2003

Multivariate geostatistical analysis with application to a Western Australian nickel laterite deposit

Ellen Bandarian
Edith Cowan University

Edith Cowan University

Copyright Warning

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

MULTIVARIATE GEOSTATISTICAL ANALYSIS WITH APPLICATION TO A WESTERN AUSTRALIAN NICKEL LATERITE DEPOSIT

BY

ELLEN BANDARIAN

A Thesis Submitted to the

Faculty of Computing, Health and Science

Edith Cowan University

Perth, Western Australia

In Partial Fulfilment of the Requirements for the Degree of

Bachelor of Science Honours (Mathematics)

June 2003

Supervisors: Associate Professor Lyn Bloom

Dr Ute Mueller

ABSTRACT

Although it can be time consuming and computationally more expensive to work with multivariate data, it is often desirable to exploit the relationships among and between the variables sampled across a study region. In most cases the availability of this secondary information can enhance the estimation of the primary variable(s). The aim of this research is to describe and then demonstrate the use of multivariate statistical methods in geostatistical analysis. The methods will be illustrated by application to a multivariate data set from an actual mineralisation. The data suite is known as MM22D and comes from the Murrin Murrin nickel mine near Laverton in Western Australia. Two four variable subsets of the MM22D data suite are used in the application. The first is MM22DHC4 and consists of nickel, cobalt, iron and zinc, chosen for their high correlations with each other. The second is MM22DTOP4 and consists of nickel, cobalt, magnesium and iron, chosen for their economic importance to the mining company.

This thesis presents the theory and the process of the modelling and estimation of multivariate data. We demonstrate the use of principal component analysis in a geostatistical environment, in particular for the detection of intrinsic correlation. We illustrate the modelling and estimation of an intrinsically correlated data set (MM22DHC4) and estimate the variables in this data set using ordinary cokriging, principal component kriging and ordinary kriging. In addition we illustrate the derivation of a general linear model of coregionalisation and estimate the variables of this data set (MM22DTOP4) using ordinary cokriging and ordinary kriging. The grade control data from the MM22D data suite, which were considered reality, were used as a comparison and assessment of the accuracy of all of the estimates. As the data used in this study were isotopic it was anticipated that there would be little difference in the estimates obtained which was indeed the case.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text; or

(iii) contain any defamatory material.

ACKNOWLEDGEMENTS

I wish to thank the Faculty of Computing, Health and Science, in particular Professor Patrick Garnett for the honour of receiving the Dean's Scholarship in 1997. I would like to express my appreciation to Anaconda Operations Pty Ltd for making the Murrin Murrin data available to me. I would also like to extend my thanks to The Statistical Society of Australia Inc for the privilege of being awarded the Honours Scholarship 2003. My thanks go also to my supervisors Associate Professor Lyn Bloom and Doctor Ute Mueller whose patience, motivation and support made this project possible.

To my Mum and Maryann I convey my sincere appreciation for all of the extra babysitting they have done. Thanks also to Dad and Odette for their past and ongoing encouragement. Finally and most importantly my heartfelt thanks goes to my husband Norm and my two beautiful daughters Katherine and Emily. Their support, encouragement and enthusiasm for my studies has been and continues to be unwavering.

TABLE OF CONTENTS

Abstract 2

Declaration 3

Acknowledgements 4

1 Introduction 7

1.1 Background and Significance 7

1.2 Objectives 9

1.3 Thesis Outline 10

1.4 Software 10

1.5 Notation 11

1.6 Acronyms and Abbreviations 13

2 Theoretical Framework 15

2.1 Principal Component Analysis 15

2.2 Multivariate Random Function Model 18

2.3 Modelling the Coregionalisation 21

2.4 Kriging 27

2.4.1 Simple and Ordinary Kriging 28

2.4.2 Ordinary Cokriging 30

2.4.3 Principal Component Kriging 32

3 Data Analysis 34

3.1 MM22D Data Suite 34

3.2 Exploratory Data Analysis 36

3.3 Spatial Data Analysis 43

3.4 Principal Component Analysis 48

3.4.1 Principal Component Analysis of MM22DEXP 50

3.4.2 Principal Component Analysis of MM22DHC4 55

3.4.3 Principal Component Analysis of MM22DTOP4 58

4 Ordinary Cokriging 62

4.1 Variography and Cokriging of MM22DHC4 62

4.1.1 Variography and the Intrinsic Coregionalisation Model 62

4.1.2 Cross Validation for Cokriging of MM22DHC4 67

4.1.3 Cokriging Estimates of MM22DHC4 70

4.2 Variography and Cokriging of MM22DTOP4 74

4.2.1 Variography and the Linear Model of Coregionalisation 75

4.2.2 Cross Validation for Cokriging of MM22DTOP4 79

4.2.3 Cokriging Estimates of MM22DTOP4 81

5 Principal Component Kriging 84

5.1 Variography of the Principal Components 84

5.2 Cross Validation for Ordinary Kriging of the Principal Components 89

5.3 Principal component Kriging Estimates 92

6 Ordinary Kriging 97

6.1 Variography of Nickel, Cobalt, Magnesium, Iron and Zinc 97

6.2 Cross Validation for Ordinary Kriging 104

6.3 Ordinary Kriging Estimates of Nickel, Cobalt, Magnesium and Iron 107

7 Comparison of Estimates 109

8 Discussion and Conclusion 126

References 131

Appendices 133

Appendix A Principal Component Analysis of Transformed Data 134

Appendix B Cross Validation Results 139

Appendix C ISATIS Parameter Files 152

1 INTRODUCTION

1.1 Background and Significance

Geostatistics enables us to analyse spatially dependent data, that is, data where location as well as value is important. Such data arise naturally in the earth, mining, petroleum and environmental sciences. Geostatistics can be described as a set of statistical tools that allow us to describe, interpret and model the spatial continuity that is a fundamental feature of many natural phenomena. Furthermore once we have a mathematical representation of the spatial continuity of the data we can use this model to estimate values at unsampled locations across the study region. Geostatistical methods generally provide better estimates than traditional methods and most importantly provide us with a measure of the accuracy of the estimates.

This rapidly advancing area of applied mathematics has its origins in the mining industry in South Africa in the 1950's. Mining engineer D. G. Krige and statistician H. S. Sichel were among the first to implement new statistical methods that did not rely on traditional methods based on normally distributed data. In the 1960's G. Matheron further developed and formalised these innovative ideas and introduced the concept of a regionalised variable which he defined as a spatially distributed phenomenon that exhibits a particular spatial structure consisting of both random and structured aspects (Matheron, 1970, p. 5).

It is often the case that data collection in the earth sciences consists of observations of many variables, some more densely sampled than others. While it is more cumbersome and computationally more expensive to work with multivariate data the availability of auxiliary information can enhance the interpretation and estimation of a primary variable. Multivariate geostatistics takes into account the relationships between and among the variables as well as incorporating their spatial distribution across a region. The relationships between these variables can be identified and summarised using methods of multivariate statistical analysis.

One classical, and probably the most commonly used, multivariate statistical technique is principal component analysis which dates back to the early 1900's, in particular to the work of Pearson (1901) and Hotelling (1933). Principal component analysis involves the construction of linear combinations of correlated variables into (preferably) fewer uncorrelated factors that account for the majority of the variation in the original data (Afifi & Clarke, 1996, pp. 330-331). More specifically, principal component analysis is concerned with explaining the variance–covariance structure of a set of variables by means of a few linear combinations of these variables (Johnson & Wichern, 2002, p. 426).

From a geostatistical perspective principal component analysis is a particularly valuable tool as it can be used for reducing the number of variables in a data set and for the detection of intrinsic correlation. One of the benefits of reducing the number of variables in a multivariate data set is the practical advantage of modelling fewer semivariograms. Furthermore, if the orthogonality of the principal components extends to any separation vector $\mathbf{h}$ we may proceed with an intrinsic coregionalisation model. Under this assumption we may perform classical kriging individually on the principal components. This technique is known as principal component kriging and is computationally less expensive than other methods such as cokriging (Goovaerts, 1997, pp. 233-234).

When the data are not intrinsically correlated we must consider not only the spatial variability of the individual variables but we must also take into account the joint variability of each pair of variables. The linear model of coregionalisation is a mathematical model that characterises the spatial variation of a multivariate system at different spatial scales. The requirement of a linear model of coregionalisation is that all direct and cross semivariograms or covariances are jointly modelled and share a common set of basic structures. This then calls for the inference and modelling of $N_v(N_v+1)/2$ direct and cross semivariograms. While this can be a tedious procedure the problem lies in whether the model fits adequately in the mathematical sense (Goulard & Voltz, 1992, p. 269), more precisely the coregionalisation matrices need to exhibit positive semi-definiteness in order for the model to be permissible.

1.2 Objectives

The aim of this study was to describe and demonstrate the use of multivariate statistical techniques in geostatistical analysis. The theory of principal component analysis in a traditional multivariate statistical environment and the use of this technique as it applies to geostatistics are presented. We examine also the theory of the general form of the linear model of coregionalisation as well a particular case known as the intrinsic coregionalisation model. In addition we discuss the theory of the estimation techniques to be used, that is, ordinary kriging, ordinary cokriging and principal component kriging.

For the application of the theory we had four main objectives. Firstly we aimed to exhibit the use of principal component analysis in a geostatistical environment, namely to determine whether the data were intrinsically correlated. Secondly we wished to show examples of both the intrinsic coregionalisation model and the linear model of coregionalisation. Thirdly we aimed to demonstrate the estimation techniques of principal component kriging and ordinary cokriging. Finally we wished to include a comparison of these multivariate techniques to the univariate estimation method of ordinary kriging as well as with 'reality'.

To demonstrate these objectives we analysed a mineralisation (MM22D), which comprises three dimensional grade and thickness measurements on eight variables: nickel, cobalt, magnesium, iron, aluminium, chromium, zinc and manganese. These data came from the Murrin Murrin nickel mine near Laverton in Western Australia. This data suite consists of two data sets: a grade control data set MM22DGC and an exploration data set MM22DEXP. The set MM22DGC comprises grade accumulations for each of the eight variables and will be considered reality for assessment and comparative purposes. The set MM22DEXP is a subset of MM22DGC and will be used to perform the analysis. This data set is isotopic and comprises accumulations for each of the eight variables jointly sampled at 125 locations.

As the data were measured on different scales and have vastly different ranges, means and variances we used the standardised values for the majority of the analyses. Initially we performed a principal component analysis on the MM22DEXP data set and determined that the data were not intrinsically correlated. We then

investigated two four variable subsets of MM22DEXP, denoted as MM22DHC4 and MM22DTOP4. The former subset was intrinsically correlated; hence we used an intrinsic coregionalisation model and preformed the estimation using ordinary cokriging. In addition we individually modelled the principal components from this subset and performed the estimation using principal component kriging. For the latter subset the variables were not intrinsically correlated hence we used a linear model of coregionalisation and performed the estimation using ordinary cokriging. The estimates were then compared to the grade control data, MM22DGC, and the estimates obtained from ordinary kriging.

1.3 Thesis outline

Chapter two of this thesis discusses the theoretical framework relevant to this study. This includes principal component analysis from both classical and geostatistical perspectives, the multivariate random function model and the various methods of kriging, namely ordinary kriging, ordinary cokriging and principal component kriging. Chapter three presents a detailed exploratory data analysis of the MM22D data suite. In chapter four we demonstrate the application of ordinary cokriging using both the MM22DHC4 and MM22DTOP4 data sets. Chapter five presents the application of principal component kriging using the principal components extracted in the eigenanalysis of MM22DHC4. Chapter six presents ordinary kriging of the nickel, cobalt, magnesium, iron and zinc variables. Chapter seven presents a comparison of the estimation techniques used and chapter eight presents a discussion and conclusions of the research.

1.4 Software

There are various software packages available to assist in geostatistical analysis. My study has used primarily the packages listed below. It is appropriate at this point to mention one package in particular, ISATIS, which is recognised as an industry standard. While ISATIS a relatively new package in the workplace, we are in the fortunate position of having it available at this university.

I have taken this opportunity to familiarise myself with the capabilities of ISATIS, to the point where I would now be considered proficient in implementing this package in the workplace. ISATIS offers the geostatistician analytical features and capabilities not previously available. Some of these features include simultaneous modelling of direct and cross semivariograms, sequential cokriging of numerous variables, alternative measures of spatial variability and alternative kriging methods.

3PLOT (Kanevski et al, 1998): post plots of data and estimates

ISATIS (Bleines et al, 2000): estimation

MATLAB 5.3 (1999): matrix manipulation and calculation

MICROSOFT EXCEL (2002): graphical representation of data; spreadsheet calculations

MINITAB 12.1 (1998): summary statistics, principal component analysis, residual analysis

VARIOWIN 2.2 (Pannatier, 1996): semivariogram inference and modelling

1.5 Notation

The notation used throughout this study is a combination of that used by Goovaerts (1997), Wackernagel (1998b) and Johnson and Wichern (2002).

