases of unequal sampling the cross-covariance is applicable since the cross-variogram cannot be calculated without equal sampling (Wawruch,2003).

The variogram and covariance can be related to each other under an assumption of first and second order stationarity:

$$\gamma_{k,k'}(\mathbf{h}) = C_{k,k'}(0) - C_{k,k'}(\mathbf{h}), k, k' = 1, ..., K \tag{2.8}$$

It is somewhat confusing to be switching back and forth between variogram and covariance, but it is important in terms of heterotopic data. Without collocation the cross-variogram cannot be calculated. The sill, or cross-covariance at lag $\mathbf{h} = 0$, $(C_{k,k'}(0))$, can only be calculated with equally sampled data. However, the cross-covariance at lags greater than 0 can be calculated and then it is possible to extrapolate C(0). This allows inference of the cross-variogram if needed.

2.2.3 The correlation coefficient

Another measure of dependence is the coefficient of correlation, ρ. The correlation coefficient is the standardized cross-covariance:

$$\rho_{k,k'}(\mathbf{h}) = \frac{C_{k,k'}(\mathbf{h})}{\sigma_k \times \sigma_{k'}}, k, k' = 1, ..., K \tag{2.9}$$

A small correlation between data types suggests there will be little gained from developing the relationship required for cokriging. However, significant

correlation indicates the possibility of improvement. The range of correlation that these decisions are based on is not well defined. Minnitt and Deutsch(2014) suggest a correlation coefficient of less than 0.4 is poor quality while above 0.7 should be of significant quality. Cuba, Babak, and Leuangthong(2009) provide a slightly larger range with a lower end of 0.2. A correlation of greater than 0.8 may allow primary and secondary data types to be directly combined. This would bypass the issue of cross variograms and cokriging.

2.3 Kriging

Kriging is a well known estimation method that considers the variogram, covariance, or correlation measures in defining the variability of a random function. Kriging is a minimum error variance linear estimator that is exact, that is, an estimate at a data location will return the known data value. Also, the linear weights defined by kriging depend only on the data configuration and not the data values (Isaaks and Srivastava, 1989).

2.3.1 Normal equations

The kriging equations were developed in parallel by Gandin(1963) and Matheron(1963) under the name of optimal interpolation and kriging, respectively. These equations are modifications of the least squares linear regression model based on the Gauss-Markov Theorem (Surumi,2015). The equations allow the minimization of variance in estimation.

2.3.2 Kriging types

There are many modifications to kriging that allow for different estimators. Those required for this dissertation include simple kriging and ordinary kriging.

Simple kriging(SK) is an unconstrained version of kriging. The simple kriging estimator is defined as:

$$z^*_{sk}(\mathbf{u}) - \mu = \sum_{i=1}^{n} \lambda_i^{sk}[z(\mathbf{u_i}) - \mu] \tag{2.10}$$

where $z^*_{sk}(\mathbf{u})$ is the estimate at the unsampled location $\mathbf{u}$ and λ^{sk}_i represents the unconstrained weight applied to the residuals of the data value at any location $\mathbf{u_i}$. The mean value of the domain is assumed to be stationary and known based on the data values within the domain.

The weights applied in estimation are calculated to minimize estimation variance with the well known kriging equations:

$$\sum_{j=1}^{n} \lambda^{sk}_j C(\mathbf{u_i}, \mathbf{u_j}) = C(\mathbf{u_i}, \mathbf{u}), i = 1, ..., n \tag{2.11}$$

The minimized estimation variance, or the simple kriging variance, is defined as:

$$\sigma^2_{sk} = C(0) - \sum_{i=1}^{n} \lambda^{sk}_i C(\mathbf{u_i}, \mathbf{u}) \tag{2.12}$$

The kriging variance decreases as additional data are considered.

Ordinary kriging(OK) is a form of kriging that considers the mean value of the function to be unknown (Journel and Huijbregts, 1978). In this case, the kriging weights are constrained to equal 1.0 for unbiasedness. The OK estimator is:

$$z^*_{ok}(\mathbf{u}) = \sum_{i=1}^{n} \lambda^{ok}_i z(\mathbf{u_i}) \tag{2.13}$$

$$\sum_{i=1}^{n} \lambda^{ok}_i = 1.0 \tag{2.14}$$

which, when applied to point estimation within varying search neighbourhoods, amounts to a re-estimate of the mean value that is dependant on location and somewhat non-stationary. This is a widely used estimation algorithm in the mining industry as evidenced by the abundance of recent NI43-101 statements reviewed for this thesis.

There are many variants of kriging including universal kriging, kriging with external drift, factorial kriging, indicator kriging, disjunctive kriging, and multi-Gaussian kriging. These methods differ in their consideration of different constraints and/or different transforms of data.

In the simplest case of a completely second order stationary random function, SK will provide the best estimate with minimum error variance. When the stationarity of the random function is suspect, which is often the case in natural processes, OK is widely used in practice to account for the variability in the mean value across the domain.

2.4 Linear model of coregionalization

Modelling the spatial relationship for one variable can be accomplished using the variogram, covariance, or correlation. A variogram represents a model of regionalization. When considering more than one variable a model that accounts for the relationship between variables is required. This entails modelling of the direct spatial relationship for each data type as well as the spatial relationship between data types. The number of variograms that must be modelled is given by $K(K+1)/2$ where K is the number of variables being considered. As this number gets larger (greater than 3) the requirements of modelling so many variograms manually becomes untenable. Spatial anisotropy is also a concern in most studies. This increases the number of variograms required by three times to reflect each of the three principal spatial dimensions (Leuangthong et al., 2008).

2.4.1 LMC structure

The LMC consists of an organized combination of direct variograms as well as cross variograms. However, covariances are most often applied over variograms in matrix calculation because they allow the development of the cross-covariance without collocation. The LMC is defined by the set of known structures (e.g. spherical, exponential, Gaussian...) that model the shape of the variogram. Each of these structures accounts for part of the of variance or covariance. In the same way, the cross variograms are modelled with the same set of known structures:

$$\gamma_{k,k'}(\mathbf{h}) = \sum_{l=0}^{L} b^l_{k,k'} \Gamma^l(\mathbf{h}), k, k' = 1, ..., K \tag{2.15}$$

where b represents the portion of variance/covariance, Γ defines the structure type for $l = 0, .., L$, where the $l = 0$ structure is the nugget effect by convention. Also, k and k' are indices for the $1, ..., K$ variables.

Each of the L+1 matrices of b coefficients $(K \times K)^2$ must be positive definite to ensure that the final LMC is positive definite.

The practitioner is advised to standardize their data prior to modelling variograms. Standardization of data refers to:

$$Z_k^{std} = \frac{Z_k - \mu_k}{\sigma_k} \tag{2.16}$$

Development of the LMC is much easier when all of the variances have been standardized to one and the sills of the cross variograms are the correlation coefficients between the data types.

2.4.2 Assumptions and requirements

There are several assumptions underlying the LMC. One must assume that each variable is a linear combination of $L + 1$ common factors. The weights of these factors define how each variable is related to the other (Leuangthong et al., 2008).

Additionally, the variograms may be fit less well than a simple direct fitting because of the additional constraints.

2.4.3 Small example

Expanding on the previous example could help understand the LMC. An LMC with two data types consists of three equations. Again the cross variogram is addressed first. This follows the standard LMC structure and assumes standardization of the data. Consider a cross-variogram with a sill value of 0.50. We set the nugget effect structure to zero as it will not be seen with unequally sampled data. Consider two spherical structures. The first structure accounts for 0.30 and the second for the remainder to a sill of 0.5, the cross variogram equation will be:

$$\gamma_{k,k'}(\mathbf{h}) = 0.0 + 0.3Sph_1(\mathbf{h}) + 0.2Sph_2(\mathbf{h}) \tag{2.17}$$

The direct variogram of the first variable "k" shows a nugget effect 0.20, the first structure accounts for 0.55, and the second 0.25. These values sum to a unit variance and result in:

$$\gamma_{k,k}(\mathbf{h}) = 0.2 + 0.55Sph_1(\mathbf{h}) + 0.25Sph_2(\mathbf{h}) \tag{2.18}$$

Next, assume that a second variable "k'" shows a nugget effect of 0.1 and contributions of 0.5 and 0.4 for structures 1 and 2. The result is:

$$\gamma_{k',k'}(\mathbf{h}) = 0.1 + 0.5Sph_1(\mathbf{h}) + 0.4Sph_2(\mathbf{h}) \tag{2.19}$$

The modelled variogram describing the LMC can be seen in Figure 2.2. The resulting matrices would be:

$$\begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.10 \end{pmatrix} \begin{pmatrix} 0.55 & 0.30 \\ 0.30 & 0.50 \end{pmatrix} \begin{pmatrix} 0.25 & 0.20 \\ 0.20 & 0.40 \end{pmatrix} \tag{2.20}$$

where each array represents the corresponding matrix of variance contributions.

Checking the determinant of each matrix is a quick check of the validity of the LMC. The difficulty arises when the sill of the cross variogram is closer in value to the sill of the direct variograms. This means that all of the variograms converge to a single model and there are few degrees of freedom in fitting.

2.4.4 Application to estimation

Application of the LMC to the cokriging system is performed through covariance matrices. However, the LMC is generally developed using variograms. The variogram values may be quickly converted to covariances via the relation:

$$\gamma_{k,k'}(\mathbf{h}) = C_{k,k'}(0) - C_{k,k'}(\mathbf{h}) \tag{2.21}$$

where $C_{k,k'}(0)$ is the sum of the $\Gamma_{k,k'}$ values from 1 to L, assuming that the $l = 0$ value is zero. The LMC can be developed using covariances. This is required in terms of cross covariance delineation with completely unequally sampled data.

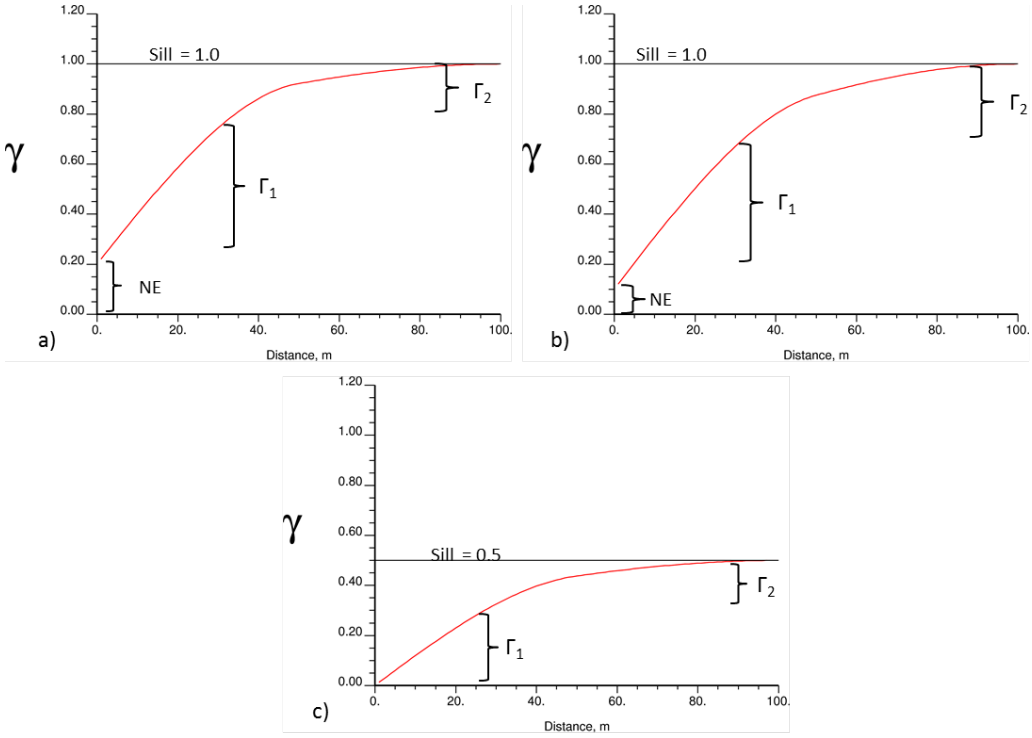

Figure 2.2: An example LMC showing the structure of a) the direct primary variogram, b) the direct secondary variogram, and c) cross-variogram

The covariance matrices coupled with the LMC equations provide the required information to the cokriging system. The cokriging weights applied to each known datum are dependant on the location being estimated and the search parameters.

2.5 Cokriging

The cokriging system is the multivariate extension of the kriging system. This increases in the complexity of nomenclature, but is essentially the same as kriging. This is displayed in Figure 1.2.

One potential problem in the application of cokriging is the screening effect. Secondary data in the vicinity of primary data can have negative weights applied. This may cause unstable estimates.

2.5.1 Simple cokriging

Simple cokriging(SCK) places no restriction on the weights in estimation. The original requirements remain in terms of positive definiteness of covariance matrices. This is not a concern if the LMC is valid. Simple cokriging minimizes estimation variance. The simple cokriging estimator is:

$$z_1^*(\mathbf{u}) - \mu_{z_1} = \sum_{i=1}^{n_{z_1}} \lambda_i(\mathbf{u})[z_1(\mathbf{u_i}) - \mu_{z_1}] + \sum_{k'=2}^{K}\sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j}(\mathbf{u})[z_{k'}(\mathbf{u_{k',j}}) - \mu_{z_{k'}}] \quad (2.22)$$

with the cokriging weights calculated by:

$$\sum_{j=1}^{n} \lambda_j C(\mathbf{u_i}, \mathbf{u_j}) + \sum_{k'=2}^{K}\sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j} C(\mathbf{u_i}, \mathbf{u_{k',j}}) = C(\mathbf{u_i}, \mathbf{u}), i = 1, ..., n \quad (2.23)$$

$$\sum_{j=1}^{n} \lambda_j C(\mathbf{u_i}, \mathbf{u_{k',j}}) + \sum_{k'=2}^{K}\sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j} C(\mathbf{u_{k',i}}, \mathbf{u_{k',j}}) = C(\mathbf{u_{k',i}}, \mathbf{u}), i = 1, ..., n_{z_{k'}} \quad (2.24)$$

and the estimation variance:

$$\sigma^2_{sck} = C_{z_1 z_1}(0) - \sum_{i=1}^{n_{z_1}} \lambda_i(\mathbf{u}) C_{z_1 z_1}(\mathbf{u_i} - \mathbf{u}) - \sum_{k'=2}^{K}\sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j}(\mathbf{u}) C_{z_1, z_{k'}}(\mathbf{u_{k',j}} - \mathbf{u}) \quad (2.25)$$

The difference between simple kriging and simple cokriging equations is the heavier notation required to index the additional data.

2.5.2 Ordinary cokriging

Ordinary cokriging is the multivariate extension of ordinary kriging. There are further conditions placed on the weights of estimation. The primary data weights must sum to one and the secondary weights must sum to zero. This is known as traditional ordinary cokriging. The restriction on secondary weights can cause the marginalization of the secondary data.

Standardized ordinary cokriging requires the transformation of the secondary variable to have the same mean as the primary variable. In this case, all weights, primary and secondary, must sum to one.

The standardized ordinary cokriging estimator is:

$$z_1^*(\mathbf{u}) = \sum_{i=1}^{n_{z_1}} \lambda_i(\mathbf{u}) z_1(\mathbf{u_i}) + \sum_{k'=2}^{K} \sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j}(\mathbf{u}) [z_{k'}(\mathbf{u_{k',j}}) + \mu_{z_1} - \mu_{z_{k'}}] \tag{2.26}$$

under the condition:

$$\sum_{i=1}^{n_{z_1}} \lambda_i(\mathbf{u}) + \sum_{k'=2}^{K} \sum_{j=1}^{n_{z_{k'}}} \lambda_{k',j}(\mathbf{u}) = 1 \tag{2.27}$$

and the estimation variance, σ^2_{cok} as defined in Equation 2.22.

In review of NI43-101 reports from 2012 and 2013 no application of ordinary cokriging was seen.

2.5.3 Further cokriging options

Variants of the cokriging system have been developed as practical short-cuts to the process of fitting an LMC. These include collocated cokriging, Bayesian updating, intrinsic cokriging, and others. These short-cuts can make the application of cokriging much easier due to the fact that a full LMC need not be developed. The primary direct covariances and some measure of cross correlation are required.

Many of these methods require that the secondary data are at the same location as the primary, but this means the secondary data will be marginalized by the collocated primary. These methods are commonly applied when the secondary is exhaustive, e.g. geophysical measurements of a related variable.

This thesis considers no collocation between primary and secondary data. This leads to cokriging with a full LMC.

Chapter 3

Data integration

Integration of primary and secondary data is not a straight-forward process. Practitioners can choose solutions that are relatively easy to implement but may lead to poor estimates. These include user-defined weighting systems and simply ignoring the difference between primary and secondary data. In many cases data integration is better achieved by correctly considering the spatial correlation for all data types with cokriging.

3.1 Correlation between data types

The method advocated here is simple cokriging with a full LMC. In order to employ this method a degree of correlation between data types is needed. This correlation defines the quality of secondary information that is applied to the estimate.

3.1.1 Defining correlation between data types

The correlation between primary and secondary data summarizes the relationship. In situations of complete homotopy the correlation is easily calculated; there is a primary-secondary data pair at each data location. When the data types show partial unequal sampling the correlation can still be calculated from the collocated pairs of primary and secondary data.