$\forall$: for all

$\mathcal{A}$: study region

a: range parameter

b^l: coefficient of the basic covariance model $c_l(\mathbf{h})$ or semivariogram model $g_l(\mathbf{u})$ in the linear model of regionalisation of the random variable $Z(\mathbf{u})$

b_{ij}^l: coefficient of the basic covariance model $c_l(\mathbf{h})$ or semivariogram model $g_l(\mathbf{u})$ in the linear model of coregionalisation of the random variables $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$

$\mathbf{B}_l$: coregionalisation matrix including the coefficients b_{ij}^l of the basic

covariance model $c_l(\mathbf{h})$ or semivariogram model $g_l(\mathbf{h})$ in the corresponding linear model of coregionalisation $\mathbf{C}(\mathbf{h}) = \sum_{l=0}^{L} \mathbf{B}_l c_l(\mathbf{h})$ or $\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{l=0}^{L} \mathbf{B}_l g_l(\mathbf{h})$

$C(0)$: covariance value at separation distance $|\mathbf{h}| = 0$

$C(\mathbf{h})$: stationary covariance of the random function $\mathcal{Z}$ for lag vector $\mathbf{h}$

$\mathbf{C}(\mathbf{h})$: covariance function matrix of size N_v x N_v

$C_{ij}(\mathbf{h})$: stationary cross covariance between the two random functions $\mathcal{Z}_i$ and $\mathcal{Z}_j$ for a lag vector $\mathbf{h}$

$\mathrm{Cov}\{\cdot\}$: covariance

$c_l(\mathbf{h})$: lth basic covariance model in the linear model of (co)regionalisation

$\mathrm{E}\{\cdot\}$: expected value

$\in$: is an element of

$\boldsymbol{\Gamma}(\mathbf{h})$: semivariogram function matrix of size N_v x N_v

$g_l(\mathbf{h})$: lth basic semivariogram model in the linear model of (co)regionalisation

$\gamma(\mathbf{h})$: stationary semivariogram of the random function $\mathcal{Z}$ for lag vector $\mathbf{h}$

$\gamma_{ij}(\mathbf{h})$: stationary cross semivariogram between the two random functions $\mathcal{Z}_i$ and $\mathcal{Z}_j$ for a lag vector $\mathbf{h}$

$\mathbf{h}$: separation vector

$\lambda_\alpha(\mathbf{u})$: kriging weight associated to z–datum at location $\mathbf{u}_\alpha$ for estimation of the attribute z at location $\mathbf{u}$

$\lambda_{\alpha i}(\mathbf{u})$: cokriging weight associated to z–datum at location $\mathbf{u}_\alpha$ for estimation of the attribute z at location $\mathbf{u}$

m: stationary mean of the random function $Z(\mathbf{u})$

$\mathbf{m}$: vector of stationary means

$m(\mathbf{u})$: expected value of the random variable $Z(\mathbf{u})$

N_v: number of variables Z_i

n: number of data values available over the study region $\mathcal{A}$

$n(\mathbf{u})$: number of data values $z(\mathbf{u}_\alpha)$ used for estimation of the attribute z at location $\mathbf{u}$

$n_i(\mathbf{u})$: number of data values $z_i(\mathbf{u}_\alpha)$ used for estimation of the attribute z at location $\mathbf{u}$

$\mathbf{Q}$: matrix of eigenvectors extracted in the principal component analysis

r_{ij}: linear correlation coefficient between variables Z_i and Z_j

$\mathbf{u}$: coordinate vector

$\mathbf{u}_\alpha$: datum location

$\mathrm{Var}\{\cdot\}$: variance

$Y_k^l(\mathbf{u})$: kth regionalised factor corresponding to the $(l+1)$ basic covariance model in the linear model of coregionalisation $C_{ij}(\mathbf{h}) = \sum_{l=0}^{L} b_{ij}^l c_l(\mathbf{h})$

$Z(\mathbf{u})$: generic continuous random variable at location $\mathbf{u}$

$\mathcal{Z}$: univariate random variable valued function

$\boldsymbol{\mathcal{Z}}$: multivariate random variable valued function

z: continuous variable (attribute)

$z(\mathbf{u})$: true value at unsampled location $\mathbf{u}$

$z(\mathbf{u}_\alpha)$: z-datum value at location $\mathbf{u}_\alpha$

$z_i(\mathbf{u}_\alpha)$: z_i-datum value at location $\mathbf{u}_\alpha$

1.6 Acronyms and Abbreviations

The following is a list of acronyms and abbreviations used in the Figures and Tables throughout this thesis.

AL: aluminium

CO: cobalt

CR: chromium

FE: iron

ICM: intrinsic coregionalisation model

LMC: linear model of coregionalisation

MAE: mean absolute error

MG: magnesium

MN: manganese

MSE: mean square error

NI: nickel

OCK: ordinary cokriging

OK: ordinary kriging

PC: principal component

PCK: principal component kriging

PCK2: principal component kriging (two retained principal components)

ZN: zinc

2 THEORETICAL FRAMEWORK

In this chapter we discuss the theory of principal component analysis from both classical and geostatistical perspectives. We then introduce the multivariate random function model and the linear model of coregionalisation. Finally we discuss the various kriging algorithms employed in this study: ordinary kriging, ordinary cokriging and principal component kriging. The development of the theory and the notation used follows that of Goovaerts (1997), Lay (1997) and Wackernagel (1998a).

2.1 Principal Component Analysis

Almost any kind of data collection in the earth sciences involves simultaneous measurements on many variables. Suppose that we are dealing with N_v variables, that is N_v dimensional data, measured at n locations $\mathbf{u}$ in the region $\mathcal{A}$. While multivariate geostatistics gives us the tools to incorporate this additional information, in practice one seldom considers coregionalisations of N_v greater than three. The reasons for this include notational and computational complexity and difficulties of statistical inference and modelling of the cross covariance or cross semivariograms (Journel & Huijbregts, 1978, p. 173). It is often desirable to exploit the interrelationships among and between the N_v variables by representing them through a few linear combinations of these variables.

One multivariate technique used to achieve this is known as principal component analysis. This is one of the most commonly used methods of multivariate analysis. It is simple to implement and interpretation of the results is often straightforward. In its simplest form a principal component analysis consists of defining a linear transformation that maps a set of N_v correlated variables into N_v uncorrelated principal components (Wackernagel, 1998a, p. 127). From a purely mathematical perspective Johnson and Wichern (2002, p. 426) describe a principal

component analysis as a means of explaining the variance–covariance structure of a set of variables by means of a few linear combinations of these variables.

Each principal component is a linear combination of the original variables and the amount of information conveyed by each principal component is measured by its variance (Afifi & Clarke, 1996, p. 330). The principal components are arranged in order of decreasing variance, thus the first principal component is the most informative and the least informative is the last principal component. One can then choose to retain only the first few principal components that account for the majority of the variability of the original data, making the subsequent analysis simpler. A principal component analysis can also be used to test for normality of the data; if a selected principal component is not normal then neither are the original data. Other classical uses of principal component analysis are to identify outliers and reveal relationships among and between variables that may not have previously been identified. One of the most important geostatistical applications of principal component analysis is to detect intrinsic correlation. This is done by examining the cross correlograms or cross semivariograms of the first few principal components and will be discussed further in section 2.3, Modelling the Coregionalisation.

The basic features of a principal component analysis consist of the extraction of the eigenvalues and eigenvectors of a square, symmetric covariance or correlation matrix. Principal component analysis does not require multivariate normality though the interpretation and application are enhanced when this condition is met. Algebraically the principal components are specific linear combinations of the random variables $Z_i(\mathbf{u})$ for $i = 1, \ldots, N_v$. Geometrically the linear combinations represent the choice of a new coordinate system by rotating the original system with $Z_i(\mathbf{u})$ as the coordinate axes (Johnson & Wichern, 2002, pp. 426-427).

If the variables have widely differing ranges, if they are measured on differing scales or if the units of measurement are not consistent, it is advisable to standardise the variables. A principal component analysis performed on the covariance matrix can be severely affected by large or inconsistent variances. Hence by standardising the original data (or equivalently using the correlation matrix as opposed to the covariance matrix) we ensure that the assigning of the weights in a principal component analysis is not influenced by variables with large variances. It is

important to note that all interpretations of a principal component analysis based on the correlation matrix must be in terms of the standardised variables.

Let **R** be the correlation matrix of the N_v random variables $Z_i(\mathbf{u})$ for $i = 1, ..., N_v$. The spectral decomposition of the N_v x N_v symmetric matrix **R** is given by

$$\mathbf{R} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + ... + \lambda_{N_v} \mathbf{e}_{N_v} \mathbf{e}_{N_v}^T \quad (1)$$

where $\lambda_1, \lambda_2, ..., \lambda_{N_v}$ are the eigenvalues of **R** and $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{N_v}$ are the associated normalised eigenvectors with $\mathbf{e}_i^T \mathbf{e}_i = 1$, $i = 1, ..., N_v$ and $\mathbf{e}_i^T \mathbf{e}_j = 0$ when $i \neq j$. In matrix notation the spectral decomposition of **R** then is

$$\mathbf{R} = \mathbf{Q} \Lambda \mathbf{Q}^T \quad (2)$$

where **Q** is the orthogonal matrix whose columns are the corresponding unit eigenvectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{N_v}$ (note $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) and Λ is the diagonal matrix $[\lambda_k]$ of eigenvalues of **R** for $k = 1, ..., N_v$ such that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{N_v}$. This orthogonal change of variables does not change the total variance of the data. Furthermore the eigenvalues determined in the spectral decomposition are the variances of the principal components.

The first principal component then is the eigenvector corresponding to the largest eigenvalue of **R**, the second principal component is the eigenvector corresponding to the second largest eigenvalue of **R** and so on. The kth principal component is given by a linear combination of the set of N_v original variables:

$$Y_k(\mathbf{u}) = \mathbf{e}_k^T \mathbf{Z}(\mathbf{u}) = q_{1k} Z_1(\mathbf{u}) + q_{2k} Z_2(\mathbf{u}) + ... + q_{N_v k} Z_{N_v}(\mathbf{u}) \text{ for } k = 1, ..., N_v.$$

The set of all N_v principal components are linear combinations of the set of N_v original variables:

$$Y_k(\mathbf{u}) = \sum_{i=1}^{N_v} q_{ik} Z_i(\mathbf{u}) \quad (3)$$

for $k = 1, ..., N_v$. In matrix notation, $\mathbf{Y}(\mathbf{u}) = \mathbf{Q}^T \mathbf{Z}(\mathbf{u})$ and has variance matrix $\mathbf{Q}^T \mathbf{R} \mathbf{Q} = \Lambda$. The total variance of $Y_k(\mathbf{u})$, $k = 1, ..., N_v$, is equal to the total variance of $Z_i(\mathbf{u})$, $i = 1, ..., N_v$, and is given by $\text{tr}(\mathbf{R}) = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + ... + \lambda_{N_v} = N_v$. Hence the variance of $Y_k(\mathbf{u})$ is λ_k and the fraction of the total variance that is explained by $Y_k(\mathbf{u})$ is measured by $\lambda_k / \text{tr}(\mathbf{R})$.

The set of N_v principal component scores is computed at each datum location $\mathbf{u}_\alpha$ in $\mathcal{A}$ for $\alpha = 1, ..., n$ as linear combinations of the standardised sample values $z_i(\mathbf{u}_\alpha)$ at that location multiplied by the loading of variable $Z_i(\mathbf{u})$ or the kth principal component $Y_k(\mathbf{u})$

$$y_k(\mathbf{u}_\alpha) = \sum_{i=1}^{N_v} q_{ki} \left(\frac{z_i(\mathbf{u}_\alpha) - m_i}{\sigma_i} \right) \tag{4}$$

for $k = 1, ..., N_v$ with m_i and σ_i being the mean and standard deviation respectively of the $z_i(\mathbf{u})$ data and q_{ki} obtained from the matrix of eigenvectors $\mathbf{Q}$ in equation (2).

2.2 Multivariate Random Function Model

One of the fundamental aims of geostatistics is to characterise the behaviour of the population of a sampled attribute over a study region $\mathcal{A}$. In order to do this we are required to model the statistical characteristics of the population using only the available sample data. In essence geostatistics uses a probabilistic approach to model the uncertainty about how the attribute behaves between the sample locations.

Consider a set of n sample data values denoted by $z(\mathbf{u}_\alpha)$, where $\alpha=1, ..., n$. These observations may be considered as a subset of a larger, possibly infinite, collection of observations. The value $z(\mathbf{u}_\alpha)$ can be thought of as one possible realisation of a random variable $Z(\mathbf{u}_\alpha)$. Similarly the value $z(\mathbf{u})$ at an unsampled location can be thought of as a particular realisation of the random variable $Z(\mathbf{u})$ for each location $\mathbf{u}$ in the region $\mathcal{A}$. The characterisation of a random variable $Z(\mathbf{u})$ is determined completely by the cumulative distribution function, that is

$$F(\mathbf{u};\ z) = \text{Prob}\{Z(\mathbf{u}) \leq z\} \tag{5}$$

for all z.

The set of random variables $Z(\mathbf{u})$, denoted by $\mathcal{Z}$, for all locations $\mathbf{u}$ in the region $\mathcal{A}$, $\{ Z(\mathbf{u}), \mathbf{u} \in \mathcal{A}\}$, is called a random function and is characterised by the set of all its N-variate cumulative distribution functions

$$F(\mathbf{u}_1, ..., \mathbf{u}_N; z_1, ..., z_N) = \text{Prob}\{Z_1(\mathbf{u}) \leq z_1, ..., Z_N(\mathbf{u}) \leq z_N)\} \tag{6}$$

for any locations $\mathbf{u}_k$, where k=1, ..., N. In general it is impossible to infer the entire spatial law of a random function so it is necessary to obtain an approximate solution to most of the problems encountered. In the most commonly used geostatistical procedures we assume that the random function is stationary, that is, the characteristics of the random function remain invariant under translation. A random function is said to be *strictly stationary* if for any set of N points $\mathbf{u}_1$, ..., $\mathbf{u}_N$ and for any vector $\mathbf{h}$, the two vectors of random variables $[Z(\mathbf{u}_1), ..., Z(\mathbf{u}_N)]$ and $[Z(\mathbf{u}_1 + \mathbf{h}), ..., Z(\mathbf{u}_N + \mathbf{h})]$ have the same multivariate cumulative distribution function

$$F(\mathbf{u}_1,...,\mathbf{u}_N; z_1,...,z_N) = F(\mathbf{u}_1+\mathbf{h},...,\mathbf{u}_N+\mathbf{h}; z_1,...,z_N) \tag{7}$$

for all locations $\mathbf{u}_1$, ..., $\mathbf{u}_N$ and any vector $\mathbf{h}$.

As it is not possible to verify strict stationarity from experimental data we usually require only *second-order stationarity* where the first two moments (mean and covariance) are constant (Armstrong, 1998, p. 18). Hence we require, for all locations $\mathbf{u}$, that the mean exists and is constant

$$E\{Z(\mathbf{u})\} = m \tag{8}$$

and the two-point covariance exists and depends only on the separation vector $\mathbf{h}$

$$C(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})\} - E\{Z(\mathbf{u})\} \cdot E\{Z(\mathbf{u}+\mathbf{h})\} \tag{9}$$

In many cases this assumption is not appropriate so we assume *intrinsic stationarity* where the increments $Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u})$ are assumed to be second-order stationary.

$$E\{Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})\} = 0 \tag{10}$$

$$\mathrm{Var}\{Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})\} = E\{[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2\} = 2\gamma(\mathbf{h}) \tag{11}$$

The function $\gamma(\mathbf{h})$ is called the semivariogram and is the basic tool for the interpretation of the spatial variability of the attribute being investigated.