A problem occurs when the data types are completely heterotopic. In this case the collocated correlation is unknown. One can calculate a correlation coefficient of pairs within a certain lag distance tolerance. The variance at

distances less than the chosen lag tolerance is unknown. The correlation between data types may be artificially understated and the importance of the secondary data may be understated.

The solution suggested by Wawruch, Deutsch, and McLennan (2002) lies in the cross-covariance. When looking at the the cross variogram under complete unequal sampling there is no direct determination of the cross variogram sill or $C_{k,k'}(0)$, where k is the primary variable and k' is the secondary variable. The approach is to calculate experimental cross covariance points instead of the experimental cross variogram points. The experimental cross covariance should converge to 0.0 at the maximum range of correlation. The cross covariance can be extrapolated back to the origin, $C(0)$. This can be done by interpreting the shape of the covariance curve as well as applying any knowledge of the underlying process.

Once $C_{k,k'}(0)$ is estimated the maximum covariance between data types is defined. The correlation can be established by applying Equation 2.8. This is displayed in Figure 3.1 where the standardized $C_{k,k}(0)$ can be compared to the extrapolated standardized $C_{k,k'}(0)$.

The concern is the unknown nugget effect on the cross variogram/covariance that contributes to the collocated correlation. In a small study of the cross variogram nugget effect this author selected five collocated groups of data each with two variables. In collocation the nugget effect can be defined with certainty. In this way the actual nugget effect on the cross variogram is known and compared to that found from the direct variograms. It is noted that there seems to be no repeatable relationship between direct and cross nugget effects. While we know the cross variogram/covariance nugget effect exists we will assume that it is zero for this method of $C(0)$ extrapolation. This assumption is reasonable since the nugget would only enter into calculations when estimating exactly at the location of a secondary data.

3.1.2 Correlation requirements for cokriging

When the correlation between the primary and secondary data is zero then the secondary data contributes nothing to estimation. Practitioners may see

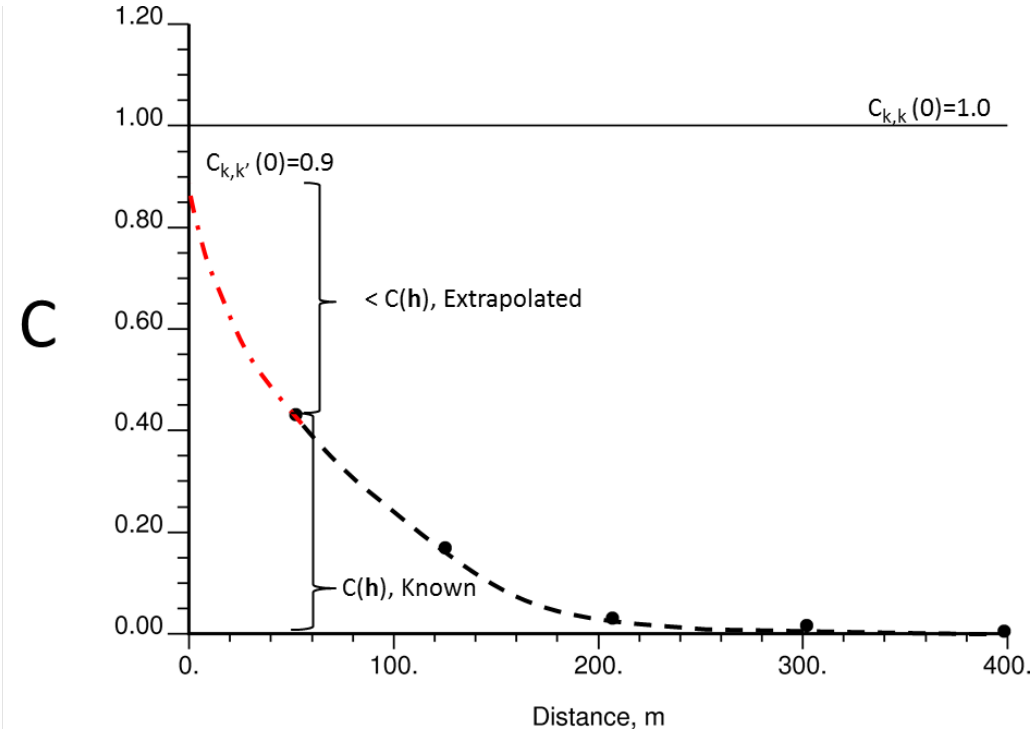

Figure 3.1: Extrapolation of the cross-covariance back to collocation through continuation of the covariance curve.

little positive outcome for the amount of work required to develop a full LMC when the correlation is close to zero. As noted by Minnitt and Deutsch (2014) minimal application quality may be signified by a correlation coefficient below 0.4 while Cuba, Babak, and Leuangthong (2009) suggest a lower limit of 0.2. A correlation above 0.8 could indicate that the data types should be considered equal. This window represents the range in which cokriging with a full LMC may have a significant effect on estimates.

The lower and upper range of correlation mentioned above are not definitive values. This is considered a soft range that could bear greater exploration.

Although only positive correlation has been discussed this can be applied to negative correlation as well. A large degree of correlation, positive or negative, shows the possible utility of secondary data in estimation.

These specific issue will be addressed in the case study in Chapter 5.

3.1.3 Applying correlation to the decision of data type integration

Applying the extrapolation method provides a better definition of collocated correlation than pairing close data. In this case estimated collocated correlation can be applied to assist in the decision of data integration.

A practitioner may decide to discard the secondary data or integrate it directly with the primary due to high positive or negative correlation. It is advised to also apply the cokriging comparison used here. Due to the inexact range of applicable correlation the secondary can provide more assistance than expected.

3.2 Problem of bias and error in secondary data

Bias and error in data can introduce increased variance to analysis and prediction of the regionalized variable. This is of particular concern in situations of unequal sampling as bias and error in secondary data is somewhat more difficult to quantify.

3.2.1 Sources of bias and error

The sources of bias and error in secondary data are related to several issues. These include the data collection process, analysis of the samples, interpretation of the samples or data cleaning prior to use.

Bias and error may develop in the sample collection process for several reasons. Primarily the issue lies in the lack of a quality sampling procedure. When the sampling scheme involves human interaction there is a great opportunity for the sampler to introduce bias and error. An ill-conceived sampling method can introduce bias and error by design. The example is chip sampling expanding on that mentioned in Chapter 1. The process may involve multiple geologists with differing assessments. Therefore, intervals defined based on 'significant' grade or lithology change may not be consistent. The term 'significant' is itself ambiguous. This relies on consistency in the visual definition of grade and contacts. Even in the case of a single geologist bias and error can be introduced through the process and visual inconsistency. There is an

opportunity for correction or calibration of visual grade estimates by comparing with returned assays. This may increase accuracy in the future. However, contacts can be gradational and sample division can have a significant affect on assays.

Sample analysis is of concern but problems can be easily avoided with proper methods. Most often a bias introduced during analysis is a process error, mechanical or human. Most labs have well defined methods and procedures but human error can still be an issue. Mechanical failure can also be a concern. One partial solution is continual running of check and blank samples to ensure the assay process is proceeding as designed. Cokriging methods will also improve bias and error filtering assuming only the secondary is affected. Bias and error introduction during this stage can be avoided or limited to minimal spoiled samples.

Finally, bias and error may be introduced by the practitioner in preparation of the data for analysis. Samples may be modified or removed if the practitioner believes there is cause to do so. An example is the application of a cut-off value in a high tail metals deposit such as Carlin-style gold. The decision should be designed to limit the impact of a few very high samples. Also, outlying samples that are considered to be in error may be removed from data sets. When these decisions are made in error the user can impose bias and error on an otherwise appropriate data set. It is important for the practitioner to maintain caution in the data cleaning process.

3.2.2 Bias and error identification

Bias and error in secondary data can be defined by comparison to primary data. Both data types should be standardized to allow for direct comparison. This is important to correct bias in the secondary data. Error is accounted for in the cross covariance.

Bias and error are defined in separate ways. Bias in the secondary can be defined as the mean difference between the primary and secondary data. Error could be thought of as unbiased random noise added to the true underlying grade. Error does not change the covariance, but increases the variance of the

secondary data.

3.2.3 Accounting for bias and error in estimation

Bias and error can be accounted for easily in the full cokriging process. This is an attractive features of simple cokriging with an LMC.

The estimates are calculated on residual values. The last step for each point estimate is to reverse standardization. In this way the bias in the secondary data is not communicated to the final estimate.

Error in the secondary is filtered through the proper modelling of the relationship between data types. The filtering is done through application of the cross covariance where the weights applied to secondary data are restricted based on the cross covariance equation. Secondary weights depend heavily on $C_{k,k'}(0)$.

3.3 Checking results

Estimation results must checked. In a theoretical exercise such as that performed by Minnitt and Deutsch(2014) this is not an issue. The underlying distribution from which each data type was sampled is known. In practical situations the estimates must be compared against known data points in order to confirm that known values are reproduced.

3.3.1 Cross-validation

A practical check is cross-validation. Each known primary data point should be estimated without bias and with reasonable error. The actual values should be plotted against the estimates and the result should be checked for bias and error.

A jacknife cross-validation may also be applied when additional data is available in the domain being considered. Jacknife allows for the comparison with alternate data sources. The results of the estimate can be compared to the alternate known data values. Large deviation from the known values may suggest an implementation issue in the estimate.

3.3.2 Model comparison

An OK estimate using the primary data only allows for the direct comparison of estimates and may provide a gauge of the advantage provided by SCK. This is done for the case study in Chapter 5.

The models may be compared in terms of cross validation(Figure 3.2a) and jackknife(Figure 3.2b) results and directly in view of estimation variance. It is possible that, despite the time spent modelling the LMC, the cokriged estimate may show little to no advantage over the OK estimate. When looking at Figure 3.2 it is noted that cross validation shows fairly good reproduction for the example OK estimate. However, the jacknife cross validation with alternate data from the same domain shows poor reproduction. This is a possibility in jacknife usage because the alternate data set used for the cross validation may contain bias and error that is not account for.

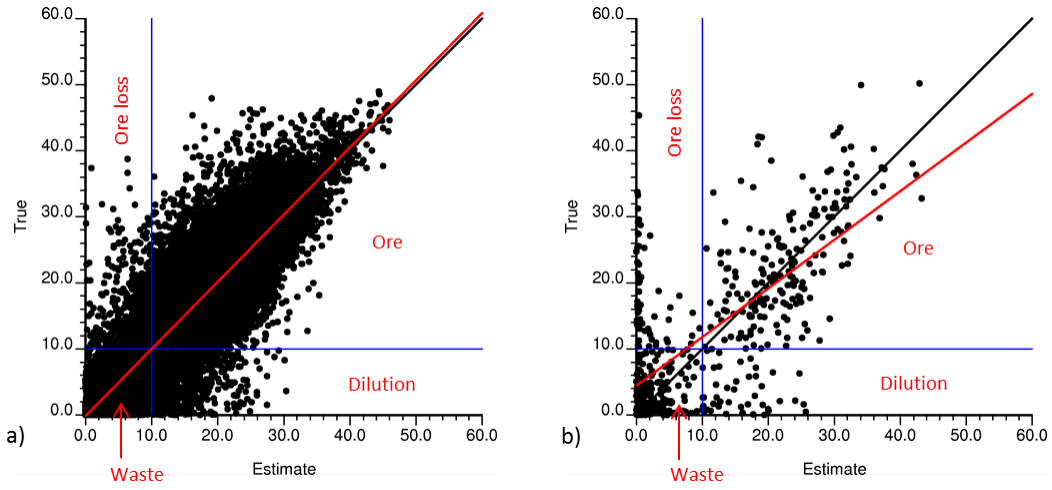

Figure 3.2: An example of cross validation as performed between estimates and original known values with a) direct cross validation using primary data and b) jacknife cross validation with secondary data. Note the cut off lines that divide each plot into four quadrants indicating the quality of each value in terms of production value. The red lines in each figure represent the slopes of regression, .

3.3.3 Mock data and comparison

Also considered in the case study by Minnitt and Deutsch(2014), production cut-off values allow for good comparison of estimates. They can be compared in terms of 'ore-waste-ore loss-dilution'. The categories are based on the min-

imum economic grade value of the deposit. This is seen in Figures 3.2 *a* and *b* where the plotted cut-off grade, ten percent in this case, divide the plot into four quadrants. The percentage of metal can then be defined for each category. Categories are then compared for SCK and OK and should display discrepancies between the two estimates. This can provide a clearer display of advantage when data and histogram reproduction does not show a clear distinction between estimates.

Chapter 4

Resource classification with secondary data

Resource classification is a practical concern for companies in the resource industry, particularly those that are publicly listed. Guidelines for reporting of natural resources have developed independently but most follow a similar framework. These guidelines are relatively general leaving method selection and burden of proof to the practitioner.

When multiple data types are used together and result in an estimate with reduced estimation variance an extension can be made to resource classification. The approach here is to establish an effective primary data spacing resulting from the combination of data types.

4.1 Established guidelines in resource classification

The guidelines for resource classification have been developed regionally based on satisfaction of public faith in value reported. All have similar requirements based on measures of data quantity and relation.

Assurances are required when companies are looking for investors to develop a given deposit or to maintain cash flow to continue operations on a current deposit. Resource reporting is meant to communicate the degree of possible positive or negative future change in the grade and tonnage values defined by an estimate.

4.1.1 Motivation

Resource classification guidelines are developed to protect investors in publicly traded companies. Previous problems in reporting of resources, particularly the Bre-x fraud (Andrews, 2005), required legal rules for reporting of resource quantity, grade, and extent. Claims need to be backed-up with legal responsibility on the part of the company signatories. This is meant to prevent improper reporting and provide confidence to investors.

A resource is divided into various portions based on the amount of knowledge available about each portion. The guidelines are designed to impart the amount of confidence that the company has in the values being reported. This begins with feasibility studies and ends with extraction. In these terms it is easier for the investor to make decisions on the viability of resource.

4.1.2 Resource-reserve definition

As defined by NI43-101 a deposit is divided into five related, and overlapping, categories. These include three resource categories and two reserve categories (Canadian Securities Administrators, 2011).

Resources are divided into inferred, indicated and measured resources. Inferred resources are those that have enough information to imply a certain geological quality of resource. However, this quality is unable to be defined. When more information is gained through increased sampling this resource may move into the indicated category. This includes resources that could support a study to determine the economics of mining the resource. Finally, measured resources include those that are known to the extent that detailed economic studies and mine planning can be performed with confidence.

Reserves are divided into probable and proven categories. Probable reserves refer to the portion of indicated resources that are economic to extract. This may also include some portion of measured resources. Proven reserves represent the economic portion of measured resources. As reserves move from probable to proven there is an increase in the confidence and accuracy of the grade and tonnage values. There is a continuous relationship between indi-

cated and measured resources and probable and proven reserves. The flow between categories is seen in Figure 4.1

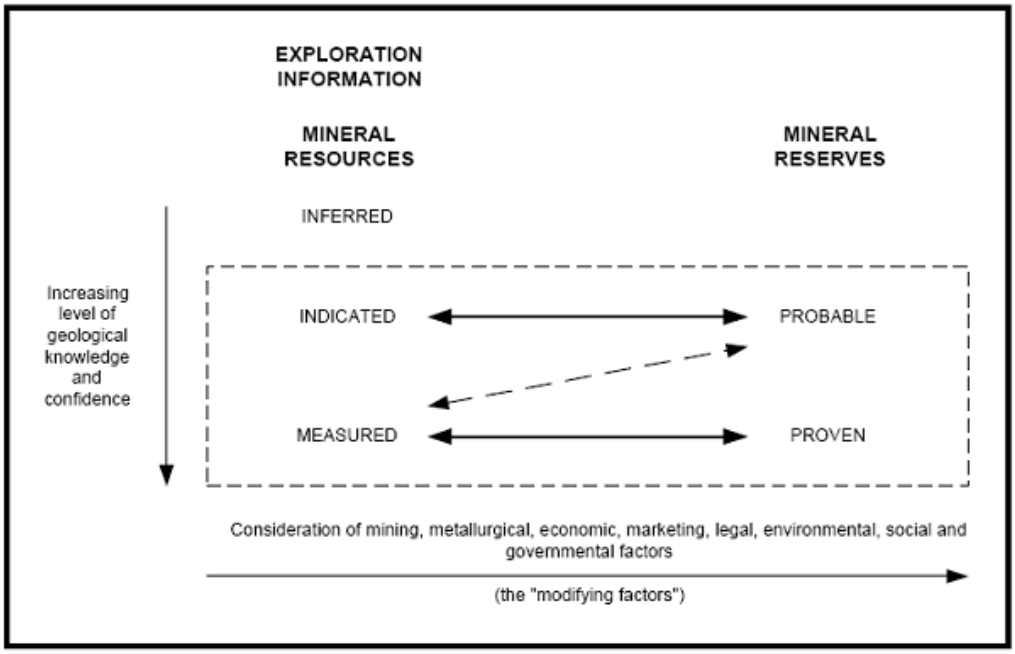

Figure 4.1: The flow of resource-reserve categories as knowledge of the deposit increases (Ramcharan, 2015).

The QP determines the volumes that fall into each category. The various resource reporting guidelines do not give strict rules around how the result is achieved. This depends on theoretically sound and repeatable methods for quantifying the uncertainty around resource values.