The probabilistic approach to a coregionalisation (a regionalised phenomenon that can be represented by several intercorrelated variables) is similar to that of the requirements of a single variable whose concepts can be easily broadened to incorporate a multivariate random function. The vector of unsampled values of N_v variables $[z_1(\mathbf{u}), ..., z_{N_v}(\mathbf{u})]$ can be considered as a particular realisation of the vector of N_v random variables $[Z_1(\mathbf{u}), ..., Z_{N_v}(\mathbf{u})]$ for all locations $\mathbf{u}$ over a region $\mathcal{A}$. Similarly this vector of random variables may be thought of as one particular realisation of a multivariate random variable valued function:

$\{[Z_1(\mathbf{u}), ..., Z_{N_v}(\mathbf{u})]; \mathbf{u} \in \mathcal{A}\}$ denoted by the vector $\mathcal{Z}$.

In order to define a cross covariance function that depends only on the separation vector $\mathbf{h}$ the direct (auto) and cross (joint) covariance functions $C_{ij}(\mathbf{h})$ of a set of N_v continuous random functions are defined in the framework of *joint second-order stationarity*. That is, for all locations $\mathbf{u}$ over the study region $\mathcal{A}$ the mean of each variable $Z_i(\mathbf{u})$ exists and is constant and the covariance of a variable pair $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$exists and is translation invariant

$$E[Z_i(\mathbf{u})] = m_i \tag{12}$$

$$C_{ij}(\mathbf{h}) = E[(Z_i(\mathbf{u}) - m_i) \cdot (Z_j(\mathbf{u}+\mathbf{h}) - m_i)] \tag{13}$$

for all $i, j = 1, ..., N_v$. The mean–value vector of a random function $\mathcal{Z}$ is defined as:

$$\mathbf{m}(\mathbf{u}) = E[\mathbf{Z}(\mathbf{u})] \tag{14}$$

Hence the cross covariance functions $C_{ij}(\mathbf{h})$ may be written as the covariance matrix $\mathbf{C}(\mathbf{h}) = \mathrm{E}\{[\mathbf{Z}(\mathbf{u}) - \mathbf{m}] \cdot [\mathbf{Z}(\mathbf{u}+\mathbf{h}) - \mathbf{m}]^{\mathrm{T}}\}$, that is:

$$\mathbf{C}(\mathbf{h}) = \begin{bmatrix} C_{11}(\mathbf{h}) & \cdots & C_{1N_v}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ C_{N_v1}(\mathbf{h}) & \cdots & C_{N_vN_v}(\mathbf{h}) \end{bmatrix} \tag{15}$$

As with the univariate case it may be that in practice the covariance function between any two locations $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$ does not exist in which case it is necessary to weaken the joint second-order stationarity hypothesis to the joint intrinsic hypothesis. The assumption here is that there is only weak stationarity of the first two moments (mean and variance) of the difference of a pair of values located at $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$, that is

$$E\left[Z_i(\mathbf{u}+\mathbf{h}) - Z_i(\mathbf{h})\right] = 0 \tag{16}$$

$$Cov\left[Z_i(\mathbf{u}+\mathbf{h}) - Z_i(\mathbf{h}), Z_j(\mathbf{u}+\mathbf{h}) - Z_j(\mathbf{h})\right] = 2\gamma_{ij}(\mathbf{h}) \tag{17}$$

for all $i, j = 1, ..., N_v$.

Hence the cross semivariogram functions $\gamma_{ij}(\mathbf{h})$ may be written as the semivariogram matrix $\boldsymbol{\Gamma}(\mathbf{h}) = \frac{1}{2}\mathrm{E}\{[\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}+\mathbf{h})] \cdot [\mathbf{Z}(\mathbf{u}) - \mathbf{Z}(\mathbf{u}+\mathbf{h})]^{\mathrm{T}}\}$, that is:

$$\Gamma(\mathbf{h}) = \begin{bmatrix} \gamma_{11}(\mathbf{h}) & \cdots & \gamma_{1N_v}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ \gamma_{N_v1}(\mathbf{h}) & \cdots & \gamma_{N_vN_v}(\mathbf{h}) \end{bmatrix} \tag{18}$$

It is important to note that while the cross variogram is symmetric in (**h**, -**h**), this is not necessarily so for the cross covariance function, that is $C_{ij}(\mathbf{h}) \neq C_{ij}(-\mathbf{h})$ and $C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$. Goovaerts (1997, p.73) states however that in practice this assumption in ignored as the usual tools for description of the spatial variability are symmetric and the verification of the presence of a lag effect is generally not possible as a result of insufficient data.

2.3 Modelling the Coregionalisation

The regionalised variable possesses a local, random, erratic aspect which accounts for local irregularities as well as a general structured aspect which reflects large scale tendencies (Armstrong, 1998, p. 15). Both of these aspects need to be taken into account in the process of developing a representation of the spatial variability of the regionalised variable. Our goal when modelling the semivariogram or covariance is to obtain a suitable interpretation of the spatial structure that characterises the association and causal relationships and main features of the coregionalisation. Our need for a model of the coregionalisation arises from the fact that it is likely that for estimation purposes we will require a semivariogram or covariance value for some distance and/or direction for which we do not have a value (Isaaks & Srivastava, 1989, p. 371). Inference of the semivariogram or covariance model provides a set of functions that allow us to compute semivariogram or covariance values for any possible separation vector **h**.

Only certain functions may be used to model the cross covariance and cross semivariogram. Let $Z_i(\mathbf{u})$ for $i = 1, ..., N_v$ be a set of intercorrelated random functions, $\mathbf{u}_\alpha$ for $\alpha = 1, ..., n$ be a set of n data locations and Y be a finite linear combination of the random variables $Z_i(\mathbf{u}_\alpha)$, where $\mathbf{u}_\alpha$ is a sample location in the study region $\mathcal{A}$ and $i = 1, ..., N_v$. The variance of Y must be non-negative and can be written as the linear combination of cross covariance values

$$\mathrm{Var}\{Y\} = \sum_{i=1}^{N_v}\sum_{j=1}^{N_v}\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_{\alpha i}\lambda_{\beta j}C_{ij}(\mathbf{u}_\alpha - \mathbf{u}_\beta) \geq 0 \tag{19}$$

where $C_{ij}(\mathbf{h})$ denotes the cross covariance at a lag distance $\mathbf{h}$. The variance in terms of the matrix $\mathbf{C}(\mathbf{h})$ is

$$\mathrm{Var}\{Y\} = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha^{\mathrm{T}}\mathbf{C}(\mathbf{u}_\alpha - \mathbf{u}_\beta)\lambda_\beta \geq 0 \tag{20}$$

where $\lambda_\alpha = [\lambda_{\alpha 1},\ldots,\lambda_{\alpha N_v}]^{\mathrm{T}}$ and $\lambda_\alpha^{\mathrm{T}}$ denotes the transpose of the vector λ_α. Thus the matrix of covariances $\mathbf{C}(\mathbf{h})$ must be positive semi-definite to ensure that the variances of Y are non-negative.

Using the relation $\mathbf{C}(\mathbf{h}) = \mathbf{C}(0) - \boldsymbol{\Gamma}(\mathbf{h})$, the variance in terms of the matrix $\boldsymbol{\Gamma}(\mathbf{h})$ is:

$$\mathrm{Var}\{Y\} = \mathbf{C}(0)\sum_{\alpha=1}^{n}\lambda_\alpha^{\mathrm{T}}\sum_{\beta=1}^{n}\lambda_\beta - \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha^{\mathrm{T}}\boldsymbol{\Gamma}(\mathbf{h})\lambda_\beta \geq 0 \tag{21}$$

When a semivariogram is unbounded and has no covariance counterpart the variance of Y is defined on a condition that the vectors of weights λ_α sum to the null vector.

$$\mathrm{Var}\{Y\} = -\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha^{\mathrm{T}}\boldsymbol{\Gamma}(\mathbf{h})\lambda_\beta \geq 0 \text{ with } \sum_{\alpha=1}^{n}\lambda_\alpha = 0 \tag{22}$$

Thus the matrix of semivariograms $\boldsymbol{\Gamma}(\mathbf{h})$ must be conditionally negative semi-definite to ensure that the variances of Y are non-negative.

Recognising whether a function is positive definite or conditionally negative definite is not easy, nor is it simple to test for positive definiteness. Hence it is common practice to model a coregionalisation by using only a few basic structures that are known to be admissible. The following list is not exhaustive but includes the most commonly used admissible models in their standarised form.

- Nugget effect model

$$g(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1 & \text{otherwise} \end{cases} \tag{23}$$

- Spherical model with range a

$$g(h) = \mathrm{Sph}\left(\frac{h}{a}\right) = \begin{cases} 1.5\cdot\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3 & \text{if } h \leq a \\ 1 & \text{otherwise} \end{cases} \tag{24}$$

- Exponential model with practical range a

$$g(h) = 1 - \exp\left(\frac{-3h}{a}\right) \tag{25}$$

- Gaussian model with practical range a

$$g(h) = 1 - \exp\left(\frac{-3h^2}{a^2}\right) \tag{26}$$

- Power model

$$g(h) = h^{\omega} \text{ with } 0 < \omega < 2 \tag{27}$$

These models are considered to be the 'basic models' and are expressed in their isotropic form, that is they are independent of direction ($h = |\mathbf{h}|$). These basic models can be modified to incorporate anisotropy if required.

The nugget effect model is a bounded transition model and is characterised by its discontinuous behaviour at the origin. The spherical and exponential models are also bounded transition models whose behaviour at small separation distances near the origin is linear. The spherical model reaches its sill at the distance a, also known as the actual range. The exponential model reaches its sill asymptotically with a practical range a, the distance at which the semi-variogram value is 95% of the sill. The Gaussian model is a bounded transition model with quadratic behaviour at the origin and also reaches the sill asymptotically with practical range a, the distance at which the semi-variogram value is 95% of the sill (Isaaks & Srivastava, 1989, pp. 373-375). Finally the power model is an unbounded model and has no covariance counterpart. The behaviour of the power model at the origin is dependent on the value of the parameter ω, that is, linear when $\omega = 1$ and parabolic as ω approaches two.

The linear model of coregionalisation is a mathematical model that characterises the spatial variation of a multivariate system at different spatial scales. The requirement of a linear model of coregionalisation is that all direct and cross semivariograms or covariances are jointly modelled and share a common set of basic structures. The linear model of coregionalisation consists of a set of intercorrelated random functions $\mathcal{Z}_i$ so that their corresponding semivariogram matrix $\boldsymbol{\Gamma}$ or covariance matrix $\mathbf{C}$ is by construction admissible. This means that each random

function $\mathcal{Z}_i$ is a linear combination of (L+1) spatially uncorrelated components Y_k^l, for l = 0, ..., L. Each of these components acts at a particular characteristic spatial scale and has a covariance function c_l and zero mean (Journel & Huijbregts, 1978, p. 172).

$$Z_i(\mathbf{u}) = \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{ik}^l Y_k^l(\mathbf{u}) + m_i \tag{28}$$

for all i = 1, ..., N_v with

- $\mathrm{E}\{Z_i(\mathbf{u})\} = m_i$ (a)
- $\mathrm{E}\{ Y_k^l(\mathbf{u})\} = 0$ for all k = 1, ..., n_l and l = 0, ..., L (b)
- $\mathrm{Cov}\{ Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u}+\mathbf{h})\} = \begin{cases} c_l(\mathbf{h}) & \text{if } k = k' \text{ and } l = l' \\ 0 & \text{otherwise} \end{cases}$ (c)

If follows then that the cross covariance function associated with the spatial components $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u} + \mathbf{h})$ can be written as the linear combination of the cross covariance between any two random variables $Y_k^l(\mathbf{u})$ and $Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})$.

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^{L} \sum_{l'=0}^{L} \sum_{k=1}^{n_l} \sum_{k'=1}^{n_{l'}} a_{ik}^l a_{jk'}^{l'} \mathrm{Cov}\{Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u}+\mathbf{h})\} \tag{29}$$

As the random variables $Y_k^l(\mathbf{u})$ in equation (28c) are uncorrelated (orthogonal) except when $l = l'$ and $k = k'$ simultaneously, equation (29) reduces to a linear combination of (L+1) basic covariance models $c_l(\mathbf{h})$. Hence we define the linear model of coregionalisation as the set of N_v x N_v direct and cross covariance models $C_{ij}(\mathbf{h})$

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^{L} b_{ij}^l c_l(\mathbf{h}) \tag{30}$$

for all i, j = 1, ..., N_v, where the sill of the basic covariance model $c_l(\mathbf{h})$ is matrix-valued and defined by

$$b_{ij}^k = \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l \tag{31}$$

for all i, j = 1, ..., N_v and l = 0, ..., L. Similarly we can define the linear model of coregionalisation in terms of the set of N_v x N_v direct and cross semivariogram models $\gamma_{ij}(\mathbf{h})$ such that:

$$\gamma_{ij}(\mathbf{h}) = \sum_{l=0}^{L} b_{ij}^{l} g_l(\mathbf{h}) \tag{32}$$

The coregionalisation matrix of size N_v x N_v , $\mathbf{B}_l = \left[b_{ij}^{l}\right]$, is by construction a positive semi-definite matrix and is the variance-covariance matrix which describes the multivariate correlation at each of the characteristic scales l for $l = 0, ..., L$ (Wackernagel, 1998b, p. 27). Thus for a linear model of coregionalisation to be admissible it must satisfy two conditions:

i) the functions $c_l(\mathbf{h})$ ($g_l(\mathbf{h})$) are admissible covariance (semivariogram) models and

ii) the $(L + 1)$ coregionalisation matrices $\mathbf{B}_l$ are positive semi-definite.

This second condition is readily checked by confirming that the eigenvalues of each of the $(L+1)$ coregionalisation matrices $\mathbf{B}_l$ is real and non-negative.