4.1.3 Regional guidelines

There are many codes that define best-practice for resource reporting. Each generally follows a similar format. Over-arcing statements are made concerning the knowledge of a deposit. No specific guide is given on how to establish the comparative level of knowledge. It is accepted that the uncertainty around the tonnage and grade decreases as a resource moves through the categories to proven reserve. All guidelines are based in local mining costs and so regional economics set the baseline for assessment of performance in terms of feasibility studies. It is left to the QP to decide how the guidelines are applied and what specific methods are employed to meet the guidelines. The QP must

also prove that the chosen methods provide acceptable outcomes under the regional guidelines.

NI43-101 is the guideline developed by the Canadian Securities Administrators. It applies to any company that is listed publicly within Canada. These guidelines apply whether the company operates inside or outside Canadian borders (Canadian Securities Administrators, 2011). The JORC code was established by the Joint Ore Reserves Committee(JORC) and applies to those companies listing themselves in Australia and many Asian countries. It has similar definitions as in the NI43-101 code (Joint Ore Reserves Committee, 2012). Industry Guide 7 is the guide defined by the Securities and Exchange Commission of the United States. This regulates any companies listing themselves within the United States. This guide follows the same resource-reserve definition as above (United States Securities, and Exchange Commission, 2015). Additionally, the SAMREC (South Africa) and the PERC reporting code(European Union and the United Kingdom) are widely used and follow a similar format (The Southern African Institute of Mining, and Metalurgy, and the Geological Society of South Africa, 2009), (Pan-European Reserves & Resources Reporting Committee, 2008). The only significant departure is the reporting system employed in Russia. Although there have been modifications in the last twenty years this reporting code is generally a hold-over from the Soviet era. It provides strict rules for calculating and reporting of resources and removes the judgement of the QP (Henley, 2004).

4.2 Current practice

In order to define the resource-reserve categories the QP must apply repeatable methods. These should allow the division of the categories based on some measure of knowledge of the deposit. There are several regularly applied options. These include distance to nearest sample, number of samples used for an estimate, estimation variance measures, among others. Any options could be considered when they are theoretically sound and repeatable.

An important aspect of all current practice is the development of thresh-

olds. Each of the following resource-reserve definition schemes relies on defined thresholds. This requires the QP to be familiar with the deposit and the relative requirements for quantification of that deposit type.

4.2.1 Sample distance

The 'nearest sample' definition relies on establishment of thresholds for the distance to the nearest sample used in an estimate of the resource. A sample must be within a certain distance in order for the volume to be considered within a certain category. As the distance to the nearest sample reaches smaller thresholds then the known information is greater. If a closer distance is reached the volume should move into a category that is recognized as better defined.

The 'number of samples' definition refers to the number of samples found in the search radius of a given estimate. Threshold values for number of samples need to be defined for each category. Classification is dependant on the search radius defined by the QP and need not be isotropic. It should be based on knowledge of the deposit and the underlying geological process.

Sample spacing is also used as a signifier of resource-reserve categories. This is prominent in very continuous deposits, such as coal beds. Distance between samples is used directly with some parameter adjustments based on known basin anisotropy (Bertoli et al., 2012).

4.2.2 Estimation variance

This classification scheme uses average estimation variance values to classify portions of a deposit. Appropriate threshold values are established through orientation of drill hole composites. When there is no grid of evenly spaced composites the thresholds can be established visually. However, it can be difficult to define clear cut-off values (Rossi and Deutsch, 2014).

This may be considered an extension of the number of samples definition with consideration of the covariance relationship. It allows the spatial relationship between points to be accounted for instead of the simple distance.

4.2.3 Further options

There are other options. All relate in some way to the quality and quantity of sampling across the domain.

Some practitioners use a multi-pass kriging classification scheme. In this method several iterations of a chosen kriging type are performed with varying minimum/ maximum number of samples per estimate and varying search radius. The variations in kriging plan create the thresholds that divide the classification categories. If the first pass has no minimum or maximum number of samples and a wide search radius then many data will be found and many blocks will be estimated. This is the base for classification. In the next pass the minimum samples can be increased, the maximum decreased, and the search radius decreased. Less blocks will be estimated during this pass. However, this also means that more information was available for estimation of these blocks. Blocks that are estimated in only the first pass are considered inferred. Blocks estimated in the second pass are indicated. This continues for a third pass with increased samples required and reduced search radius to establish measured resources.

Models of uncertainty can also be used for classification. In this method thresholds are defined by probability intervals. This is not suggested as a primary method as the parameters that go into developing the required simulation affect the outcome of the simulation and it is not simply data based (Rossi and Deutsch, 2014).

4.3 'Effective' data spacing

All methods for resource classification discussed above have a relation to data spacing. In some cases the relationship is direct, such as data spacing classification. In cases such as multi-pass kriging it is related in terms of the 'pass' in which a block is estimated. A smaller data spacing around a particular cell is more likely to produce an estimate in a more restricted pass. Data spacing has a great affect on classification.

'Effective' data spacing is the term applied to the modified data spacing

resulting from the combination of primary and secondary data types. Unless these data are so well correlated that they can be treated as the same data type then data spacing will not be a simple measured average distance. 'Effective' data spacing accounts for this through correlation and the effect of secondary data on reducing the estimation variance through cokriging.

4.3.1 Data spacing

Data spacing of primary data is often defined by the engineering and geological parameters of the project. It is often laid out in a configuration that provides a balance between ore body definition and exploration cost. Primary data gathering is often expensive in comparison to secondary methods. Due to cost primary data are often widely spaced.

Less expensive secondary sampling allows more samples to be collected at a similar or reduced cost as compared to the primary. However, the trade off is in the possibility of bias and error. The secondary data can show reduced correlation to the primary due to bias and error even when they represent the same variable.

If primary and secondary data types are used in estimation it does not make sense to simply combine them for a new average data spacing. The correlation between data types should be considered. If the data are not completely correlated then the secondary should not have the same contribution as the primary. This is best illustrated when considering the development of an LMC for the primary and secondary data. Assuming standardization, the correlation coefficient will be the sill of the cross variogram. Without standardization it can be calculated from Equation 2.9. If the correlation coefficient is 0.6 then the standardized cross variogram structure contributions will be limited to 0.6. The weights calculated and assigned to the secondary will then be restricted by the cross variogram. With the sill of the standardized primary direct variogram being 1.0 the primary weights will not be restricted.

In terms of resource classification the 'effective' data spacing lies between that of the primary data only and that of the primary plus secondary. The f factor below specifies modification of the data spacing; $f = 0$ gives primary

only and $f = 1$ gives primary plus secondary:

$$ds_{eff} = ds_{prim} + f \times (ds_{prim+sec} - ds_{prim}) \tag{4.1}$$

where ds_{eff} is the calculated 'effective' data spacing, ds_{prim} is the primary data spacing, $ds_{prim+sec}$ is the data spacing of combined primary and secondary data, and f is a modification factor related to the quality of the secondary data. Initially, f was suspected to be in line with correlation. However, the relation proves to be more complicated.

In the case of zero correlation an estimate is made with just primary data. In the opposite case of complete correlation the data spacing would be an average of the combined primary and secondary spacing.

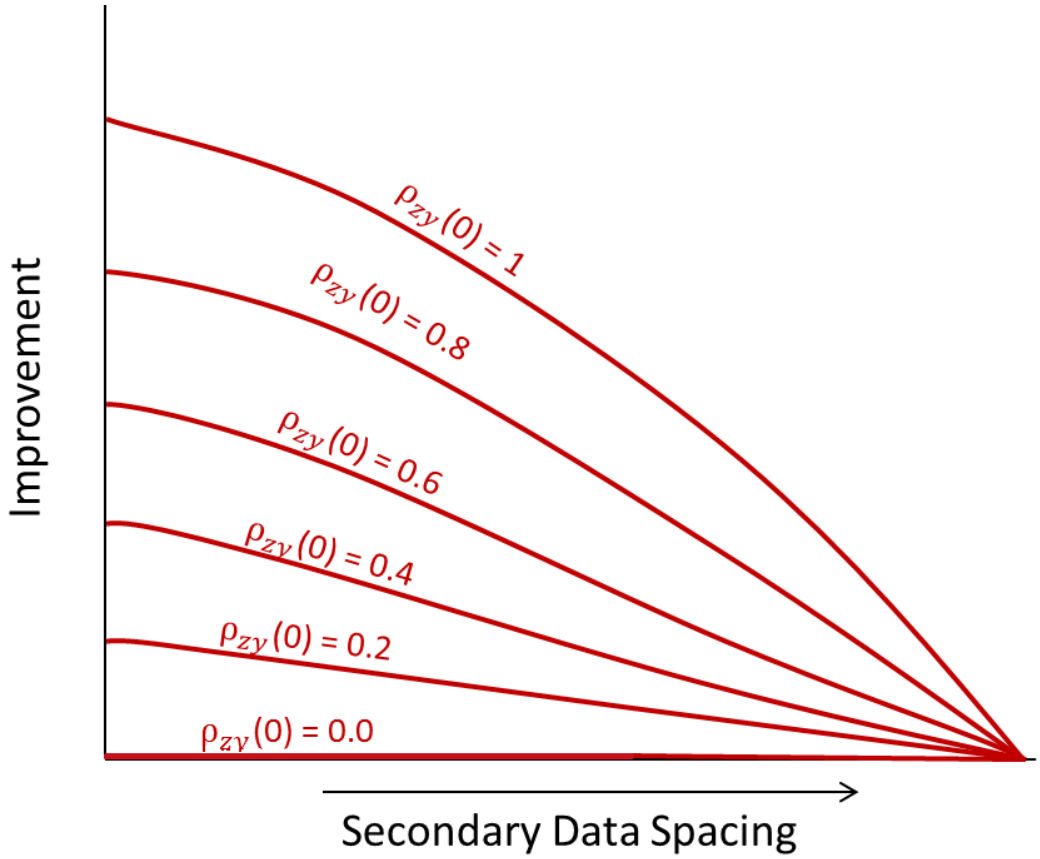

Figure 4.2: Schematic illustration of improvement in estimation due to secondary data of differing quality. Note that the plotted curves represent possible utility of secondary data at increasing increments of correlation.

Complexity exists in the primary and secondary data spacing relationship.

If there is abundant primary data then there is no help from the secondary. Additionally, there are considerations for the variogram between data types, and the scale of primary and secondary data. These issues must be considered in any practical study.

4.3.2 Correlation-data spacing relation

As data spacing decreases there will be a decrease in the average estimation variance. The best way to display change of variance is to consider the degree of improvement over the primary-only estimation variance. This is seen on the vertical axis of Figure 4.2. The scale is developed such that the cokriging average estimation variance with correlation coefficient of 1 is the maximum. The minimum is represented by the average kriging variance of the primary data only or the primary+secondary spacing with correlation of 0. Simplicity requires rescaling between 1 and 0. This variance relation allows the link between correlation and data spacing resulting from the combination of the primary and secondary grids.

The definition of 'effective' data spacing as defined by Equation 4.1 suggests a linear relationship with the correlation coefficient, but this proves not to be the case.

4.3.3 Establishing 'effective' data spacing

A numerical approach is considered to quantify the value of the secondary data in 'effective' data spacing. Scenarios of primary data spacing are coupled with many scenarios of secondary data with an LMC for estimation. This process is repeated for 11 equal increments of correlation (from 0.0 to 1.0). The resulting estimation variances are then compared for each data spacing combination.

The data is compared for each of 35 increments of average data spacing. The minimum to maximum average estimation variance is rescaled between zero and one for each plot. Figure 4.3 displays the plot for every fifth data spacing. It can be seen that the experimental points for each data spacing prove to be very similar. This relationship extends to the data spacings not shown. There is a minimal decrease in average estimation variance at lower

correlation but a great increase when moving above correlation of 0.5. This does not include the final data spacing (35) as it represents primary data spacing only and so no improvement in average estimation variance is seen.

The result is an understanding of the function for improvement in average estimation variance with the correlation coefficient. A fitted power law relation:

$$f* = 1 - (\rho_{k,k'}(0))^{2.8} \tag{4.2}$$

where $f*$ is a possible factor for scaling of the secondary data.

This produces an adjustment to data spacing that shows the previously assumed linear relation with correlation coefficient to be optimistic.

4.3.4 Application of 'effective' data spacing

Applying the 'effective' data spacing is done over an entire domain. However, there are concerns for unequal sampling.

Figure 4.4 displays the case of secondary data overlapping with primary data in one area of the domain. In this case only the area covered by primary and secondary data would be considered for data spacing recalculation.

Success in applying these methods is dependent on data configuration. If the secondary data points are very close to the primary points (Figure 4.5a) then there is little to gain due to redundancy because there is minimal increase in information content provided by the secondary. The data types can be highly correlated but will achieve minimal spacing reduction. The best results are achieved in a minimum combined spacing configuration (Figure 4.5b) as the secondary information content is maximized.

Note that areas with no primary data with the search range will require separate assessment by the QP.

When considering the methods of resource classification from Section 4.2 the 'effective' data spacing can be applied to methods where data spacing directly defines classes. However, as most methods are, in some way, connected to data spacing there may be opportunities to modify the methods to allow 'effective' data spacing application.

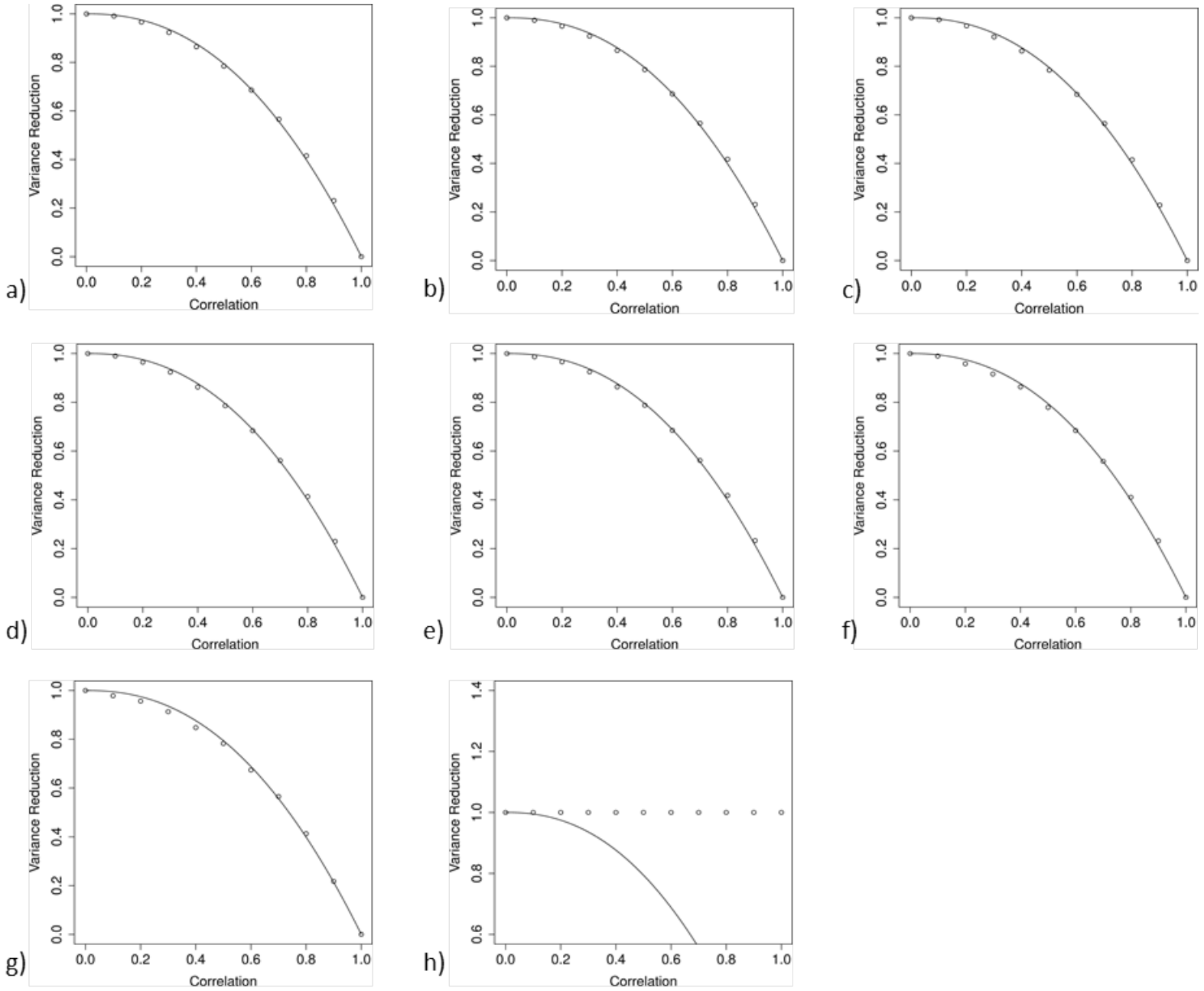

Figure 4.3: A display of the modelled relationship between correlation coefficient and average estimation variance improvement. Each plot represents one average data spacing of combined primary and secondary data. The average data spacings shown are a)58.15 ft, b)60.45 ft, c)65.17 ft, d)72.01 ft, e)85.40 ft, f)100.39 ft, g)117.66 ft, and h)137.16 ft. Note that these spacings are a representative sample of the 35 increments of spacing analysed.