The multivariate nested covariance function model $C_{ij}(\mathbf{h})$ with positive semi-definite coregionalisation matrices $\mathbf{B}_l$ expressed in matrix notation is

$$\mathbf{C}(\mathbf{h}) = \sum_{l=0}^{L} \mathbf{B}_l c_l(\mathbf{h}) \tag{33}$$

with $\mathbf{B}_l = \mathbf{A}_l \mathbf{A}_l^T$ and $\mathbf{A}_l = \left[a_{ij}^{l}\right]$. Correspondingly, the multivariate nested model associated with a linear model of coregionalisation of intrinsically stationary random functions expressed in matrix notation is

$$\boldsymbol{\Gamma}(\mathbf{h}) = \sum_{l=0}^{L} \mathbf{B}_l g_l(\mathbf{h}) \tag{34}$$

where $\mathbf{B}_l$ are positive semi-definite matrices and $g_l(\mathbf{h})$ are the semivariogram models.

If the multivariate correlation structure of a set of N_v variables is independent of the spatial correlation the multivariate correlation is said to be intrinsic. According to Chiles & Delfiner (1999, p. 337) Matheron introduced the intrinsic coregionalisation model in order to validate the use of the correlation coefficient from a geostatistical perspective. The problem arises with the fact that variance (or covariance) of spatially correlated data within a finite domain $\mathcal{A}$, depends on $\mathcal{A}$ (Chiles & Delfiner, 1999, p. 337). The intrinsic coregionalisation model may only be implemented when the correlation r_{ij} between the random functions $\mathcal{Z}_i$ and $\mathcal{Z}_j$ does not depend on spatial scale, that is:

$$r_{ij} = \frac{b_{ij}\gamma(\mathbf{h})}{\sqrt{b_{ii}b_{jj}}\gamma(\mathbf{h})} = \frac{b_{ij}}{\sqrt{b_{ii}b_{jj}}} \tag{35}$$

The intrinsic coregionalisation model is a special case of the linear model of coregionalisation where the coefficients (sills) of any basic structure $c_l(\mathbf{h})$ or $g_l(\mathbf{h})$ that constitute the model are proportional to each other. That is, $b_{ij}^l = \varphi_{ij} \cdot b^l$ for $i, j = 1, \ldots$ and $l = 0, \ldots, L$. The intrinsic coregionalisation model is the simplest multivariate model used in geostatistics as all direct and cross covariance and semivariogram models are proportional to the basic standardised covariance, $C(\mathbf{h}) = \sum_{l=0}^{L} b^l c_l(\mathbf{h})$, or semivariogram, $\gamma(\mathbf{h}) = \sum_{l=0}^{L} b^l g_l(\mathbf{h})$, functions

$$C_{ij}(\mathbf{h}) = \varphi_{ij} C(\mathbf{h}) \tag{36}$$

$$\gamma_{ij}(\mathbf{h}) = \varphi_{ij} \gamma(\mathbf{h}) \tag{37}$$

for all $i, j = 1, \ldots, N_v$ and $\sum_{l=0}^{L} b^l = 1$. In matrix notation the intrinsic coregionalisation model is written

$$\boldsymbol{\Gamma}(\mathbf{h}) = \boldsymbol{\Phi}\boldsymbol{\gamma}(\mathbf{h}) \tag{38}$$

$$\mathbf{C}(\mathbf{h}) = \boldsymbol{\Phi} C(\mathbf{h}) \tag{39}$$

where the matrix of coefficients $\boldsymbol{\Phi} = [\varphi_{ij}]$ is equal to the variance-covariance matrix under the assumption of second-order stationarity (Wackernagel, 1998b, p. 10).

While the intrinsic coregionalisation model is more restrictive than the linear model of coregionalisation it is of particular benefit as this model reduces the inference of $N_v (N_v + 1)/2$ covariance or semivariogram functions to the inference of only one covariance or semivariogram model and $N_v (N_v + 1)/2$ coefficients. One way of detecting whether variables are intrinsically correlated is by examining the codispersion between the variables. This can be done by checking graphically whether the codispersion coefficients

$$cc_{ij}(\mathbf{h}) = \frac{\gamma_{ij}(\mathbf{h})}{\sqrt{\gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h})}}$$

are constant and equal to the sample correlation coefficient. If they are the correlation of each pair of variables does not depend on spatial scale. This can

however become time consuming when the number of variables is large. An alternative method is to perform a principal component analysis on the data and compute the cross correlograms of the first few principal components which account for the majority of variability in the original data. If the cross correlograms of the principal components are zero for any separation vector **h** this implies that the orthogonality of the principal components is not dependent on spatial scale. In other words for the data to be intrinsically correlated we require the cross correlograms of the principal components to be zero for any separation vector **h**.

2.4 Kriging

Once we have a mathematical representation of the spatial continuity of the attributes of interest in the form of our random function model we are able to proceed with the estimation of those attributes at unsampled locations across the study region. There are many traditional point estimation techniques available, such as polygonal, Delaunay triangulation, inverse distance squared and moving average methods. The overriding problem with these techniques is that the 'best' one is dependent on one's choice of criteria as to what is 'best'.

In the 1950's South African mining engineer Danie Krige developed a technique of interpolation in an attempt to more accurately predict ore reserves. Based on Krige's work, Georges Matheron developed the Theory of Regionalized Variables in the early 1960's in which he combined Krige's pioneering work into a single framework which was coined "krigeage" in recognition of Krige's contribution to the field (Chiles & Delfiner, 1999, p. 150). Formally, kriging now refers to a family of least-squares linear regression algorithms that share the objective of minimising the estimation (error) variance subject to the constraint of unbiasedness of the estimator (Deutsch & Journel, 1998, p. 14). Over the past several decades kriging has become a fundamental tool in the field of geostatistics as it has established itself as a superior method of estimation.

2.4.1 Simple and Ordinary Kriging

Let $\mathcal{Z}$ be a second-order stationary random function with mean m. The estimator $Z^*(\mathbf{u})$ of $\mathcal{Z}$ is given by a linear combination of random variables $Z(\mathbf{u}_\alpha)$ with weights $\lambda_\alpha(\mathbf{u})$ chosen such that the estimator is unbiased and the estimation variance is minimised. The basic generalised spatial least-squares regression estimator $Z^*(\mathbf{u})$ is defined as

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] \tag{40}$$

where the quantities $m(\mathbf{u})$ and $m(\mathbf{u}_\alpha)$ are the expected values of $Z(\mathbf{u})$ and $Z(\mathbf{u}_\alpha)$ respectively. By making the assumption that the both the sample value $z(\mathbf{u}_\alpha)$ and unknown value $z(\mathbf{u})$ are realisations of the random variables $Z(\mathbf{u}_\alpha)$ and $Z(\mathbf{u})$ respectively we are able to define an estimation error random variable $Z^*(\mathbf{u}) - Z(\mathbf{u})$. In order for the estimator to be unbiased we require that the expected value of the estimation error be zero:

$$E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0 \tag{41}$$

The estimation error variance is then given by

$$\sigma_E^2(\mathbf{u}) = Var\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} \tag{42}$$

and is minimised under the constraint of unbiasedness of the estimator.

If we assume $m(\mathbf{u})$ to be known and constant across the study region $\mathcal{A}$, the linear estimator $Z^*(\mathbf{u})$ is known as the simple kriging (SK) estimator $Z_{SK}^*(\mathbf{u})$ where

$$Z_{SK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u})\right] m \tag{43}$$

and the $\lambda_{SK}(\mathbf{u})$ are determined such that the error variance is minimised. The simple kriging estimator automatically exhibits unbiasedness as the mean error is equal to zero. The simple kriging system then becomes, in terms of the covariance function, a system of normal equations

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{SK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{u}_\alpha - \mathbf{u}) \tag{44}$$

for all $\alpha = 1, \ldots, n(\mathbf{u})$. The simple kriging minimum error variance is given by

$$\sigma^2_{E_{SK}}(\mathbf{u}) = Var\{Z^*_{SK}(\mathbf{u}) - Z(\mathbf{u})\} = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) \tag{45}$$

In most cases however it is not acceptable to assume that the mean is known and globally constant. An alternative approach is ordinary kriging which assumes that the mean is unknown but locally constant. Ordinary kriging then limits stationarity of the mean to the local neighbourhood of the location **u** being estimated. The linear estimator (40) is then modified to incorporate the constant local mean $m(\mathbf{u})$.

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\right] m(\mathbf{u}) \tag{46}$$

In order to filter the mean $m(\mathbf{u})$ from equation (40) we must impose the constraint that the sum of the weights be equal to one. This yields the ordinary kriging estimator $Z^*_{OK}(\mathbf{u})$ which again must be solved to obtain the optimal kriging weights such that the estimation error is minimised under the unbiasedness constraint.

$$Z^*_{OK}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) Z(\mathbf{u}_\alpha) \quad \text{with} \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) = 1 \tag{47}$$

The ordinary kriging system involves $n(\mathbf{u})$ weights $\lambda_\alpha^{OK}(\mathbf{u})$ and the Lagrange multiplier $\mu_{OK}(\mathbf{u})$, which accounts for the unbiasedness constraint. Expressed in terms of the covariance function this system is

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{OK}(\mathbf{u}) = C(\mathbf{u}_\alpha - \mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases} \tag{48}$$

for all $\alpha = 1, \ldots, n(\mathbf{u})$. Unlike the simple kriging system the ordinary kriging system may also be expressed in terms of the semivariogram as

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_{OK} = \gamma(\mathbf{u}_\alpha - \mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases} \tag{49}$$

for all $\alpha = 1, \ldots, n(\mathbf{u})$. The ordinary kriging minimum error variance is given by

$$\sigma^2_{E_{OK}}(\mathbf{u}) = = Var\left\{Z^*_{OK}(\mathbf{u}) - Z(\mathbf{u})\right\} = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{OK}(\mathbf{u}) \quad (50)$$

2.4.2 Ordinary Cokriging

The linear estimator (40) is readily extended to the multivariate case where we have available N_v continuous random variables $Z_i(\mathbf{u})$. We consider $z_i(\mathbf{u})$ as a realisation of the random variable $Z_i(\mathbf{u})$, $i = 1, ..., N_v$, with $z_i(\mathbf{u}_\alpha)$ being the set of n sample data located at $\mathbf{u}_\alpha$, $\alpha = 1, ..., n$. Let $Z_1(\mathbf{u})$ be the primary random variable of interest with $E\{Z_1(\mathbf{u})\} = m_1(\mathbf{u})$, $E\{Z_i(\mathbf{u}_{\alpha_i})\} = m_i(\mathbf{u}_{\alpha_i})$ and λ_{α_1} and λ_{α_i} being the weights assigned to $z_1(\mathbf{u}_{\alpha_1})$ and $z_i(\mathbf{u}_{\alpha_i})$ respectively. In order to estimate a primary variable with N_v–1 auxiliary variables the linear estimator (40) is extended to incorporate the additional information.

$$Z_1^*(\mathbf{u}) - m_1(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha_1}(\mathbf{u})[Z_1(\mathbf{u}_\alpha) - m_1(\mathbf{u}_{\alpha_1})] + \sum_{i=2}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}(\mathbf{u})[Z_1(\mathbf{u}_{\alpha_i}) - m_i(\mathbf{u}_{\alpha_i})] \quad (51)$$

Analogous to the kriging paradigm the cokriging algorithm generally only retains the data closest to the location $\mathbf{u}$. Again we wish to determine the weights λ_{α_1} and λ_{α_i} such that the estimation variance

$$\sigma^2_E(\mathbf{u}) = Var\left\{Z_1^*(\mathbf{u}) - Z_1(\mathbf{u})\right\} \quad (52)$$

is minimised under the unbiasedness constraint that the expected error is zero.

$$E\left\{Z_1^*(\mathbf{u}) - Z_1(\mathbf{u})\right\} = 0 \quad (53)$$

The three most commonly used types of cokriging are

i) Simple cokriging where the mean $m_i(\mathbf{u})$ is known and constant throughout the study region $\mathcal{A}$

ii) Ordinary cokriging where the mean $m_i(\mathbf{u})$ is unknown but constant throughout the study region $\mathcal{A}$ and

iii) Cokriging with a trend where the mean $m_i(\mathbf{u})$ is unknown and varies as a function of the spatial coordinates $\mathbf{u}$.

Pertinent to this study is ordinary cokriging which limits stationarity of the mean to the local neighbourhood of the location $\mathbf{u}$ being estimated. In order to filter the means $m_1(\mathbf{u})$ and $m_i(\mathbf{u})$, from equation (51) we must impose the constraints that

the sum of the weights of the primary variable be equal to one and the sum of the weights of the secondary variable(s) be equal to zero. This yields the ordinary cokriging estimator

$$Z_{OCK}^{(1)*} = \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{OCK}(\mathbf{u}) Z_i(\mathbf{u}_{\alpha_i}) \tag{54}$$

which again must be solved for the kriging weights such that the estimation error is minimised subject to the unbiasedness constraints for $i = 2, ..., N_v$.

$$\sum_{\alpha_1=1}^{n_1(\mathbf{u})} \lambda_{\alpha_1}^{OCK}(\mathbf{u}) = 1$$

$$\sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{OCK}(\mathbf{u}) = 0$$

The ordinary cokriging system can be expressed in terms of the direct and cross covariances as

$$\begin{cases} \sum_{j=1}^{N_v} \sum_{\beta_j=1}^{n_j(\mathbf{u})} \lambda_{\beta_j}^{OCK}(\mathbf{u}) C_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) + \mu_i^{OCK}(\mathbf{u}) = C_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}) \\ \sum_{\beta_i=1}^{n_i(\mathbf{u})} \lambda_{\beta_i}^{OCK}(\mathbf{u}) = \delta_{ij} \quad \delta_{ij} = \begin{cases} 1 \text{ if i = 1} \\ 0 \text{ otherwise} \end{cases} \end{cases} \tag{55}$$

for $\alpha_i = 1, ..., n_i(\mathbf{u})$ and $i = 1, ..., N_v$. As for the ordinary kriging case the ordinary cokriging system can be expressed in terms of the direct and cross semivariograms as

$$\begin{cases} \sum_{j=1}^{N_v} \sum_{\beta_j=1}^{n_j(\mathbf{u})} \lambda_{\beta_j}^{OCK}(\mathbf{u}) \gamma_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) - \mu_i^{OCK}(\mathbf{u}) = \gamma_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}) \\ \sum_{\beta_i=1}^{n_i(\mathbf{u})} \lambda_{\beta_i}^{OCK} - \delta_{i1} \end{cases} \tag{56}$$

for $\alpha_i = 1, ..., n_i(\mathbf{u})$ and $i = 1, ..., N_v$. The cokriging variance is given by

$$\begin{aligned} \sigma_{E_{OCK}}^2(\mathbf{u}) &= Var\left\{Z_{OCK}^{(1)*}(\mathbf{u}) - Z_1(\mathbf{u})\right\} \\ &= C_{11}(0) - \mu_1^{OCK}(\mathbf{u}) - \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i(\mathbf{u})} \lambda_{\alpha_i}^{OCK}(\mathbf{u}) C_{i1}(\mathbf{u}_{\alpha_i} - \mathbf{u}) \end{aligned} \tag{57}$$

As the cokriging systems (55) and (56) can become unstable if the variances of the primary and secondary variables differ by several orders of magnitude it is advisable to use standardised variables if this is likely to be a concern.