4.4 Recommended practice

There are practical issues in tackling 'effective' data spacing that must be addressed for proper implementation.

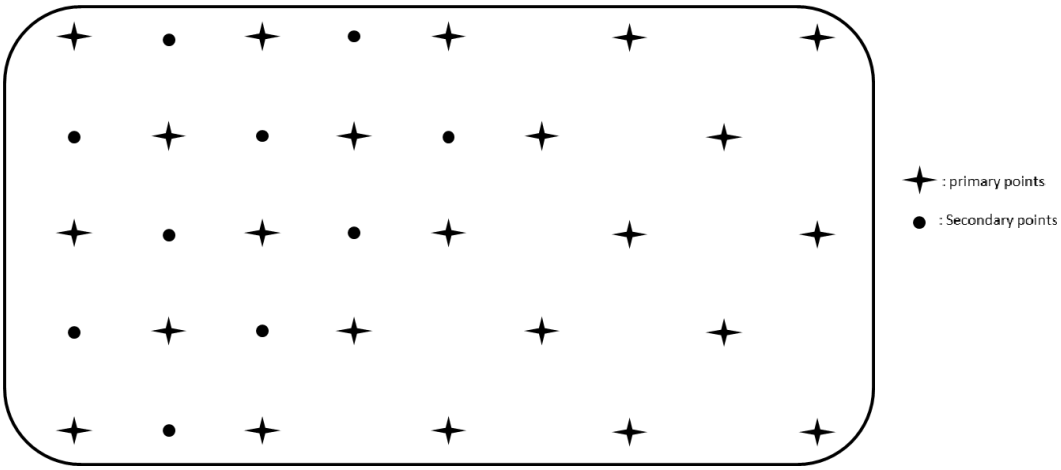

Figure 4.4: Primary and secondary data points overlapping in a 2-dimensional domain. The secondary points do not cover the entire domain. This figure is an example and not to scale.

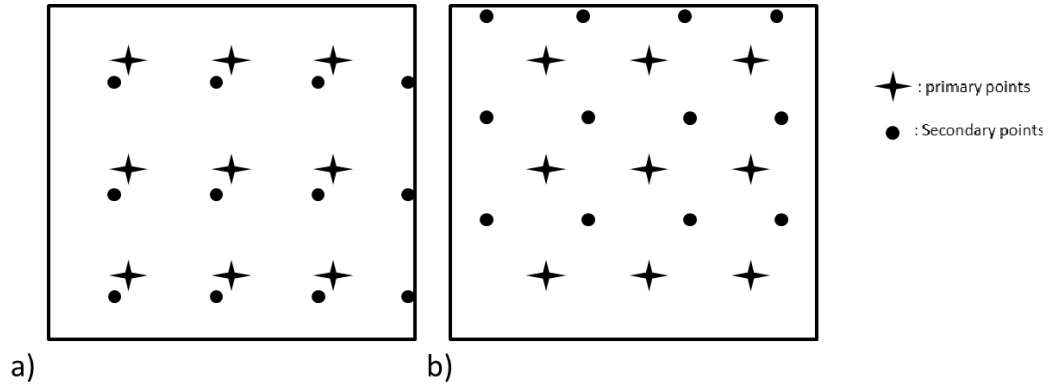

Figure 4.5: Primary and secondary data configurations for application of 'effective' data spacing. a) represents a poor configuration and b) is an optimal configuration. These figures are examples and not to scale.

4.4.1 Study of applicability

The applicability of 'effective' data spacing is based on a decrease in kriging variance as we move from OK kriging with only primary data to cokriging.

This requires the initial modelling of variograms for the primary data and subsequent OK kriged estimate. The kriging variance associated with this estimate should be the greatest. Additionally, the two data types can then be combined directly as if they were the same data or displayed very high correlation. This provides an estimate for comparison where all data are used. Finally, an LMC is developed and a cokriging estimate is performed. It should

be noted that OK with combined primary and secondary data is likely to produce an average estimation variance less than SCK. However, this result is wrong as it does not address error in secondary data.

In general:

$$\bar{\sigma}^2_{OK_P} \geq \bar{\sigma}^2_{SCK} \tag{4.3}$$

then the f factor as defined in Equation 4.3 could be considered for the 'effective' data spacing calculation.

The extent of the difference between average estimation variance values may have an effect on the applicability of the 'effective' data spacing. Even with the appropriate adherence to Equation 4.1 minimal difference between average estimation variance values could suggest that the primary data spacing be considered.

4.4.2 Applying updated data spacing

The updated spacing is helpful when applying the methods of classification based in sample distance. This is particularly helpful in situations where data spacing is the classifier.

Average spacing depends on proper domaining with reference to application of the 'effective' data spacing equation. When sampling is not equally gridded in the domain, as in Figure 4.4, averaging can be problematic. Certain areas of the domain are represented closely by the 'effective' spacing and other will be off. This is a critical point to retain when domaining.

When applying the 'effective' data spacing it must be considered a function specifically for classification. There is obviously no change in the data positioning. The combined primary and secondary data are likely to have a much smaller spacing. This simply reflects the value of the secondary information in prediction of the primary variable.

4.4.3 Decisions of the Qualified Person

The identification of 'effective' data spacing includes parameters defined by the practitioner. These parameters include domaining of the the deposit and

the kriging parameters including search and variograms. These parameters are thoroughly checked by practitioners in a well developed geostatistical study. However, a spacing study specific to the region is suggested.

The QP is relied on to consider the consequences of applying 'effective' data spacing. In the base case of equally gridded primary and secondary data the job is easier. When averaging spacing in domains where sampling is not as homogeneous there will be smoothing of 'effective' spacing. The QP may be required to prove the suitability of 'effective' data spacing. Performed correctly, the spacing study described above will provide the support information

The QP retains the ability to modify the effective data spacing. This can be done to create a more conservative adjustment. When establishing the the modification factor the equation may be adjusted to reflect a greater curvature associated with reduced impact of the secondary. There may be cause for a more aggressive approach by moving the model toward linearity but this would need thorough justification.

The decision to employ 'effective' data spacing rests with the QP. This is a method developed purely for classification. As with all new methods, clear documentation is required. There is no stated guideline for treatment of multiple data type combination in NI43-101, the JORC code, Industry Guide 7, the SAMREC code, or the PERC code. There are references to 'modifying factors' in estimates which may include the methods here. It is advised to document completely when applying this method.

Chapter 5

Case study in reclassification

This case study is developed using real data from the Red Dog Mine in Alaska (the data was provided by Teck for research purposes). The case of a primary data type with a single secondary data type is considered. Primary data is represented by diamond drill core assay values. Secondary data is blast hole bulk assay values. Blast holes (BH) samples are known to have greater bias and error based on over-drill and sample practice. Over-drill refers to the practice of drilling blast holes beyond the bench height depth to facilitate breakage of the bench to the appropriate height. Over-drill can add bias and error to blast hole assays in two ways. First, the sample taken from a blast hole may include material from the portion of the bench below. Second, material from the previous bench may be included in a blast hole sample due to the top intermingling with the bottom of the over-drilled bench from the last blast.

5.1 Data analysis and statistics

The data is assessed for division into stationary domains and each domain is studied separately. The minimum requirement for this study is two overlapping unequally sampled data sets with a significant number of secondary data. This requirement guides the data analysis. Blast hole data is available in areas already under production. This limits consideration to a segment of the deposit.

5.1.1 Geological context

The geological context for this deposit is defined by the Teck exploration geology team and can be viewed in detail in the Red Dog NI43-101 report from 2004 (Knapp, 2004). The deposit is considered to be of sedimentary exhalative origin with replacement mineralization. The Red Dog region consists of four deposits differentiated by syn-depositional thrust faulting. This faulting causes overlapping of mineralized and non-mineralized zones.

Of the four sub-deposits only one has available production data suitable for use in this study. This is the Main deposit. It has been developed as a production pit and has available blast hole assay data along with exploration diamond drilling. The majority of the Main deposit mineralization occurs in the mid thrust plate. It contains three distinct ore-types based on major constituent minerals and deposition type. These are siliceous, veined, and baritic in order of decreasing abundance with the vast majority being siliceous (70%).

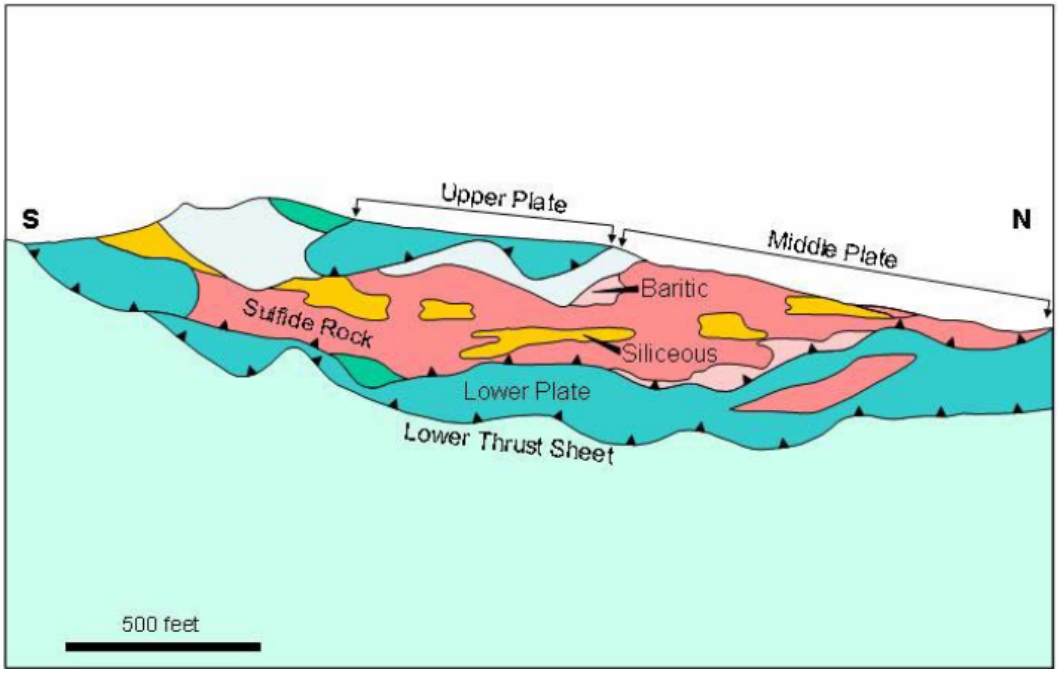

Figure 5.1: A typical west-facing section through the geology of the Main deposit within the Red Dog mine area (Knapp, 2004).

The overall deposit is said to show a north-south trend of increased conti-

nuity with maximum continuity being approximately 15 degrees from north. This is tested in the following study to ensure that this continuity applies to the chosen domain.

While the deposit is polymetalic only zinc assay values are considered for the purposes of this study. Zinc assays are in percentage of total rock.

5.1.2 Stationarity and domaining

Due to the overlapping data requirement of the study the zone chosen was the Middle Plate zone contained within the Main deposit. Limiting to the middle plate helps to ensure geological consistency while the Main zone provides production information in the form of blast holes. Collar locations for each of these data sets can be observed in plan view in Figure 5.2 *a* and *b*. Note that blast hole data is overlapping in plan view as each blast hole is drilled for a single mining bench. In this case the benches are stacked in plan.

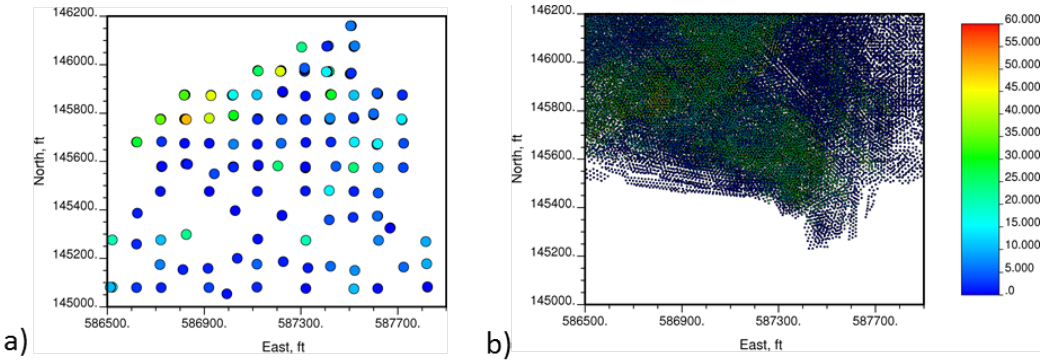

Figure 5.2: Plan view sections displaying the collar locations of a) diamond drilling data and b) blast hole data with values represented in percent zinc over the study domain measured in feet.

5.1.3 Compositing

The data sets are composited to a length of 25 feet. This represents the chosen SMU of the deposit. Only diamond drilling values required compositing as the blast holes assays are based on the bulk sample over the 25 foot bench height for which the blast hole is drilled.

The compositing is performed using the GSLIB software *Compositing*. This program computes composite values by arithmetic average within each

composite by length weighting (Silva and Boisvert, 2014).

5.1.4 Declustering and despiking

Declustering and despiking were performed on both diamond drilling and blast hole data sets.

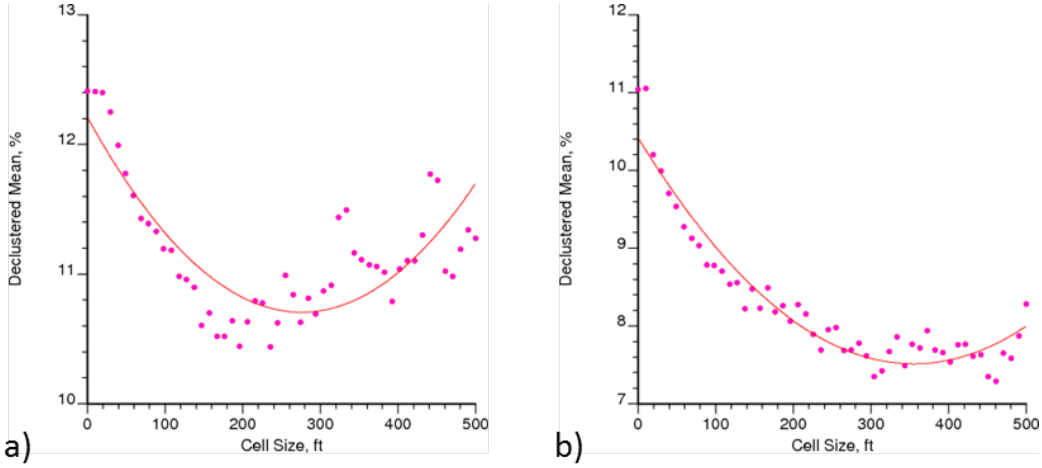

Figure 5.3: Plots of cell size versus declustered mean for a) diamond drilling samples and b) blast hole samples.

The minimum delcustered mean value for primary data corresponds to the cell size of 196 cubic feet while secondary cell size is 460 cubic feet. This can be observed in the in Figure 5.3 *a* and *b* respectively.

The declustered histograms showed that values were being slightly over-reported for both diamond drilling and blast holes due to the relative locations of higher grade samples. This is often the case due to the preferential sampling of areas known to contain higher grades. Diamond drilling data with declustering weights applied have a mean value of 10.44 percent while the clustered data mean is 12.41 percent. The blast holes show a similar response to declustering with a declustered mean value of 7.29 percent as compared to a clustered mean of 11.04 percent. However, global kriged estimates of each data set shows that the cell sizes corresponding to minimum mean value are over correcting both distributions. The global kriged mean value for diamond drilling is 10.76 and corresponds to a cell size of 157 cubic feet. The global kriged mean value for blast hole data is 7.63 and this corresponds to a cell size

of 235 cubic feet. These corrected cell sizes are used in recalculation of the declustering weights.

Prior to despiking each data set was reviewed for quantity of data with the same value. This is most often an abundance of zero values but may also be high values below a top-cut. In the case of both data sets there were many zero values. However, the number of zeros in both cases did not exceed eight percent of the total number of data. Deutsch(2014) suggests that despiking is of consequence when this total is greater than ten percent of total data. Both data sets showed little response to despiking.

5.1.5 Histograms and probability distributions

Both diamond drilling and blast hole data sets were explored visually using histograms to characterize the spread of the data. Each showed the lognormal shape characteristic of many metal deposits with the blast hole data being slightly more erratic in the high tail. Application of declustering weights produced less erratic distributions while despiking had no effect. This can be seen in Figure 5.4.

Cumulative probability plots were created for each data set using declustering weights. This can be seen in Figure 5.5. Note that both diamond drilling data and blast hole suggests that higher values may represent a different population. In both cases these high tails contain a very few values.

5.2 Kriging for comparison

The standardized diamond drilling composites are kriged using OK to provide a baseline for estimate and kriging variance from which to measure the quality of the cokriged estimate. OK is chosen as it is the industry standard in direct estimation. This is evidenced by recent NI43-101 reports in which most estimates reviewed were completed using the OK method.

Additionally, an OK estimate is made using both the primary and secondary data directly combined as a single data set. This is also compared with the subsequent cokriged estimate.

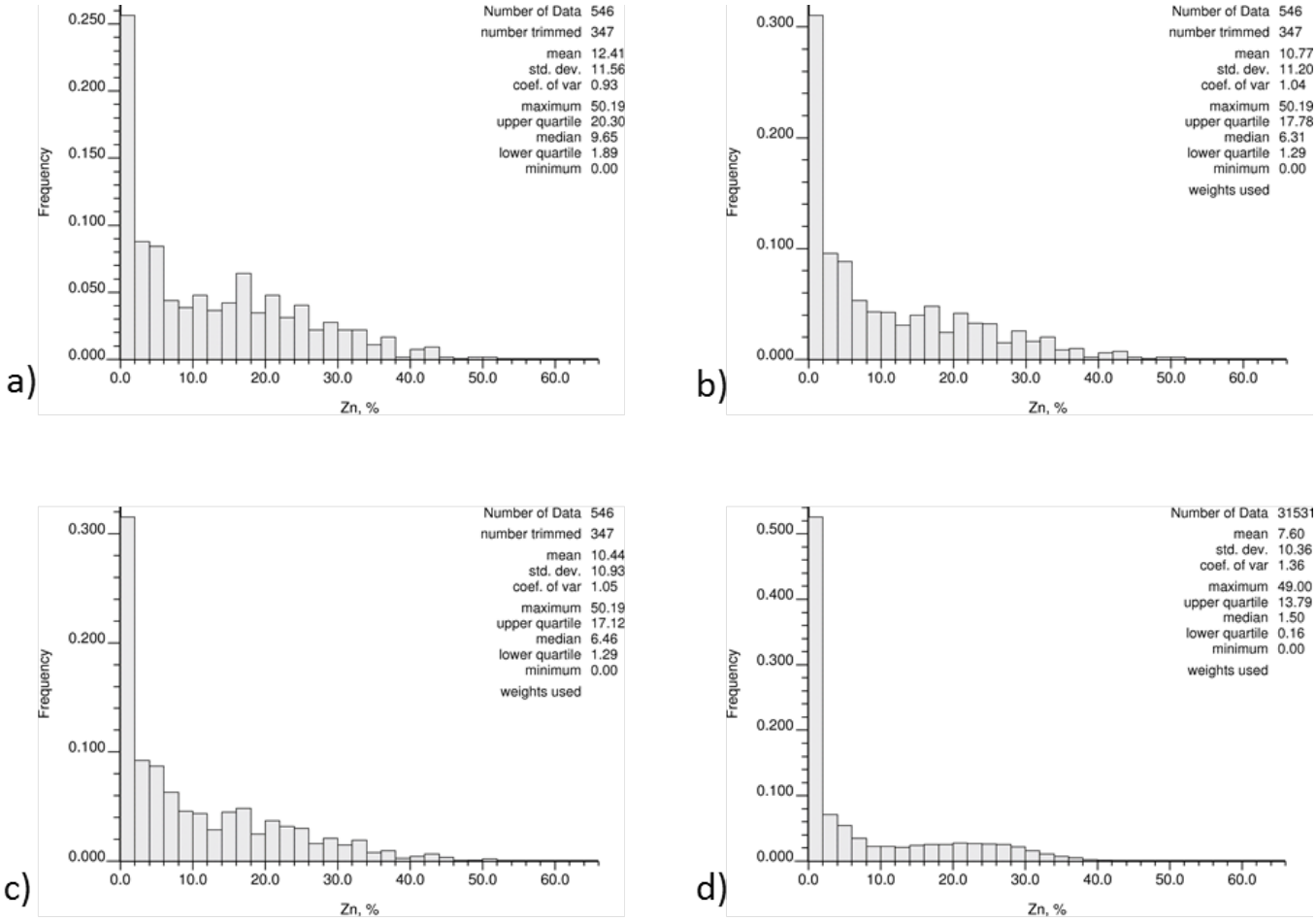

Figure 5.4: Histograms displaying the distributions of a) clustered diamond drilling zinc assays, b) declustered diamond drilling zinc assays, c) clustered blast hole zinc assays, and d) declustered blast hole zinc assays.

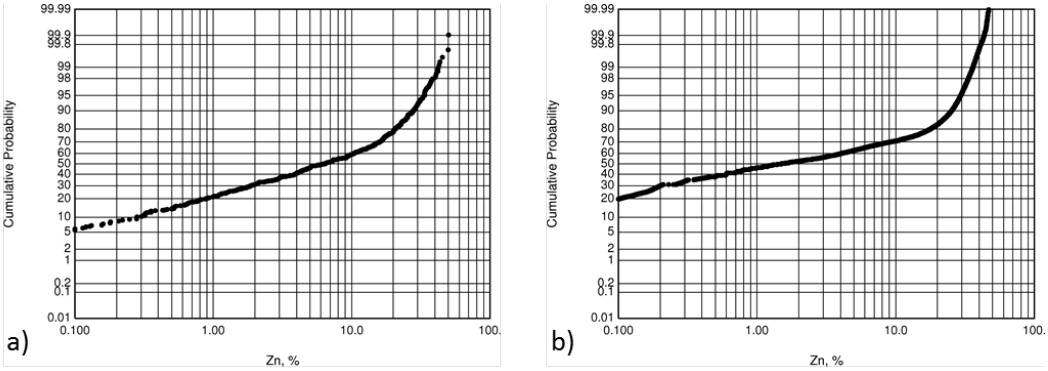

Figure 5.5: Cumulative probability plots of a)diamond drilling zinc assays and b) blast hole zinc assays. Note that curvature in each plot at values above 20 percent suggesting separate populations above this value.

Maximum/minimum continuity was checked using a global kriging approach. The results mimicked closely the directions indicated in the Red Dog technical report with maximum continuity at 15 degrees. These directions of continuity were similar for both datasets.

Prior to modelling of variograms data was standardized for ease of use and comparison. This is continued for further estimation in the case study.

5.2.1 Variogram modelling

Experimental variograms were modelled for the horizontal directions defined above and the vertical.

For OK with primary data in the horizontal directions a lag distance of 100 feet was chosen as that is the approximate spacing of of the diamond drilling grid. Due to the comparatively small number of diamond drilling values a lag tolerance of 70 feet was applied with 0 dip and a dip tolerance of 45 degrees. The bandwidth was set to 125 feet to limit to approximately 5 sample lengths above and below. This produced the the variograms seen in Figure 5.5*a* with overlaid model. The models contained no nugget effect structure and two spherical structures. Each structure contributes 0.5 to to unit variance with ranges of 120 and 350 feet.

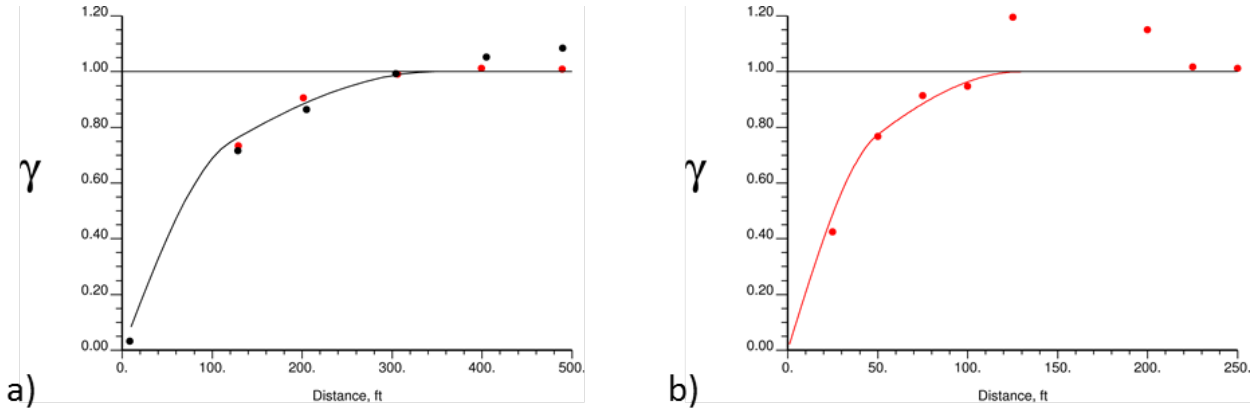

Figure 5.6: The a) horizontal and b) vertical variograms and models developed for OK estimation. Black indicates the direction of greatest continuity while red represents the least continuity.

In the vertical direction there is closer spaced information. A lag distance

of 25 feet was selected to reflect the composite length with a lag tolerance of 15 feet. Bandwidth was 300 to account for three sample distances in the horizontal. Figure 5.6*b* displays the experimental points with overlaid model. The model required no nugget effect and two spherical structures at ranges 120 and 350 feet respectively as with the horizontal models. While a trend can be noted here it is standard practice to model to the sill and ignore the trend in modelling.

The variograms for OK with combined data showed extreme trends in all directions. Despite each data type being standardized prior to combination the experimental variogram showed large interpreted sill values. To assist in estimation it was decided to apply the variogram model used for diamond drilling data. This represents the data that is assumed to be closest to the real underlying distribution of zinc values.

The variograms were built using a unix script constructed from GSLIB programs $Gamv2004$, $Vmodel$, and $Vargplt2004$.

5.2.2 Estimation

The estimation of the diamond drilling data was completed using the GSLIB program $Kt3d$ with the option of OK selected. The kriging parameters are shown in Appendix 2; Figure A.1 for OK with primary data only and Figure A.2 for OK with combined data. There was difficulty in estimating all blocks for OK with primary data only. The original search radius of 5 times the data spacing was increased to 300 feet in the vertical direction before all blocks were estimated. Estimates were made for 37,632 blocks of 25 x 25 x 25 feet. Each block was discritized by 3 in X and Y directions and by 1 in the Z direction as data was composited in that direction.

The plotted estimates and estimation variance using OK can be seen in section in Figure 5.7 *a* and *b* respectively. The same plots for the OK estimate with combined primary and secondary data can be seen in Figure 5.8 *a* and *b* respectively. The 'spotted dog effect' is clear in the plan and east-facing sections for both estimates where variance is minimal around drill hole locations.

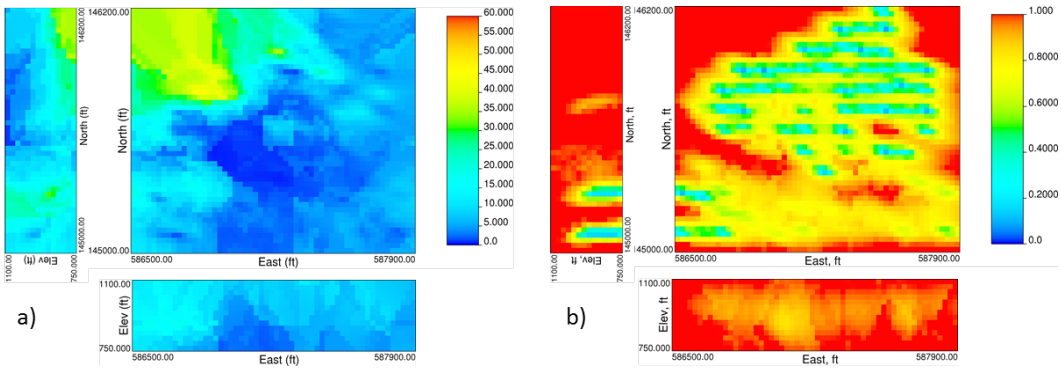

Figure 5.7: Plan, east-facing, and north-facing sections displaying the a) estimated values and b) estimation variance for the OK estimate with primary data only.

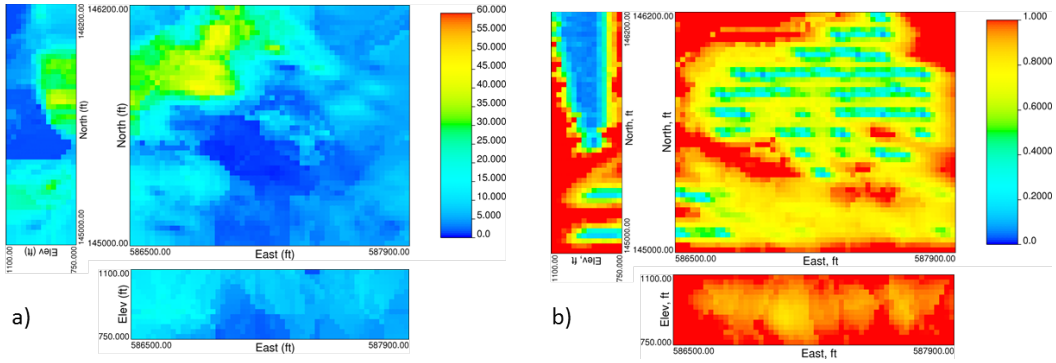

Figure 5.8: Plan, east-facing, and north-facing sections displaying the a) estimated values and b) estimation variance for the OK estimate using combined primary and secondary data.

5.2.3 Checking the kriged model

The model created above was checked against the known diamond drilling assay values by running *Kt3d* in cross-validation. The scatterplot of Figure 5.9 shows how well the estimate reproduces the known data. There is a general linear trend while the correlation of 0.665 reflects the data spacing and variogram.

The OK estimate of combined data is cross validated using the diamond drilling assay values and is displayed in Figure 5.10. There is good reproduction of true values and may be interpreted incorrectly as a good quality estimate.

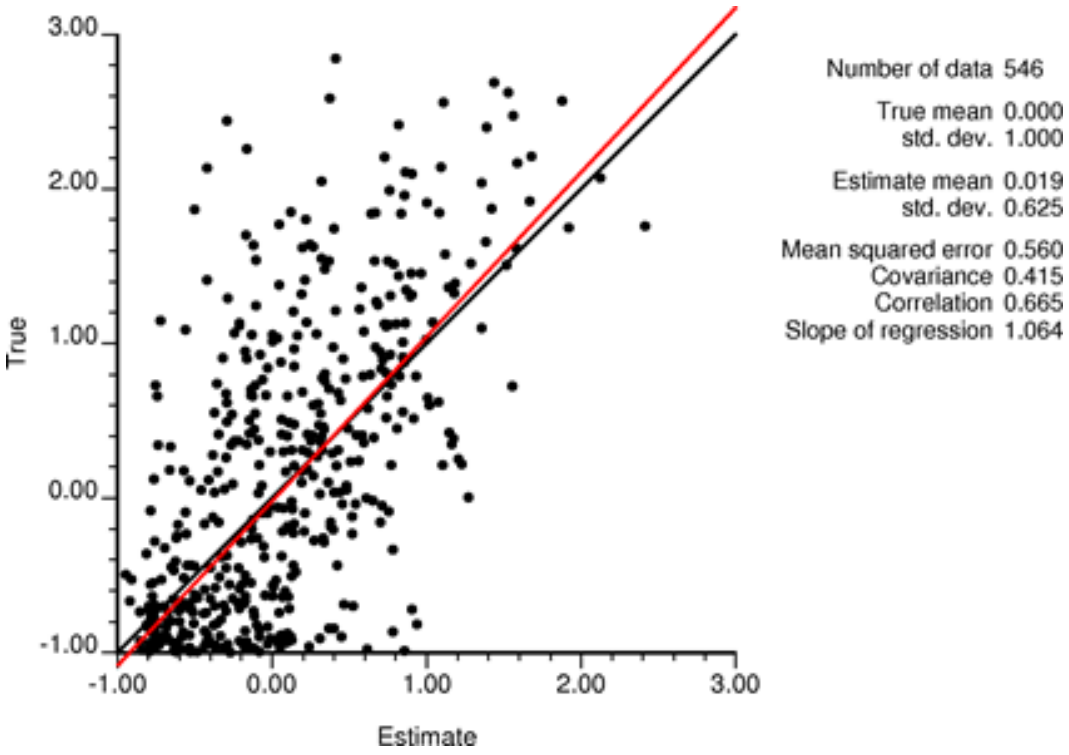

Figure 5.9: Cross validation of the OK estimate using original diamond drilling assays.

5.3 Linear model of coregionalization

The LMC was developed in variogram form as this is the expected input for the GSLIB programs that are used for this analysis; there is no fundamental difference to using covariances.

Standardization was completed and checked for both data sets prior to building the LMC. This ensures unit variance for both primary and secondary and therefore allows direct comparison and easily understandable cross-covariance values.

5.3.1 Establishing correlation

Correlation between these variables cannot be established directly due to complete unequal sampling. This requires application of the methods discussed in Section 3.1.1 and establishment of the cross-covariance experimental points. For these variables the $C_{k,k'}(0)$ value was extrapolated back to 0.85 where k is the primary and k' is the secondary. This extrapolation is seen in Figure 5.11.