In general the benefit of incorporating secondary information is fully exploited when the primary variable(s) of interest is undersampled. In the isotopic or

equally sampled case the estimates obtained from ordinary cokriging are likely to be similar to those obtained by ordinary kriging. However one advantage of cokriging a set of equally sampled variables is that we preserve the coherence of the estimators. That is, the cokriging of a sum of variables is equal to the sum of the cokrigings of each of the variables. Another advantage of cokriging in both the isotopic and heterotopic cases is that the estimation variance is less than or equal to that of the kriging estimator. In the particular case of the intrinsic coregionalisation model the ordinary cokriging estimates will be equivalent to those obtained by ordinary kriging.

2.4.3 Principal Component Kriging

As we have discussed previously, a principal component analysis transforms a set of correlated variables into a set of components that are uncorrelated at $|\mathbf{h}| = 0$. The advantage of principal component kriging is that it reduces the estimation problem of cokriging N_v variables to the kriging of N_v principal components. The overriding assumption of principal component kriging is that the principal components are mutually orthogonal for any separation vector $\mathbf{h}$

$$Cov\{Y_k(\mathbf{u}), Y_{k'}(\mathbf{u}+\mathbf{h})\} = C_{kk'}(\mathbf{h}) = 0 \tag{58}$$

for all $k \neq k'$. If the data are intrinsically correlated condition (48) is then satisfied. This means that there is no benefit in incorporating secondary information as the kriging and cokriging estimates will be identical. Hence the principal components can be kriged independently. One of the drawbacks however of principal component kriging is that the data must be isotopic, that is only those data that are jointly measured can be considered.

Having determined the N_v principal components we calculate the principal component scores $y_k(\mathbf{u}_\alpha)$ as per equation (4) and model the N_v semivariograms $\gamma_{kk}(\mathbf{h})$ from these scores. We then estimate the principal components separately at each unsampled location $\mathbf{u}$ in $\mathcal{A}$. As the mean of each principal component is zero, the ordinary kriging estimator of the kth principal component at location $\mathbf{u}$ is of the form

$$Y_{OK}^{(k)*}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha k}^{OK}(\mathbf{u}) Y_k(\mathbf{u}_\alpha) \tag{59}$$

where the kriging weights are obtained from the ordinary kriging system as displayed in expression (48). The estimate of $z_i(\mathbf{u})$ is then recalculated as a linear combination of the principal component estimates at each location plus the mean m_i of each attribute.

$$z_{PCK}^{(i)*}(\mathbf{u}) = \sum_{k=1}^{K} a_{ki} y_{OK}^{(k)*}(\mathbf{u})\sigma_i + m_i \tag{60}$$

The coefficients a_{ki} are obtained from the matrix $\mathbf{A} = [a_{ki}] = \mathbf{Q}^{-1} = \mathbf{Q}^{T}$ where $\mathbf{Q}$ is the orthogonal matrix of eigenvectors calculated in the principal component analysis.

Should sufficient variability of the original data be explained by P principal components (where P is ideally significantly less than N_v), it is possible to retain only these and yet simultaneously estimate all N_v original variables without the loss of too much information. In this case the estimates are obtained by modifying equation (60) as follows:

$$z_{PCKP}^{(i)*} = \sum_{p=1}^{P} a_{pi} y_{OK}^{(p)*}(\mathbf{u})\sigma_i + m \tag{61}$$

3 DATA ANALYSIS

In this chapter we introduce the MM22D data suite used in this study. We then proceed with an exploratory data analysis of the grade control and exploration data sets contained in the data suite. We next discuss the principal component analysis of the MM22DEXP data set and assess whether the data are intrinsically correlated. In addition we consider the principal component analysis and subsequent assessment of intrinsic correlation of two four variable subsets of MM22DEXP called MM22DHC4 and MM22DTOP4. The first subset consists of variables that are highly correlated with nickel and cobalt. The second subset consists of variables that are considered to be economically the most important to the mining company.

3.1 MM22D Data Suite

The data to be used in this study come from Anaconda's Murrin Murrin nickel mine near the town of Laverton in Western Australia. The data have been collected from an area within this mine known as MM2. Murphy, Bloom and Mueller (2002) explain that in this region "the laterite deposits are of the dry-climate type and occur as laterally extensive, undulating blankets of mineralisation with strong vertical anisotropy and near normal nickel distributions." The data suite MM22D comprises three dimensional grade and thickness measurements on eight variables: nickel, cobalt, magnesium, iron, aluminium, chromium, zinc and manganese. For the purposes of this study the data have been transformed to accumulations (average grade multiplied by total thickness) hence are treated as two dimensional.

The MM22D data suite consists of two data sets: the grade control data set MM22DGC and the exploration data set MM22DEXP. The data MM22DGC are grade control accumulations for each of the eight variables jointly sampled at 1718 locations over an irregularly shaped study region with grid spacings of 12.5m by 12.5m as displayed in Figure 3.1. This data set is considered reality and will be used to assess the accuracy of the results obtained from the estimation methods. MM22DEXP is a subset of MM22DGC and comprises accumulations for each of the

eight variables jointly sampled at 125 locations across the study region as displayed in Figure 3.2. Ninety-seven of these data are located at grid spacings of 50m by 50m. In addition there is a local cluster of 24 samples at 12.5m spacings around the location at easting 1098.03, northing 298.53. There are also four additional samples along the line of drill holes extending southwards from easting 1302.22, northing 357.59 at 25m spacings. These local clusters are sampled for the purposes of identification of the short range characteristics of the variables (Murphy et al., 2002).

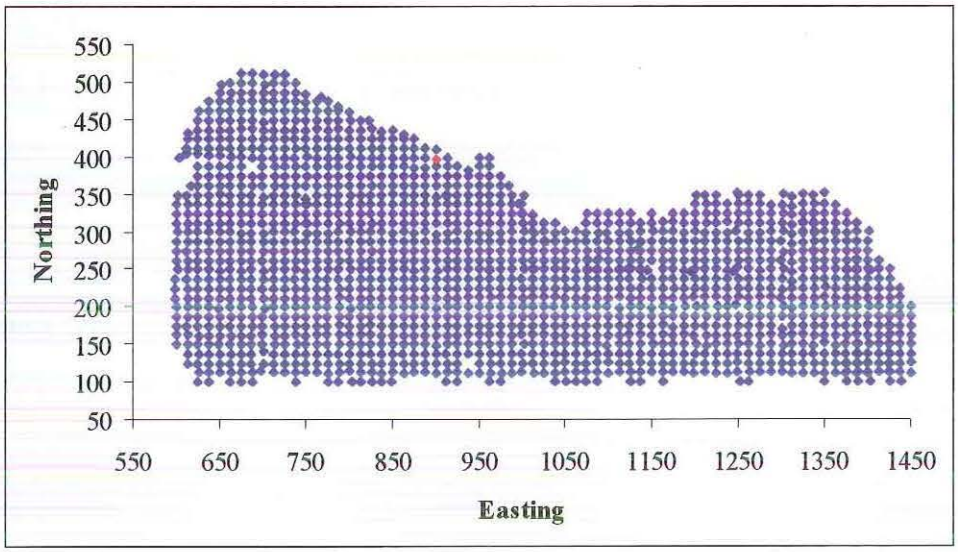


Figure 3.1: MM22DGC grade control sample locations

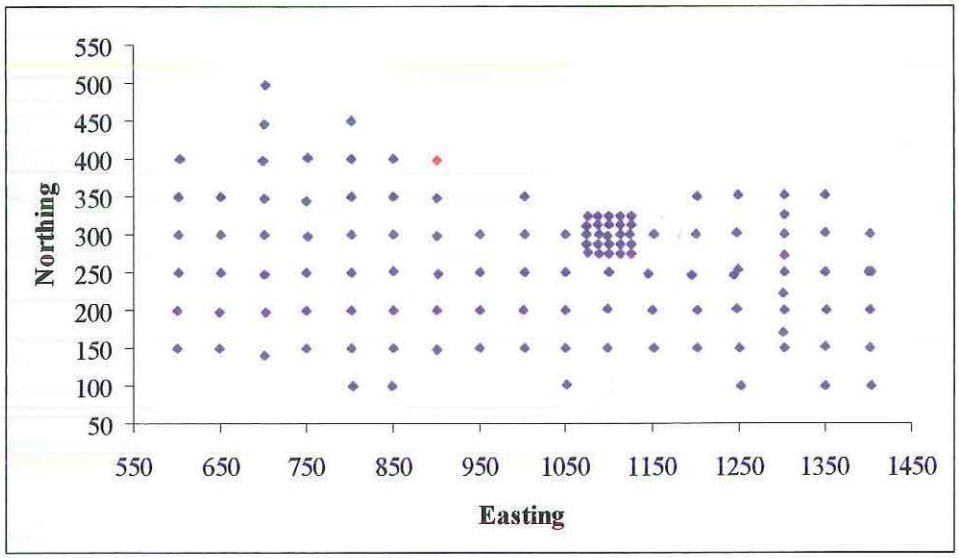


Figure 3.2: MM22DEXP exploration sample locations

3.2 Exploratory Data Analysis

Two types of measurements have been used to assess the level of the attributes. Nickel, cobalt, magnesium, iron and aluminium are measured in parts per million (ppm-metres) and chromium, zinc and manganese are measured in percent-metres (%). The only erroneous data identified in the exploration data set are the samples located at easting 900, northing 398 (denoted by a red marker in Figures 3.1 and 3.2). As the readings for each variable are incomplete, this record has been removed from the both the grade control and exploration data sets. Tables 3.1, 3.2 and 3.3 display the descriptive statistics for the grade control, exploration and declustered exploration data (using cell declustering) where SD and CV are the standard deviation and coefficient of variation respectively.

Overall the exploration data have reproduced the grade control summary statistics well. In most cases summary statistics of the declustered exploration data are very similar to the clustered values with manganese being the only mineral to show any difference. As the differences are minimal the exploration data rather than the declustered data will be utilised in this study.

Table 3.1: Summary statistics for MM22DGC grade control data

	Ni %	Co %	Mg %	Fe %	Al %	Cr ppm	Zn ppm	Mn ppm
n	1717	1717	1717	1717	1717	1717	1717	1717
Mean	13.05	0.82	62.41	313.83	46.33	120053	2312.90	38874
Median	12.98	0.78	57.39	312.19	44.49	118127	2314.20	35411
Min	0.24	0.00	0.20	9.60	1.40	345	86.0	0
Max	42.64	4.90	378.93	813.79	163.90	372240	6831.99	236803
SD	6.64	0.56	52.73	144.53	25.91	67042	1054.3	34943
Skewness	0.20	1.26	1.51	0.08	1.18	0.33	-0.03	1.57
CV	0.51	0.68	0.85	0.46	0.56	0.56	0.46	0.90

Table 3.2: Summary statistics for MM22DEXP exploration data

	Ni %	Co %	Mg %	Fe %	Al %	Cr ppm	Zn ppm	Mn ppm
n	124	124	124	124	124	124	124	124
Mean	14.07	0.85	78.46	318.6	47.99	82471	2130.10	46354
Median	13.37	0.68	58.69	314.9	44.20	81357	1989.0	36279
Min	0.56	0.02	8.07	11.0	1.40	2119	88.0	1149
Max	42.64	2.97	378.93	778.99	146.40	217900	4918.10	236803
SD	6.72	0.55	67.51	144.2	27.13	48238	944.3	43314
Skewness	0.95	1.24	2.06	0.22	1.08	0.42	0.33	1.69
CV	0.48	0.65	0.86	0.45	0.57	0.58	0.44	0.93

Table 3.3: Summary statistics for declustered MM22DEXP exploration data

	Ni %	Co %	Mg %	Fe %	Al %	Cr ppm	Zn ppm	Mn ppm
n	96	96	96	96	96	96	96	96
Mean	15.28	0.91	87.00	354.0	49.82	91059	2330.8	50985
Median	15.09	0.82	68.65	331.3	45.05	91800	2182.2	44391
Min	3.83	0.14	8.07	98.6	8.41	2119	326.0	2220
Max	42.64	2.44	378.93	778.99	146.40	217900	4918.10	200600
SD	6.47	0.50	72.57	128.4	28.06	46718	881.10	39234
Skewness	1.05	0.94	1.85	0.33	1.14	0.41	0.43	1.17
CV	0.42	0.55	0.83	0.36	0.56	0.51	0.38	0.77

Graphical summaries of the grade control data are in displayed in Figure 3.3. Cobalt, aluminium, magnesium and manganese are all strongly positively skewed (skewness coefficient greater than 1) for the grade control data.