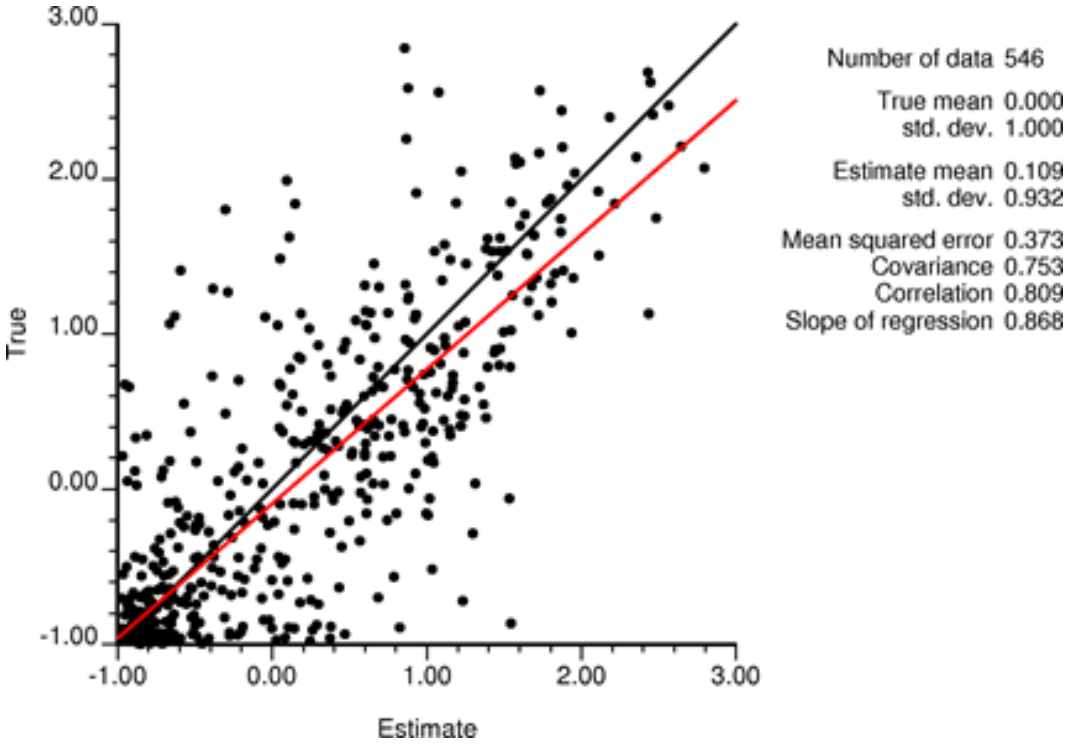

Figure 5.10: Cross validation of OK with combined data estimate using diamond drilling assays.

Referencing Equation 2.8 in Section 2.2.3 and due to data standardization $\rho_{k,k'}$ is equal to $C_{k,k'}(0)$.

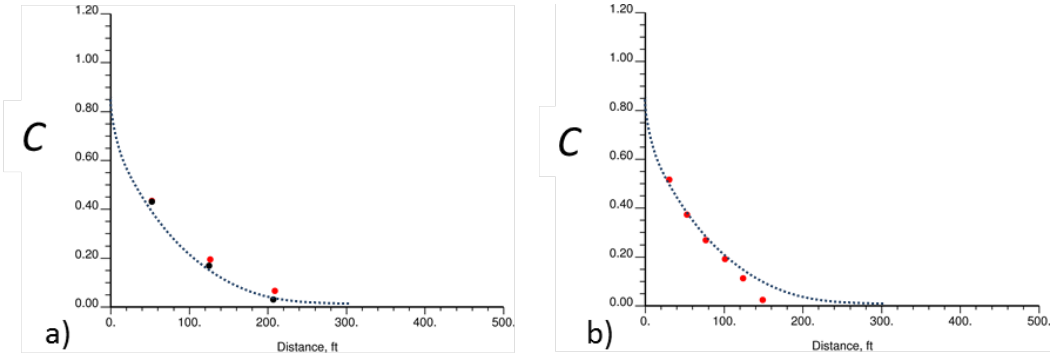

Figure 5.11: The projection of the cross covariance curve to C(0) in a) the horizontal directions and b) the vertical direction.

5.3.2 Modelling cross-variograms

Experimental points for the cross relationship are produced in the form of experimental cross-covariance. However, with $C_{k,k'}(0)$ established the maximum

variance of the cross-variogram is known.

Creating the cross-variogram involves the 'flipping' of the cross-covariance experimental points based on the established maximum variance value of 0.85. As the cross-variogram has no observed nugget effect the origin should be at 0.0 and so 0.85 is subtracted from each experimental point value. The conversion can be seen in Figure 5.12. It should be noted that some experimental points for the cross variograms are above the defined still. This corresponds to negative experimental cross covariance points. An argument could be made for a higher sill value. However, it is difficult to be sure how to treat the negative cross covariance values. A conservative approach is considered prudent.

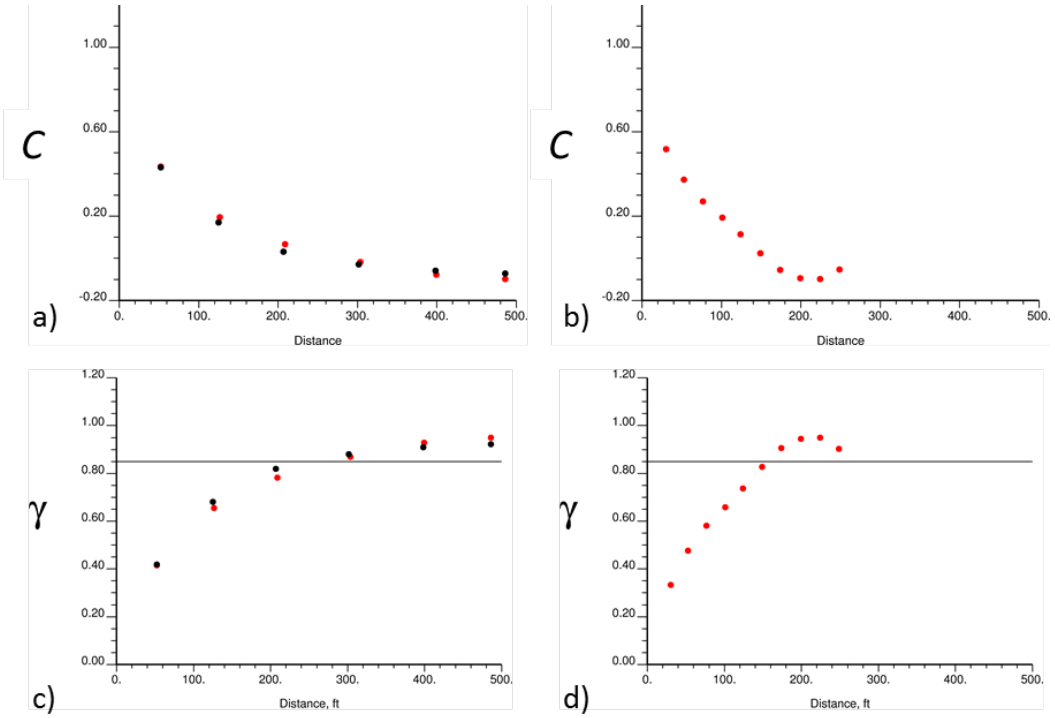

Figure 5.12: The 'flipping' of cross covariances to cross variograms for use in estimation. The a) horizontal cross covariances is converted to c) the horizontal cross variograms while the b) vertical cross covariance is converted to d) the vertical cross variogram. In figures a) and c) red points reflect direction of highest continuity while black represents the lowest

The modelling of the cross variogram is completed using two spherical structures. The parameters of the variograms are shown in Table 5.1. The modelled cross variograms are displayed in Figure 5.13 *b* and *e*.

Estimation Type	Nugget Effect	Number of Structures	Structure contribution	Anisotropy angles			Ranges		
				Max	Min	Vert	Max	Min	Vert
		1	0.3	15	105	0	100	100	100
OK w/ primary only	0	2	0.7	15	105	0	270	330	150
		1	0.5	15	105	0	120	120	50
OK w/ primary and secondary	0	2	0.5	15	105	0	350	350	130
SCK (LMC)		1	0.461	15	105	0	30	50	60
direct -primary	0.05	2	0.489	15	105	0	300	300	175
		1	0.149	15	105	0	30	50	60
direct-secondary	0.05	2	0.807	15	105	0	300	300	175
		1	0.255	15	105	0	30	50	60
cross	0	2	0.595	15	105	0	300	300	175

Table 5.1: Variogram parameters defined for each estimate.

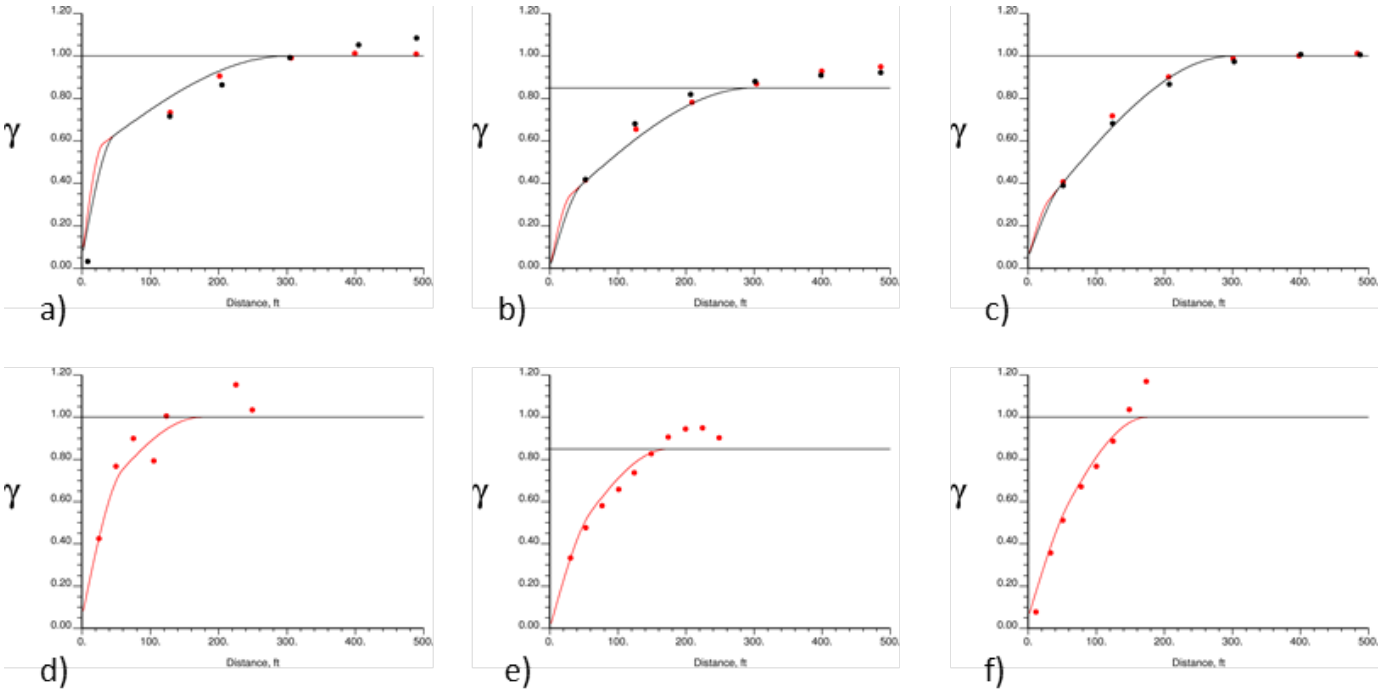

Figure 5.13: Experimental variograms and models for the a) horizontal primary, b)horizontal cross, and c) horizontal secondary as well as the d) vertical primary, e) vertical cross, and f) vertical secondary.

5.3.3 Modelling direct variogram

Experimental variograms for the primary variable, diamond drilling zinc assays, were developed for the original OK analysis. These points were used for LMC development. For the secondary variable, blast hole zinc assays, the same analysis is performed. Lag distance was chosen in each direction based on sample spacing. This produced well-behaved secondary experimental points.

Although the primary variograms have already been modelled for the OK estimate they must be remodelled to satisfy the LMC. In this case both primary and secondary direct variograms must include the same two spherical

structures used for the cross-variogram. All parameters can be seen in Table 5.1 and each direct variogram can be seen in Figure 5.13 a, c, d, and, f.

5.3.4 Checking the LMC

There are several requirements that must be met before the LMC is acceptable for use. The positive definiteness of each structure contribution matrix must be ensured by checking the determinant of each matrix. This is satisfied for the LMC.

Ranges for corresponding structures must also be the same for each direction in order for the relationship to be meaningful. This requirement is met and observed in Table 5.1

A simple check for positive definiteness is performed by running the GSLIB program $Varfit-LMC$ (Neufeld, 2004). This program automatically creates an LMC using experimental points and several user input parameters. While the outcome is not exactly what was produced manually the results were very close. This provides some measure of confidence in the variogram fit.

5.4 Cokriging

The cokriging estimate is completed using the updated $Cokb3d$ cokriging program. Additional functionality includes modified variogram input, multiple data file input, data standardization option, primary variable selection, and jacknife cross-validation. This update is documented in Appendix 1.

5.4.1 Cokriging with the LMC

In order to check the effect of secondary data on the estimate and estimation variance the parameters of the estimate are kept the same as the OK estimate. This includes blocks size, discritization, search radii, and search ellipse angles.

The estimate was performed using standardized primary and secondary data sets as these were used in LMC development.

Sections displaying the resulting estimates are seen in Figure 5.14a. Figure 5.14b displays cokriging estimation variance sections. Note that the locations of sections a and b in Figures 5.7, 5.8, and 5.14 are the same and so $5.7a$,

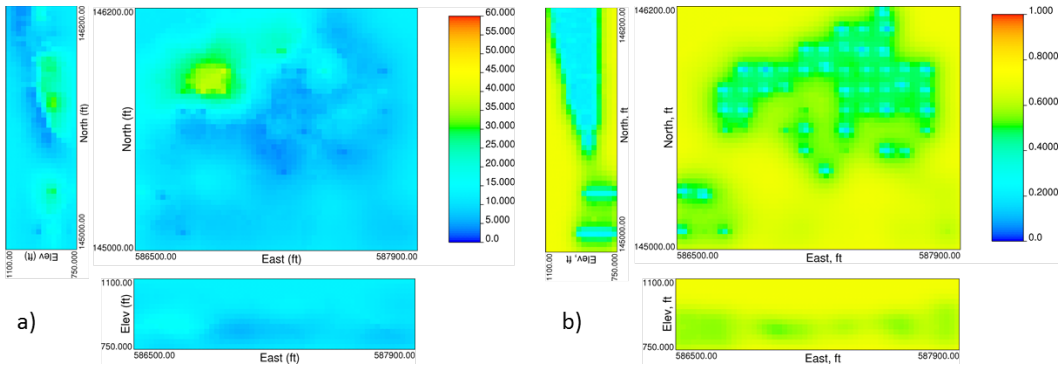

Figure 5.14: Plan, east-facing, and north-facing sections displaying the a) estimated values and b) estimation variance for the SCK estimate.

5.8a, and 5.14a estimates and 5.7b, 5.8b and 5.14b estimation variances can be compared directly.

5.4.2 Checking the cokriging model

The cokriged estimate can be checked in the same way as the OK estimate using direct cross-validation with the diamond drilling assay values. The plot seen in Figure 5.15 shows the good reproduction by the cokriged estimate as compared to the primary data with high correlation and relatively low mean squared error.

5.5 Reclassification

The reclassification of the resources is considered here in terms of data spacing. Tonnage is not reclassified, however, the extension to reclassification is made through the difference in primary data spacing and 'effective' data spacing.

5.5.1 'Effective' data spacing

The 'effective' data spacing can be calculated by applying f as calculated from Equation 4.1 with $\rho_{k,k'} = 0.85$. The 'effective' data spacing can then be used in calculation of resource-reserve categories for the domain considered. This is a reflection of blast hole information usage and not on increased data available for study of the domain.

It can be seen in Table 5.2 that the minimum data spacing is achieved using

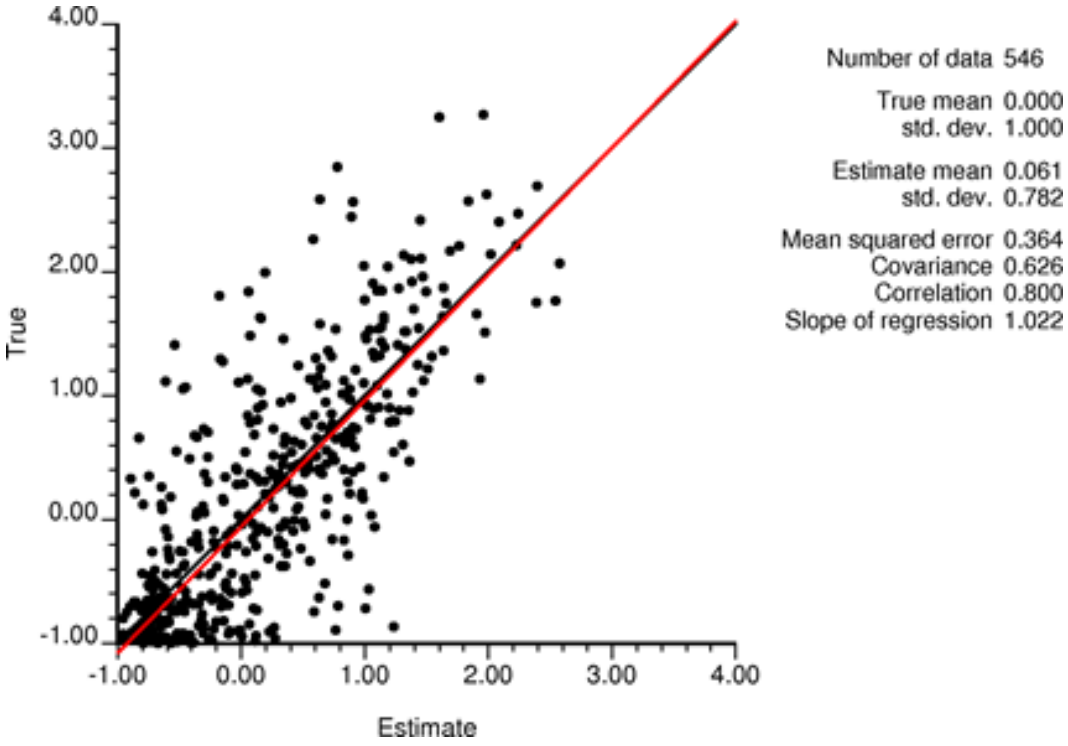

Figure 5.15: Cross validation of the SCK estimate using diamond drilling assays.

the direct combination of data. The 'effective' data spacing calculated from the average SCK results is much higher than the minimum. When compared with both the maximum spacing, as represented with diamond drilling composites only, and minimum spacing, as represented with the combination of blast holes and diamond drill composites, the 'effective' data spacing is 37 percent less than maximum. This reflects the conservative relationship between average estimation variance and correlation discussed in Section 4.3.4.

Data	Average Data Spacing
Primary	1070 ft
Secondary	663 ft
Combined	215 ft
Effective'	758 ft

Table 5.2: A comparison of the average data spacing values of both OK estimates with the 'effective' data spacing calculated for the SCK estimate.

5.6 Outcomes and discussion

The first goal of this study is to display the abilities of cokriging with unequally sampled real data. The second goal is to establish an acceptable model for calculation of 'effective' data spacing and apply this model to calculate the 'effective' spacing of the data used in cokriging.

SCK with unequally sample data proves to be a good quality application. There is a marked decrease in average estimation variance when compared with OK (primary only) as would be expected with abundant additional information but also a more accurate reproduction of known data values. OK with combined data provides a reduced average estimation variance and shows a good quality of true value reproduction but a slightly higher mean squared error. However, this result is wrong as it does not provide proper consideration of the secondary data.

The 'effective' data spacing calculation is shown to be a viable method for applying the reduced estimation variance from cokriging to support more complete reporting. Theoretical study shows a repeatable factor, as defined in Equation 4.4, allowing the scaling of combined data spacing to reflect the information content provided by the secondary data.

5.6.1 Model comparison

In model comparison there are marked differences in estimation variance between the three estimates completed. Table 5.3 shows the decrease in average estimation variance from OK with primary data and SCK. While the overall difference can be observed in the sections of estimation variance in Figures 5.7*b* and 5.14*b* the visual difference is not so easily defined for sections in Figures 5.7*a* and 5.14*a*. The difference in average estimation variance is much less for the OK (primary + secondary) and the SCK estimate than between the OK (primary only) and SCK estimates. While this could suggest motivation for the direct combination of data types it does not take into consideration the error that may be passed on to the estimated values. An OK (primary + secondary) estimate that produces an average estimation variance less than

that of SCK would still be wrong.

The models are also compared in cross validation. It can be seen in comparison of Figures 5.9, 5.10, and 5.15 that OK with primary data has the least quality reproduction of the original diamond drilling assay values. With the addition of the blast hole data both OK with combined data and SCK are better informed in cross validation. While both show fairly good reproduction the SCK estimate shows a reduced mean squared error. In addition, the true mean of the SCK estimate is closer to the true value, 0.0, than that of combined data OK.

In comparison with approximate data values observed from the location maps of Figure 5.4 show that the SCK produces estimates that are closer to honouring the original data values from diamond drilling assay composites. This is particularly obvious in comparison of mean squared error values. The OK estimate returns a value of 0.56 while SCK shows reduction to 0.36.

Average Estimation Variance	
OK w/ primary	SCK
0.77542	0.44803

Table 5.3: A comparison of average estimation variance resulting from OK with primary data and SCK.

Note that bias was not discussed here as both data types were standardized prior to estimate. Otherwise, the bias of the secondary data would be passed along to the estimate as well.

5.6.2 'Effective' data spacing scale

The reduced average estimation variance for cokriging allows the application of 'effective' data spacing as defined above. This results in a spacing of 758 feet. The average primary data spacing for the domain was 1070 feet while the additive data spacing of primary and secondary data was 215 feet. The new 'effective' data spacing represents a covariance and correlation tuned application of the secondary data. It has an obvious impact but much more conservative than the spacing from direct combination. The relatively small decrease from primary spacing is somewhat surprising considering the high

correlation between data types but can be expected due to the modelled relationship between average estimation variance and correlation at each spacing.

The 'effective' spacing is entirely dependant on correlation, as defined by f, Equation 4.1. $\rho_{k,k'}$ is required to define f. Referring back to Figure 4.2 the possible utility of the secondary data increases as correlation increases. However, it is noted that, even at lower correlation values there is some utility in the secondary. The theoretical study of reducing secondary data at increments of correlation from Chapter 4 shows that, even at increments as low as $\rho = 0.2$ there can be significant reduction in average estimation variance when secondary data are abundant as compared to primary data.

Though references have been made to 'average data spacing' it is better thought of in terms of average data density. The GSLIB program *DataSpacing* (Wilde and Deutsch, 2011) was used to define average spacing. This program uses domain volume by generic shape to calculate spacing within the domain.

Additionally, it is noted that, during the random removal of secondary data, there is a relatively small effect on average spacing until approximately half of the data is removed. The spacing increase with data removal shows an approximate exponential increase.

5.6.3 Decisions for application

The methods for establishing 'effective' data spacing have been established as sound. However, the decision to apply these methods depends on the specifics of the deposit. Here, the domain was defined with the eventual application of 'effective' data spacing in mind. This means that the domain was a subset of what is normally considered a domain for definition of resources and reserves in this deposit. This seems to be a likelihood when dealing with unequally sampled data. Overlap of sampling grids will often not be complete and areas without primary data must be considered separately.

In the example correlation is very high, 0.85. As noted previously, some would consider this high enough to be grouped together for estimation. However, it is shown from comparison of average estimation variance that direct data combination does not provide minimum variance nor is it correct in terms

of error. This correlation value does allow for a good cokriged estimate and a degree of confidence when applying 'effective' data spacing. At lower correlation values a QP may be less confident. Equation 4.1 will provide the 'effective' data spacing as defined by the relationship of correlation and data spacing. It is possible to adjust the exponent applied in Equation 4.4 to reflect a more conservative 'effective' data spacing model. If a data spacing study is completed as done here then the process is provable and defensible.

It must be pointed out that, while this study portrays the application of 'effective' data spacing well it is not a viable application. In most cases, if a deposit is at the point of having this abundant blast hole assay data available it is beyond the application of 'effective' data spacing. This process applies further upstream and would be more likely to consider secondaries such as legacy diamond drilling data sets, soil samples, or possibly RC drilling assays.

Chapter 6

Conclusions

Improving estimation is an important goal for any mining company. Achieving this at little cost is attractive. The key idea is to make full use of all data sources available while carefully dealing with bias, error, and the computational challenges of unequal sampling. Cokriging provides a practical solutions to these problems. This thesis has developed important implementation and practical details that permit use of cokriging for this purpose.

In application to actual exploration and production data in the case study of Chapter 5 it is seen that bias and error in the secondary data are not transmitted through to finals estimates. A comparison with kriging using primary data only and kriging of primary and secondary data shows that inclusion of the secondary data with the proposed methodology improves the estimates significantly.

Another development of this thesis is the use of 'effective' data spacing that considers the improvement due to secondary data. The 'effective' data spacing is shown to be reasonable in a case study with data from the Red Dog mine. This improved data spacing could support a better resource classification.

6.1 Research contributions

This work has been an extension on previously employed methods. The gains made here are in the vein of practical application. They confirm previous theory and extend to show possibilities in cokriging application.

Building on the methods shown by Minnitt and Deutsch (2014) the pro-

jection of the cross covariance to collocation is shown to work viably in a case study with real data. While this does not allow a comparison with a known underlying distribution it does allow comparison with conventional estimates. A comparison can be made with data left out of the estimation procedure; this was done and the estimates are indeed improved.

In cases where cokriging improves estimation, a reduced data spacing can be defined. This is referred to as an 'effective' data spacing. In concept and case study the relationship between average estimation variance due to correlation with secondary data is modelled and is shown to provide an improved assessment of the available data. An equation for 'effective' data spacing is developed and allows a weighting of the secondary data to effectively scale its contribution to average data spacing in the domain.

A update to the GSLIB cokriging program was undertaken to implement practical aspects that were not available in the previous programs. The additional features include the use of multiple data files, standardization of variables, cross validation, easy primary variable selection, and more natural variogram input specifications.

6.2 Future work

Throughout the course of research there have been points identified that may bear further study. Additionally, there is opportunity for extension of the research put forward here.

The results of 'effective' data spacing show a logical relationship between correlation and average estimation variance. However, a further confirmation would be suggested in the form of comparison with mining reconciliation data. It may be possible to fine tune or calibrate contribution of the secondary data at the time of the study.

Additionally, the 'effective' data spacing currently applies only to methods that directly consider classification based on spacing. There may be an opportunity to extend this to other methods of classification. Most methods are, in some way, reliant on data spacing. For example, if classification is based on

a search criteria such as a minimum number of data with fixed radius, then this could be modified to include primary and secondary data with appropriate weighting. This would require establishment of a localized 'effective' data spacing. This may also allow a much improved 'effective' data spacing calculation in reference to unequal primary and secondary data grids.

This work considers the data to be stationary in terms of the mean and other statistical parameters including the variogram. Extension to non-stationary data may increase the range of application of the developed methods.

A single secondary variable was used here to illustrate methods. There is room for implementation and case study examples that include several secondary variables.

Non-linear relationships between data types are not explored here. The data types chosen represent the same variable in the same units. However, many secondary data could provide an opportunity to explore the possible non-linear relationships between primary and secondary data.

Data imputation is a possible methodology that could be used as an alternative to cokriging with unequally sampled data. In these terms, cokriging with the cross covariance extrapolation method could be applied in imputation. The simulated collocated data could provide an opportunity to apply alternative estimation methods (Barnett and Deutsch, 2013).

Finally, there is a need for some supporting programs including the variogram input conversion from original GSLIB format to variogram matrix format and the calculation of 'effective' data spacing.

Bibliography

Andrews, G. C. (2005). *Canadian Professional Engineering and Geosciences: Practice and Ethics*. Thompson Nelson.

Babak, O. and Deutsch, C. V. (2006). Comparison of cokriging with an lmc versus the intrinsic model. *CCG Annual Report 9*.

Barnett, R. M. (2011). *CCG Guidebook Series Vol. 13: Tools for Multivariate Geostatistical Modeling*. University of Alberta.

Barnett, R. M. and Deutsch, C. V. (2013). Imputation of geologic data. *CCG Annual Report 15*, (102).

Bertoli, O., Paul, A., Casley, Z., and Dunn, D. (2012). Geostatistical drillhole spacing analysis for coal resource classification in the bowen basin, queensland. *International Journal of Coal Geology*.

Besaw, L. E. and Rizzo, D. M. (2007). Stochastic simulation and spatial estimation with multiple data types using artifical neural networks. *Water Resource Research*, 43.

Canadian Securities Administrators (2011). National instrument 43-101. Retrieved from http://web.cim.org/standards/documents/Block484_Doc111.pdf.

Chica-Olmo, J. (2007). Prediction of housing location price by a multivariate spatial method: Cokriging. *Journal of Real Estate Research*, 29(1).

Cuba, M., Babak, O., and Leuangthong, O. (2009). On the selection of secondary variables for cokriging and cosimulation. *CCG Annual Report 11*.

Deutsch, C. V. (2010). Dimodel of cross validation/jackknife results. *CCG Annual Report 12*.

Deutsch, C. V. (2014). Despiking. Mine613 course notes.

Deutsch, C. V. and Journel, A. G. (1992). *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press.

Deutsch, J. L. and Deutsch, C. V. (2013). Uncertainty checking and the assessment of spatial predictions. *CCG Annual Report 15*.

Gandin, L. (1963). *Objective Analysis of Meteorological Fields*. U.S Department of Commerce.

Goovaerts, P. (1997). *Geostatistics for Natural Resource Evaluation*. Oxford University Press.

Gotway, C. A. and Hartford, A. H. (1996). Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables. *Journal of Agriculture, Biological, and Environmental Statistics*, 1(1):17–38.

Goulard, M. and Voltz, M. (1992). Linear coregrionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3):269–286.

Henley, S. (2004). The russian reserves & resource reporting system. Technical report, Resource Computing International Ltd.

Huang, Y.-T. (2014). Integrative modeling of multiple genomic data from different types of genetic association studies. *Biostatisticss*, 15(4):587–602.

Isaaks, E. H. and Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. Oxford.

Joint Ore Reserves Committee (2012). Australasian code for reporting of exploration results, minerals resource and ore reserves. Retrieved from http://www.jorc.org/docs/jorc_code2012.pdf.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics*. Academic Press.

Knapp, R. (2004). Red dog mine closure and reclaimation plan: Supporting document b1 plans of operations. Technical report, Teck Cominco American Incorperated.

Knotters, M., Brus, D., and Oude Voshaar, J. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67:227–246.

Lark, R., Bellamy, P., and Rawlins, B. (2006). Spatio-temporal variability of some metal concentrations in the soil of eastern england, and implications for soil monitoring. *Geoderma*, 133:363–379.

Leuangthong, O., Khan, K. D., and Deutsch, C. V. (2008). *Solved Problems in Geostatistics*. Wiley.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

McLennan, J. A., Leuangthong, O., and Deutsch, C. V. (2006). Another look at the kriging equations. *CCG Annual Report 8*.

Minnit, R. and Deutsch, C. V. (2014). Cokriging for optimal mineral resource estimates in mining operations. *The Journal of The South African Institute of Mining and Metallurgy*, 114:189–203.

Myers, D. E. (1984). Co-kriging - new developments.

Neufeld, C. (2004). varfit lmc. Retrieved from https://ccgsrv.geostats.ualberta.ca/ccgkb/doku.php?id=software:varfit_lmc.

Nicholson, J. (2014). *The Concise Oxford Dictionary of Mathematics*. Oxford University Press, 5 edition.

Niven, E. B. and Deutsch, C. V. (2011). Calculating a robust correlation coefficient and quantifying its uncertainty. *Computers and Geosciences*, 40:1–9.

Odeh, I., McBratney, A., and Chittleborough, D. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, 67:215–226.

of Mining Metalurgy, C. I. and Petroleum (2014). Cim definition standards - for mineral resource and mineral reserves.

Pan-European Reserves & Resources Reporting Committee (2008). Perc reporting standard 2013: Pan-european standard for reporting of exploration results, mineral resource and reserves. Retrieved from http://web.cim.org/standards/documents/Block472_Doc93.pdf.

Pawlowsky, V. (1989). Cokriging of regionalized compositions. *Mathematical Geology*, 21(5):513–521.

Pawlowsky, V., Olea, R., and Davis, J. (1995). Estimation of regionalized compositions: A comparison of three methods. *Mathematical Geology*, 27(1):105–127.

Pyrcz, M. J. and Deutsch, C. V. (2014). *Geostatistical Reservoir Modeling*. Oxford, 2nd edition.

Ramcharan, A. J. (2015). Mineral reserves reporting standards. Retrieved from http://technology.infomine.com/reviews/reportingstandards.

Rossi, M. E. and Deutsch, C. V. (2014). *Mineral Resource Estimation*. Springer.

Sanchez-Borrego, I., Opsomer, J., Rueda, M., and Arcos, A. (2014). Nonparametric estimation with mixed data types in survey sampling. *Revista Matematica Complutense*, 27:685–700.

Silva, D. S. and Boisvert, J. B. (2014). Two new tools: Directional survey to gslib xyz format and drill hole spacing. *CCG Annual Report 16*.

The Southern African Institute of Mining, and Metalurgy, and the Geological Society of South Africa (2009). Samcode: The south african mineral codes. Retrieved from http://www.samcode.co.za/downloads/SAMREC2009.pdf.

United States Environmental Protection Agency (2003). Guidance for geospatial data quality assurance project plans. Retrieved from http://www.epa.gov/quality/qs-docs/g5g-final.pdf.

United States Securities, and Exchange Commission (2015). Indsutry guide 7: Description of property by issuers engaged or to engaged in significant mining operations. Retirved from http://web.cim.org/standards/documents/Block474_Doc32.pdf.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting a model for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:2758–294.

Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer.

Wawruch, T., Deutsch, C. V., and McLennan, J. (2002). Geostatistical analysis of multiple data types that are not available ar the same location. *CCG Annual Report 4*.

Wilde, B. and Deutsch, C. (2011). Updated program for determining data spacing. *CCG Annual Report 14*.

Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, 18:125–139.

Zhao, Q., Ge, Z., Wang, J., and Xu, J. (2006). Robust traffic matrix estimation with imperfect information: Making use of multiple data sources. *College of Computing, Georgia Institue of Technology*.

Appendix A

COKB3D update

It has been noted through this authors research that some updating of the GSLIB program Cokb3d in terms of inputs and options would be beneficial. Additional abilities, such as cross-validation via jackknife, and the option to standardize the data are noted as helpful additions. Some parameter input changes are required to facilitate program use. These include adjusted data file input, reformat of the variogram parameter inputs, and primary/secondary data selection. A specific input that is considered redundant is the collocated cokriging option.

All changes made by this update are checked using both natural and mock data sets to tests the program performance over as many data configuration as may practically be encountered.

The option for collocated cokriging was removed prior to any other changes in coding. With the updated code allowing many files for variable input this option is no longer required for gridded covariate input. Collocated cokriging can then be accomplished simply by input of the required variogram parameters. The related code has been commented out in the program but subsequent code adjustments discussed below do not consider further usage of the collocated cokriging code. This bars usage of the collocated cokriging function in conjunction with the updates unless the user is willing to code the required changes and test for validity.