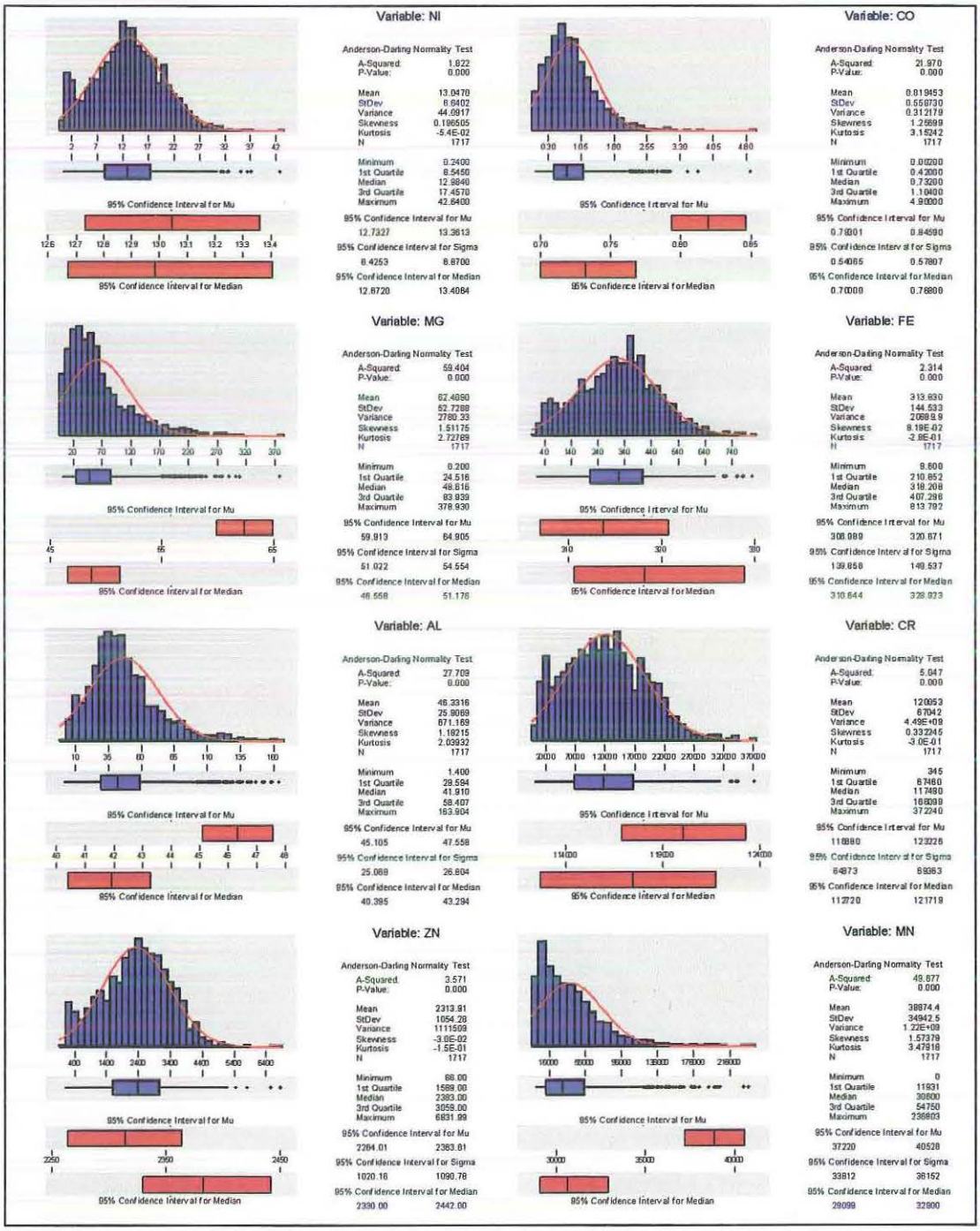


Figure 3.3: Graphical summary of MM22DGC grade control data

The skewness of the variables cobalt, aluminium, magnesium and manganese is also apparent in the exploration data as displayed in Figure 3.4. The exploration data for iron and chromium are normally distributed with nickel near normal.

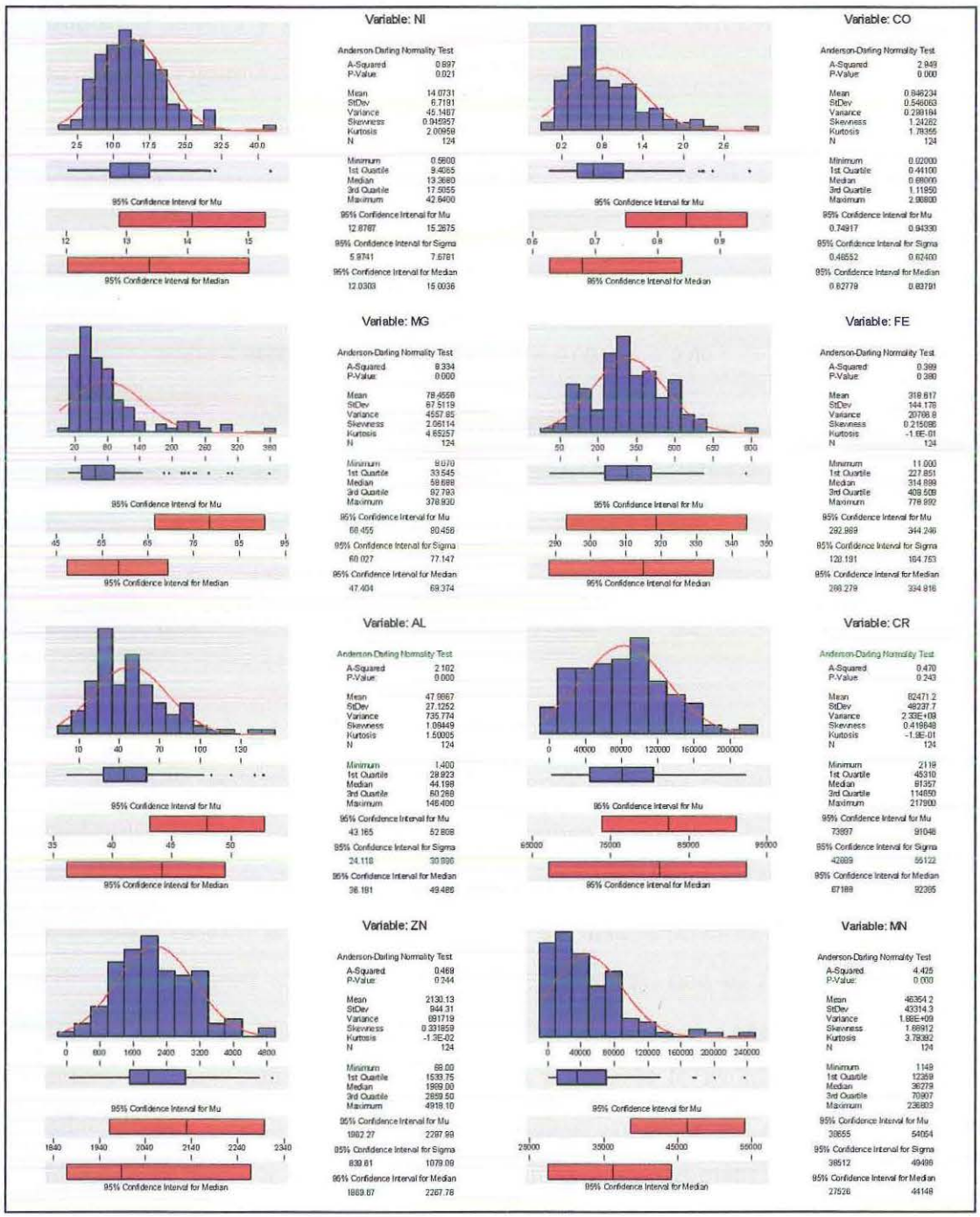


Figure 3.4: Graphical summary of exploration data

From the summary statistics displayed in Tables 3.1 and 3.2 it is evident that the range, mean and variance of the attributes vary greatly. In order to make the attributes comparable the data need to be standardised by subtracting the sample mean and dividing by the sample standard deviation. Table 3.4 displays the standardised summary statistics for the exploration data (with the erroneous data removed) and Figure 3.5 shows the graphical summary.

Table 3.4: Summary statistics for standardised MM22DEXP exploration data

	Ni %	Co %	Mg %	Fe %	Al %	Cr PPM	Zn PPM	Mn PPM
N	124	124	124	124	124	124	124	124
Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Median	-0.10	-0.30	-0.29	-0.03	-0.14	-0.02	-0.15	-0.23
Min	-2.01	-1.51	-1.04	-2.13	-1.72	-1.67	-2.16	-1.04
Max	4.25	3.89	4.45	3.19	3.63	2.81	2.95	4.40
SD	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Skewness	0.95	1.24	2.06	0.22	1.08	0.42	0.33	1.69

In many geostatistical and statistical procedures, we make the assumption of second-order stationarity and normality. Although a principal component analysis does not require normality of the data, the results are enhanced if this condition is met. In many cases it is possible to transform the data in order to remove any trend, stabilise the variance and achieve normality. However, one of the problems with transforming data is that the estimation errors can become exaggerated when the estimates are back transformed. Hence, it is preferable to avoid transformation if possible. Figure 3.6 displays the normal probability plots for each of the eight variables. The plots for cobalt, magnesium, aluminium and manganese show strong deviation from normality in the upper and lower tails. Figures 3.7 and 3.8 show the lognormal and Weibull plots for these variables. While neither transformation appears to have adequately removed the asymmetry of the raw data the lognormal plots seem to have less deviation than the Weibull plots. The decision of whether to

transform the data was further examined when we performed the principal component analysis on the exploration data.