In the previous version of the COKB3D program the variogram inputs were similar to those found in most GSLIB parameter files. This format consists of a listing of the variogram parameters (ranges, angles, contributions) for each

```
                  Parameters for COKB3D
                  *********************

START OF PARAMETERS:
somedata.dat                        -file with data
2                                   -   number of variables primary+other
0   1   2   3   4   5               -   columns for X,Y,Z and variables
-0.01     1.0e21                    -   trimming limits
0                                   -co-located cokriging? (0=no, 1=yes)
somedata.dat                        -   file with gridded covariate
4                                   -   column for covariate
3                                   -debugging level: 0,1,2,3
cokb3d.dbg                          -file for debugging output
cokb3d.out                          -file for output
50   0.5   1.0                     -nx,xmn,xsiz
50   0.5   1.0                     -ny,ymn,ysiz
10   0.5   1.0                     -nz,zmn,zsiz
1    1     1                       -x, y, and z block discretization
1   12     8                       -min primary,max primary,max all sec
25.0  25.0  25.0                   -maximum search radii: primary
10.0  10.0  10.0                   -maximum search radii: all secondary
 0.0   0.0   0.0                   -angles for search ellipsoid
2                                   -kriging type (0=SK, 1=OK, 2=OK-trad)
3.38  2.32  0.00  0.00              -mean(i),i=1,nvar
1     1                             -semivariogram for "i" and "j"
1   11.0                            -   nst, nugget effect
1   39.0  0.0   0.0   0.0          -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert
1     2                             -semivariogram for "i" and "j"
1    0.0                            -   nst, nugget effect
1   14.5  0.0   0.0   0.0          -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert
2     2                             -semivariogram for "i" and "j"
1    9.0                            -   nst, nugget effect
1   15.0  0.0   0.0   0.0          -   it,cc,ang1,ang2,ang3
         60.0  60.0  60.0          -   a_hmax, a_hmin, a_vert
```

Figure A.1: The original COKB3D parameter file.

structure in each direct and cross variogram. This format, while familiar to GSLIB users, is somewhat repetitive. The updated form uses the variogram matrix format. This allows the use of all previous information indexed by number of structures. For example, if an LMC defines two direct and one cross variogram each with a single structure then inputs will be one 2x2 nugget effect matrix and a single 2x2 variogram sill contribution matrix for the single structure. Note that all the information previously provided for each structure and each variogram is now provided for each structure only (i.e. structure type, ranges, angles, etc.). This matrix input format is familiar to all geostatisticians and can be quite helpful particularly in confirmation of the positive definiteness of the variogram matrices developed for the LMC. Depending on the number of variables being considered the input matrix parameters can sim-

ply be increased in size to the program maximum of four. This decision was made in order to constrain the extensive code required when iteration could not be performed in automated loops. Expansion of this constraint is possible if required by the user.

To facilitate the change from variogram directed inputs to structure directed inputs an additional variable was defined signifying total number of structures. This value is used to index the structure contribution matrices. The original code for reading of the variogram parameters was maintained to avoid lengthy changes to the COKB3D subroutine. The new parameters are simply read to temporary memory in the previous parameter format and then called as required by the original subroutine code.

Checking of these changes was performed using example parameters defined in Deutsch and Journel (1998). Along with the input parameters the authors provide the estimated values and estimation variances for the first and last ten estimates. These values were used for direct comparison with those produced by the updated code and were found to be in agreement.

Cross-validation via jackknife is added to the COKB3D program as part of the software update. This ability was an addition of code already available in KT3D and so the output file is of the same format as that produced by cross-validation in KT3D. This is a time saving addition that is instituted with relatively small changes to the COKB3D code. It is noted that some adjustment to the COKB3D subroutine is required in order to loop the data search correctly and return only the estimates at known data locations. As variables used in the COKB3D subroutine are indexed in arrays it is not possibly to directly simply input the code harvested from KT3D.

Checking of the cross-validation additions is done by direct comparison with the known data. This is achieved using previous estimates and by reference to the exhaustive debugging file from those previous estimates. Reproduction of known data points confirms that the cross-validation by jackknife code update is valid and performing correctly.

Adjustments related to the import of data constitute the majority of this update. In previous iterations of the COKB3D program the user was required

```
                  Parameters for COKB3D
                  *********************

START OF PARAMETERS:
2                                 -   number of variables
1                                 - primary variable (to be estimated)
0                                 - standardize data? (0=No, 1=Yes)
2                                 - number of data files
somedata1.dat                     - data file 1
1                                 - # of variables in data file
1   2   3   4   5   0  0          -   columns for DH,X,Y,Z and variables
somedata2.dat                     - data file 2
1                                 - # of variables in data file
1   2   3   4   5  0  0          -   columns for DH,X,Y,Z and variables
-0.01     1.0e21                  -   trimming limits
0                                 -option: 0=grid, 1=jackknife
xvk.dat                           -   file with jackknife data
1  2  3  4  5  6                  -   column for DH,X,Y,Z,vr and sec vr
3                                 -debugging level: 0,1,2,3
cokb3d.dbg                        -file for debugging output
cokb3d.out                        -file for output
50   0.5   1.0                   -nx,xmn,xsiz
50   0.5   1.0                   -ny,ymn,ysiz
10   0.5   1.0                   -nz,zmn,zsiz
1    1     1                     -x, y, and z block discretization
1   12     8                     -min primary,max primary,max all sec
25.0  25.0  25.0                 -maximum search radii: primary
10.0  10.0  10.0                 -maximum search radii: all secondary
 0.0   0.0   0.0                 -angles for search ellipsoid
2                                 -kriging type (0=SK, 1=OK, 2=OK-trad)
3.38  2.32  0.00  0.00            -mean(i),i=1,nvar
4.16  3.25  0.00  0.00            -std dev(i),i=1,nvar
11.0   0.0                       -nugget effect (nvar x nvar)
0.0    9.0                       -nugget effect (nvar x nvar)
1                                 -number of structures
1                                 -structure number
60.0  60.0  60.0                 - a_hmax, a_hmin, a_vert
1  0.0  0.0  0.0                 -structure 1 it, ang1,ang2,ang3
39.0   14.5                      -   contributions (nvar x nvar)
14.5   15.0                      -   contributions (nvar x nvar)
```

Figure A.2: The updated COKB3D parameter file.

to accumulate all variables in a single data file in order to facilitate the single input allowed by the program. In this update the user can directly input multiple data files each consisting of multiple variables. The input follows the same format as the previous single file input but repeated sequentially for the number of files required. Note that the user can copy the input format. Due to the limit of four variables, having multiple variables in a single file will reduce the number of files the user is able to input. To accommodate the combination of multiple files the data from each file be read into the program with the count of the data in each variable being recorded. The corresponding data is then assigned as primary and secondary variables based on order of

input. All associated coordinate values are assigned last based on the file from which they were taken.

A practitioner may have many variables of interest among the input variables. In a time-saving measure the ability to change the variable of interest is added. This is a change of the primary variable indicator based on the order in which the variables have been input. For example, if file one contains one variable and file two contains one variable then these variables are signified as one and two respectively. Estimating the second variable would require only that the primary variable indicator be changed to '2'. In code, to avoid modifying the COKB3D subroutine, the changes are made prior to estimation by redirecting which input variable is assigned to each of the variables to estimate (vr, sec1, sec2, and sec3 representing the primary and each secondary data, in that order). This requires re-assignment of all affected data, variograms, input variable means, and standard deviations. The code is designed to read which variable is selected to be the primary variable and, if it is not the first variable input, re-assign that variable and its associated data and parameters to the position of the first variable. The former first variable, along with associated data and parameters, is then re-assigned to the former position of the variable of interest. Note that primary variable selection is based on the order in which the variables are entered into the parameter file beginning with the first entered variable in the first entered file and counting up.

The code which re-assigns the data and parameters is divided into two sections based on the previous code update allowing matrix input. The data values, means, and standard deviations are reassigned using a temporary memory. The data values can be redirected just prior to the COKB3D subroutine because the number of data has been counted for each variable and so each can be re-assigned as needed. However, as mentioned, the changes made to accommodate variogram matrix input complicate the re-assignment of the required rows and columns from the matrix. The code addition re-directing the required variogram input is introduced after the reading of the initial matrices but prior to modifying the values to index format. This allows the variogram re-assignment but to also maintains the indexing used in the COKB3D sub-

routine.

The final addition to the import of data is the ability to standardize data prior to estimation. Many practitioners require standardization to allow for ease of use and comparison. This change required the addition of the standard deviation to the parameter file. The additional code requires simply that the standardization formula be applied to each data point of each variable before the COKB3D subroutine begins. Once estimation is complete the standardization is reversed prior to output of estimates. Therefore, the outputs from this estimate will be the same as the outputs from a non-standardized estimate. The only additional note is that a standard linear model of coregionalization is required if applying the data standardization option.

Changes made during this update have been checked individually as well as in final debugging. The initial stages of testing required that, after each code adjustment, the entire program be debugged, compiled, and run for testing. Estimates produced were compared with those produced from the previous edition of the COKB3D program. Adjustments, debugging and compiling was repeated until estimates were reproduced. In the case of the additional jackknife code a small mock data set was produced including eight data points assigned to two variables. The program was compiled and run with the jackknife option selected. The jackknife code was tested for validity based on appropriate number of data and covariances used in each estimate as seen in the debug file. Seven data points were used for estimate along with eight covariances for the file containing eight data points.

A final check of the completed program was done using a mock data file containing four variables of five data points each. This data file can be observed in Figure A.3. The file was called in several different ways in several different tests of the multiple file- multiple variable input update. For example, the option for four files with one variable each was tested using this single files called four time with a different variable indicated upon each call. Up to four variables over four files (and any lesser quantities) can be handled by the current code. Jackknife was tested once again to ensure proper number of samples and covariances are used. This was shown to be robust as well.

Finally, standardization was checked via comparison with estimates produced from the previous COKB3D program using previously standardized data. Estimates were shown to be reproduced using non-standardized data as the input for the updated program. Note that the same standardized linear model of coregionalization was used for both.

```
Cokriging Mock file
 7
X
Y
Z
Zn
Pb
Ba
Au
 1  1  0.5  44.344 -99 -99 -99
 1  2  0.5  40.682 -99 -99 -99
 2  1  0.5  41.069 -99 -99 -99
 2  2  0.5  42.049 -99 -99 -99
 3  1  0.5  45.936 -99 -99 -99
 0.5  0.5  0.5 -99  7.82 -99 -99
 0.5  1  0.5 -99  9.844 -99 -99
 1  0.5  0.5 -99  10.764 -99 -99
 1  1.2  0.5 -99  18.621 -99 -99
 1.5  1.7  0.5 -99  7.286 -99 -99
 0.7  0.7  0.5 -99 -99  44.482 -99
 0.7  1.4  0.5 -99 -99  47.297 -99
 1.4  0.7  0.5 -99 -99  44.62 -99
 1.4  1.4  0.5 -99 -99  25.944 -99
 2.1  0.7  0.5 -99 -99  18.676 -99
 0.3  0.3  0.5 -99 -99 -99  0.862
 0.3  1.2  0.5 -99 -99 -99  0.796
 1.2  0.3  0.5 -99 -99 -99  0.812
 1.2  1.2  0.5 -99 -99 -99  0.845
 2.1  1.2  0.5 -99 -99 -99  0.926
```

Figure A.3: The mock data file created for program validation.

Appendix B

Estimation parameters

Included here are the parameters used in each of the three estimates performed for the case study of Chapter 5. This provides transparency in the process as well as detailed information for those interested in reproducing the study. Note that OK estimates were performed using $KT3D$ version 3.0 and SCK using $COKB3D$ version 2.0.

```
                  Parameters for KT3D
                  *******************

START OF PARAMETERS:
DDHdata.dat                       -file with data
1  2  3  4  6  0                  -   columns for DH,X,Y,Z,var,sec var
-10.0   1.0e21                    -   trimming limits
0                                 -option: 0=grid, 1=cross, 2=jackknife
xvk.dat                           -file with jackknife data
1   2   0    3    0               -   columns for X,Y,Z,vr and sec var
3                                 -debugging level: 0,1,2,3
kt3d.dbg                          -file for debugging output
kt3d.out                          -file for kriged output
56   586500.0    25.0             -nx,xmn,xsiz
48   145000.0    25.0             -ny,ymn,ysiz
14   750.0    25.0                -nz,zmn,zsiz
3    3     1                      -x,y and z block discretization
4    8                            -min, max data for kriging
0                                 -max per octant (0-> not used)
500.0  500.0  300.0               -maximum search radii
 0.0   0.0   0.0                  -angles for search ellipsoid
1     12.41                       -0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0                 -drift: x,y,z,xx,yy,zz,xy,xz,zy
0                                 -0, variable; 1, estimate trend
extdrift.dat                      -gridded file with drift/mean
4                                 -  column number in gridded file
2    0.0                          -nst, nugget effect
1    0.5 15.0   105.0   0.0       -it,cc,ang1,ang2,ang3
      120.0  120.0  50.0          -a_hmax, a_hmin, a_vert
1   0.5  15.0   105.0   0.0       -it,cc,ang1,ang2,ang3
      350.0  350.0  130.0         -a_hmax, a_hmin, a_vert
```

Figure B.1: Input parameters for the OK estimate using diamond drilling composite data only.

```
                  Parameters for KT3D
                  *******************

START OF PARAMETERS:
DataSTDall.dat                    -file with data
0  1  2  3  4  0                  -   columns for DH,X,Y,Z,var,sec var
-5.0   1.0e21                     -   trimming limits
0                                 -option: 0=grid, 1=cross, 2=jackknife
xvk.dat                           -file with jackknife data
1   2   0    3    0               -   columns for X,Y,Z,vr and sec var
3                                 -debugging level: 0,1,2,3
kt3d_STDall.dbg                   -file for debugging output
kt3d_STDall.out                   -file for kriged output
56   586500.0    25.0             -nx,xmn,xsiz
48   145000.0    25.0             -ny,ymn,ysiz
14   750.0    25.0                -nz,zmn,zsiz
3    3      1                     -x,y and z block discretization
4    8                           -min, max data for kriging
0                                 -max per octant (0-> not used)
500.0  500.0  300.0               -maximum search radii
 0.0   0.0   0.0                  -angles for search ellipsoid
1     12.41                       -0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0                 -drift: x,y,z,xx,yy,zz,xy,xz,zy
0                                 -0, variable; 1, estimate trend
extdrift.dat                      -gridded file with drift/mean
4                                 -  column number in gridded file
2    0.0                          -nst, nugget effect
1    0.5 15.0   105.0   0.0       -it,cc,ang1,ang2,ang3
      120.0  120.0  50.0           -a_hmax, a_hmin, a_vert
1   0.5  15.0   105.0   0.0       -it,cc,ang1,ang2,ang3
      350.0  350.0  130.0          -a_hmax, a_hmin, a_vert
```

Figure B.2: Input parameters for the OK estimate using combined diamond drilling assay composites and blast hole assay data.

```
                  Parameters for COKB3D
                  *********************

START OF PARAMETERS:
2                                 -   number of variables
1                                 - primary variable (to be estimated)
0                                 - standardize data? (0=No, 1=Yes)
1                                 - number of data files
SDdata.dat                        - data file 1
2                                 - # of variables in data file
1   2   3   4   17  19            -   columns for DH,X,Y,Z and variables
-10.0     1.0e21                  -   trimming limits
0                                 -option: 0=grid, 1=jackknife
data.dat                          -   file with jackknife data
1  2  3  4  9  0                  -   column for DH,X,Y,Z,vr and sec vr
3                                 -debugging level: 0,1,2,3
cokb3d.dbg                        -file for debugging output
cokb3d.out                        -file for output
56   586500   25.0                -nx,xmn,xsiz
48   145000   25.0                -ny,ymn,ysiz
14   750   25.0                   -nz,zmn,zsiz
3    3     1                      -x, y, and z block discretization
1   8     4                       -min primary,max primary,max all sec
600.0  600.0 300.0                -maximum search radii: primary
600.0  600.0 300.0                -maximum search radii: all secondary
 15.0   0.0   0.0                 -angles for search ellipsoid
0                                 -kriging type (0=SK, 1=OK, 2=OK-trad)
0.00  0.00  0.00  0.00            -mean(i),i=1,nvar
1.00  1.00  1.00  0.00            -std dev(i),i=1,nvar
0.05   0.0                        -nugget effect (nvar x nvar)
0.0   0.05                        -nugget effect (nvar x nvar)
2                                 -number of structures
1                                 -structure number
30.0  50.0  60.0                  - a_hmax, a_hmin, a_vert
1  15.0  0.0  0.0                 -structure 1 it, ang1,ang2,ang3
0.461   0.255                     -   contributions (nvar x nvar)
0.255   0.149                     -   contributions (nvar x nvar)
2                                 -structure number
300.0  300.0  175.0               - a_hmax, a_hmin, a_vert
1      15.0  0.0  0.0             -structure 2 it, ang1,ang2,ang3
0.489   0.595                     -   contributions (nvar x nvar)
0.595   0.801                     -   contributions (nvar x nvar)
```

Figure B.3: Input parameters for the SCK estimate.