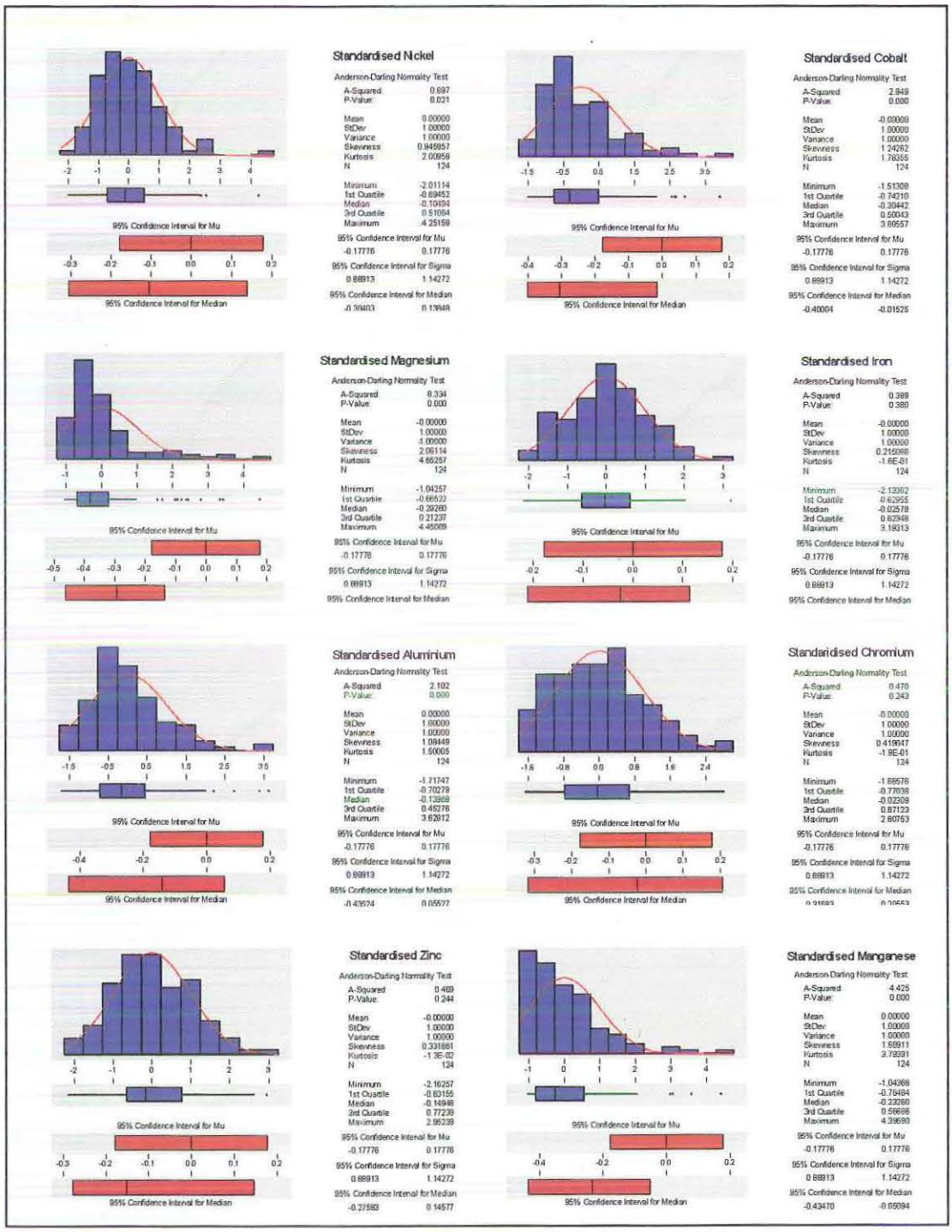


Figure 3.5: Graphical summary for standardised MM22DEXP exploration data

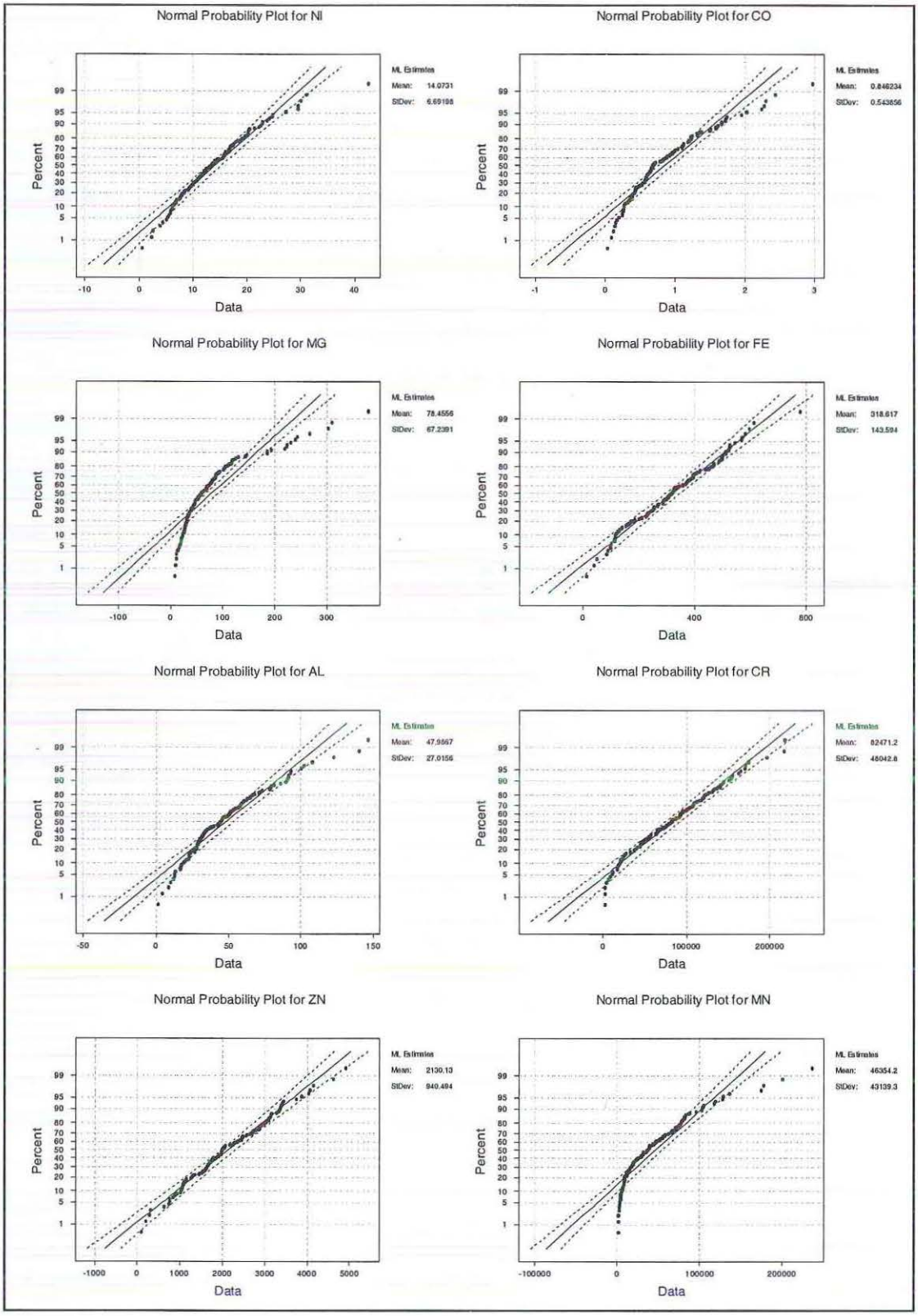


Figure 3.6: Normal probability plots of raw exploration data

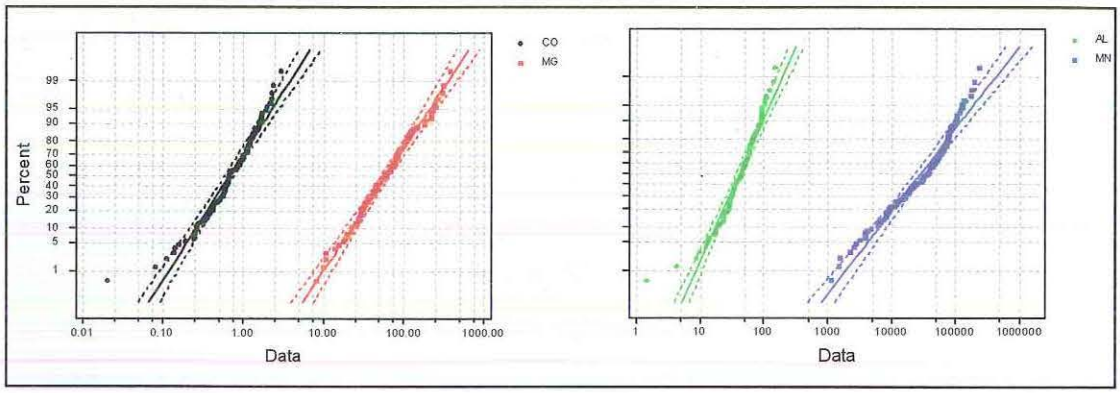


Figure 3.7: Lognormal probability plots for cobalt, magnesium, aluminium and manganese

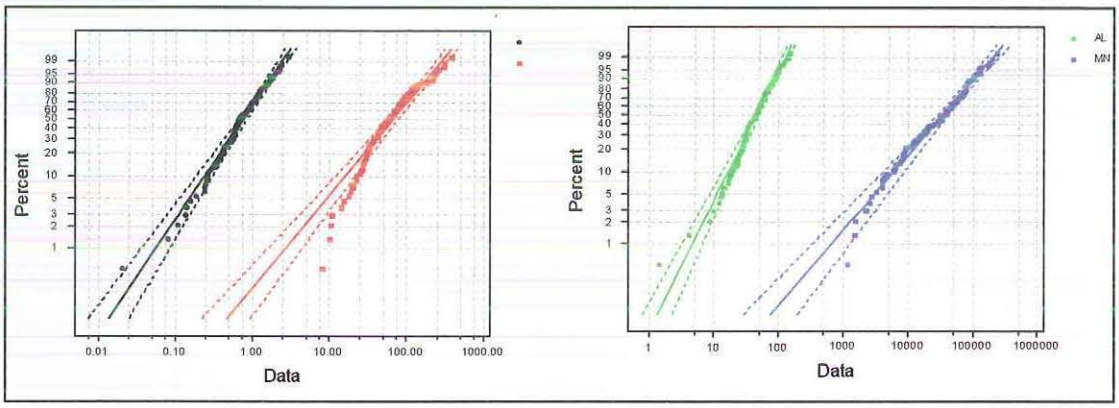


Figure 3.8: Weibull probability plots for cobalt, magnesium, aluminium and manganese

3.3 Spatial Data Analysis

The post plots for each of the eight variables for both grade control and exploration data are displayed in Figures 3.9 and 3.10 respectively. The post plot for the grade control nickel reveals a large concentration of high values across the lower eastern half of the study region. These high values drop to moderately high across the lower western half of the study region. Another smaller concentration of medium to high nickel values is located at the north-western corner. There is a long band of low nickel values along the northern perimeter of the study region and another along the

eastern perimeter. The exploration nickel data reflect the spatial distribution of the grade control nickel very well with only the low values on the eastern perimeter under represented.

The cobalt grade control values seem to be predominately high in small pockets across the southern half of the study region, along the western perimeter and in the north-western corner. The entire south-western corner appears to have medium to high levels of cobalt. There are low values along the northern perimeter of the region as well as in the south-eastern corner. The exploration cobalt values also reflect the spatial distribution of the grade control data very well, although the low values in the south-eastern corner and the high values in the north-western corner are not well represented.

The post plot of the grade control magnesium shows small pockets of high values in the north-western corner, central eastern region, eastern perimeter, south-eastern and the south-western corners. The concentrations of low values are located predominately across the centre of the study area with pockets along the western and northern perimeters. The spatial distribution of the exploration data is also very representative of the grade control data with the only exception being an under representation of low values along the south south-eastern and south south-western perimeters of the study region.

The post plot of the grade control iron indicates that there is a large area of high values across the central southern region extending across toward the south-western corner and up along the western perimeter. There is also a small pocket of high values in the north-western corner. The low values extend in a long band along the northern perimeter of the region with a large area of low values in the south-eastern corner. Overall the spatial distribution of the grade control iron is very well represented by the exploration data, the only exception being an under representation of low values in the south-eastern corner.

The post plot of the grade control aluminium shows two large areas of high concentrations in the south-eastern and south-western regions of the study area. There is also a pocket of high values in the north-western corner. The low values are predominately located across the western side of the study region; in addition there are small pockets of low values in the south-eastern corner. The spatial distribution of the exploration data is also very representative of the grade control data with the

only exception being an under representation of high in the north-western corner of the study region.

The post plot of the grade control chromium shows clear cut areas of high and low concentrations with the high values along the south-eastern perimeter extending across the central southern region into the lower western side. The low values extend across the entire northern boundary extending down along the east south-eastern perimeter. The spatial distribution of the exploration data is in this case not very representative of the grade control data. The large regions of high concentrations appear to be greatly under represented across the entire study region.

The post plot of the grade control zinc shows high concentrations across the entire southern half of the study region and in the north-western corner. The low values are located along the northern and south-eastern perimeters. The spatial distribution of the exploration data is very representative of the grade control data with all high and low values well represented across the entire study region.

The post plot of the grade control manganese shows a large concentration of high values occupying the entire south-western region, extending east across the lower southern half up into the south-eastern corner. There is also a pocket of high values in the north-western corner of the study region. Three regions of low values are evident, on the central north-western side, across the centre of the study area and across the central south-eastern region. Again, the spatial distribution of the exploration data is very representative of the grade control data with all high and low values well represented across the entire study region.

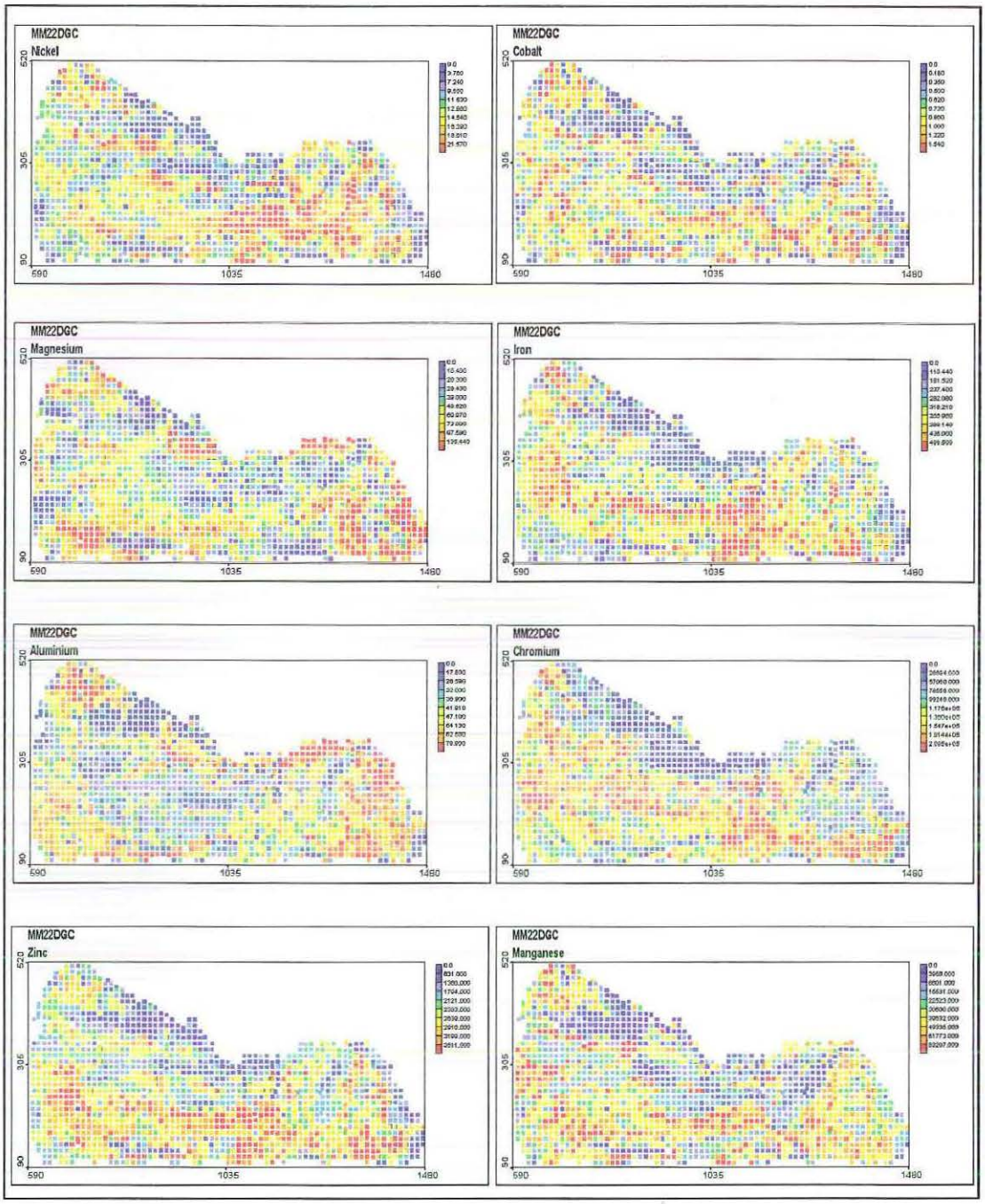


Figure 3.9: Post plots of MM22DGC data

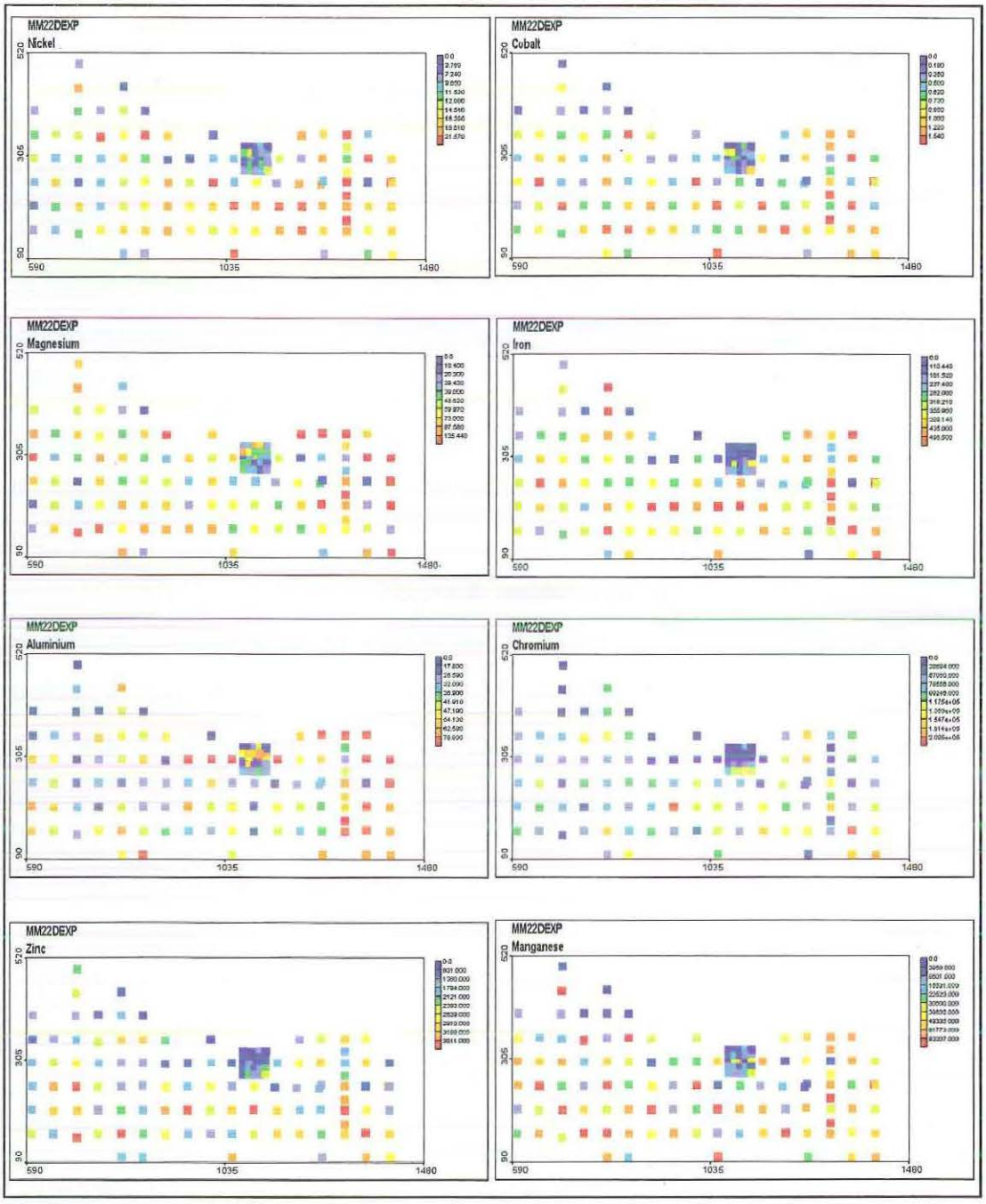


Figure 3.10: Post plots of MM22DEXP data

3.4 Principal Component Analysis

In a principal component analysis one assumes that the relationships between pairs of variables is linear, hence it is desirable that the scatterplots between variables exhibit linearity and the correlation coefficients are significantly large. Figure 3.11 displays the scatterplots between each of the variables. There appear to be moderate to strong linear relationships between nickel-cobalt, nickel-iron, nickel-chromium, nickel-zinc, nickel-manganese, cobalt-zinc, cobalt-manganese, iron-chromium, iron-zinc and zinc-chromium.

Table 3.5 displays the lower half of the correlation matrix and corresponding p-values (in parenthesis) for the exploration data. The use of a two tailed hypothesis test

$$H_0: \rho = 0 \quad \text{versus} \quad H_1: \rho \neq 0$$

(where ρ is the correlation between a pair of variables) at a significance level of 0.05, indicates that all pairs of variables, except for magnesium-chromium and magnesium-zinc, are linearly related. However as a principal component analysis does not require *all* pairs of variables to have a strong linear correlation, merely that there be *some* significant linear correlations, we conclude that the exploration data meet the assumption of linearity for a principal component analysis.

Table 3.5: Lower half of the correlation matrix for exploration data, with p-values in parenthesis

	NI	CO	MG	FE	AL	CR	ZN	MN
NI	1 (0.00)							
CO	0.72 (0.00)	1 (0.00)						
MG	0.36 (0.00)	0.23 (0.01)	1 (0.00)					
FE	0.64 (0.00)	0.58 (0.00)	0.18 (0.04)	1 (0.00)				
AL	0.30 (0.00)	0.34 (0.00)	0.38 (0.00)	0.41 (0.00)	1 (0.00)			
CR	0.52 (0.00)	0.42 (0.00)	0.13 (0.16)	0.74 (0.00)	0.31 (0.00)	1 (0.00)		
ZN	0.64 (0.00)	0.60 (0.00)	0.11 (0.21)	0.83 (0.00)	0.28 (0.00)	0.65 (0.00)	1 (0.00)	
MN	0.62 (0.00)	0.84 (0.00)	0.27 (0.00)	0.46 (0.00)	0.23 (0.01)	0.20 (0.03)	0.49 (0.00)	1 (0.00)

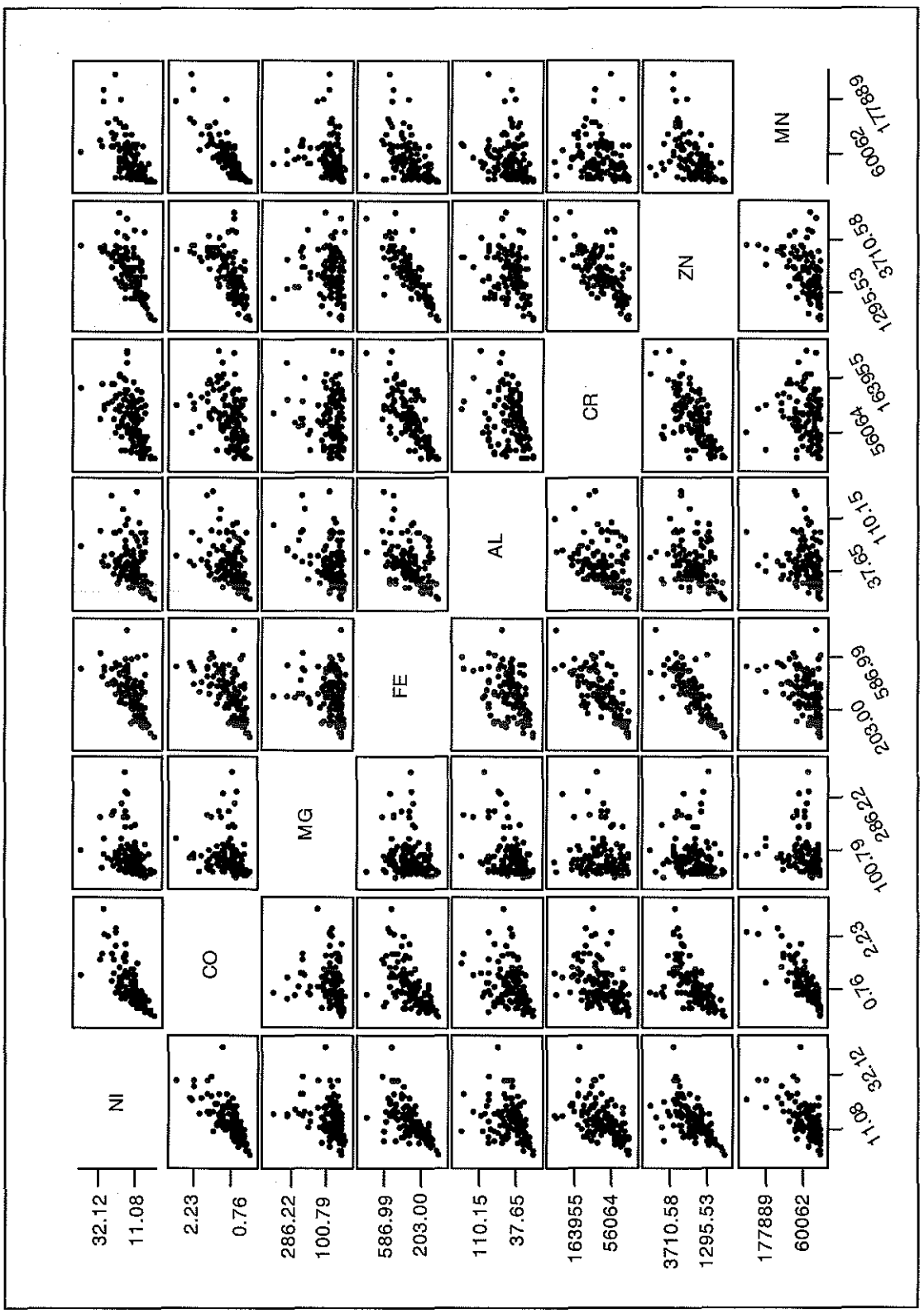


Figure 3.11: Scatterplots of the MM22DEXP variables

3.4.1 Principal Component Analysis of MM22DEXP

The results of the principal component analysis performed on the correlation matrix of all eight variables are shown in Table 3.6. The principal components extracted 53.7%, 14.8%, 13.1%, 7.6%, 4.1%, 3.5%, 1.9% and 1.4% of the total variance.

Table 3.6: Eigenanalysis of the correlation matrix of MM22DEXP

Eigenvalue	4.29	1.18	1.05	0.61	0.33	0.28	0.15	0.11
Proportion	0.54	0.15	0.13	0.08	0.04	0.04	0.02	0.01
Cumulative	0.54	0.69	0.82	0.90	0.94	0.98	0.99	1.00
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
NI	-0.41	0.08	0.13	-0.28	-0.40	-0.74	0.10	0.12
CO	-0.41	0.15	0.37	0.20	-0.26	0.23	-0.21	-0.69
MG	-0.18	0.65	-0.37	-0.57	0.19	0.20	-0.06	-0.10
FE	-0.42	-0.29	-0.17	-0.00	0.34	0.04	0.73	-0.25
AL	-0.24	0.30	-0.57	0.70	-0.06	-0.15	-0.07	0.11
CR	-0.34	-0.42	-0.33	-0.20	-0.51	0.46	-0.13	0.26
ZN	-0.40	-0.32	0.01	-0.06	0.58	-0.17	-0.60	0.07
MN	-0.35	0.31	0.51	0.16	0.13	0.31	0.17	0.60

As the variance of each standardised variable is equal to one, we seek components whose corresponding eigenvalues are greater than one. This ensures that the explanatory value of the component exceeds that of any single variable. The first three components have eigenvalues greater than one and account for 81.5% of the variability of the original data. However, consideration of the scree plot in Figure 3.12 reveals that the eigenvalues appear to level off after the fifth component. As the associated eigenvalue for the fifth component is only 0.33 it would be of no benefit to retain this component. If data reduction were the objective here we would be required to retain the first four principal components of this data set. The decision of

how many components to retain in order to account for "enough" of the variability of the original data is a subjective and strongly debated area (Afifi & Clarke, 1996). In general however one would aim to retain as many components as required to account for at least 85% of the variability. However some authors consider that as little as 80% (Johnson & Wichern, 2002) is adequate, whereas others debate that at least 90% of the variability should be accounted for (Dunteman, 1984). In this case the first four components would then account for 89% of the variability of the original data which would certainly be "enough" for even the most stringent authors.

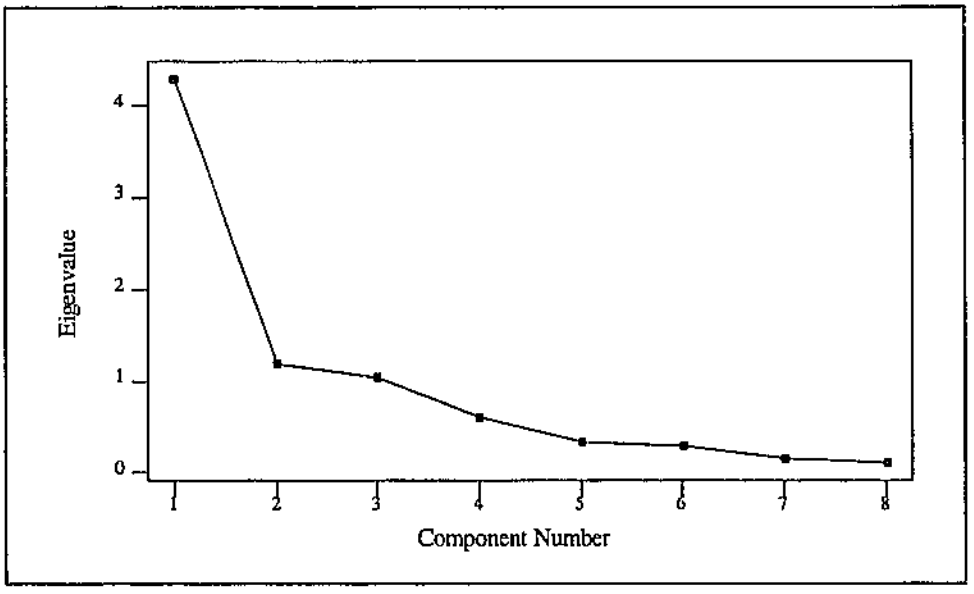


Figure 3.12: Scree plot of eigenvalues for the principal component analysis of MM22DEXP

The first principal component has the greatest weighting on nickel, cobalt, iron and zinc with manganese and chromium following. The lowest weights are associated with aluminium and magnesium. While all the weights are negative there is little to be interpreted from this component. The second principal component is dominated by magnesium and has positive weights for this variable as well as manganese, aluminium, cobalt and nickel. The weight of nickel for this component is only 0.08, so this variable barely contributes to this component. The negative weights, in order of magnitude, are associated with chromium, zinc, and iron. The

third principal component is composed of positively weighted manganese, cobalt, nickel and zinc. The contribution of zinc to this principal component is negligible with a weight of only 0.01. The negative weights are associated with aluminium, magnesium, chromium and iron. The fourth principal component is dominated by aluminium, whose weight is positive as are the weights for cobalt and manganese. The negative contributors are magnesium, nickel, chromium, zinc and iron, with the latter two being almost zero. There does not seem to be any clear interpretation for any of the principal components from this data set.

Figure 3.13 displays the scatter plots of the scores of the first four principal components. The plots between PC1-PC2, PC2-PC3 and PC3-PC4 reveal potential outliers. The corresponding sample data was checked and it was ascertained that these data were genuine, hence they could not be treated as erroneous and were left in the data set.

There does appear to be some slight curvature in some of the scatter plots, in particular those between PCS1, PCS2 and PCS3, suggesting that the strongly skewed variables cobalt, magnesium, aluminium and manganese, may benefit from a transformation. These variables were transformed logarithmically and the principal component analysis repeated. The results were compared to those obtained in the principal component analysis of the raw data and are displayed in Appendix A. The eigenvalues in each are case very close in value with the cumulative proportion of variability almost identical by the fourth component. The eigenvalues of the first and second principal components are somewhat higher for the transformed data while the third eigenvalue is somewhat lower. There have been numerous small changes in some of the elements of the coefficient vectors, though they are all small.

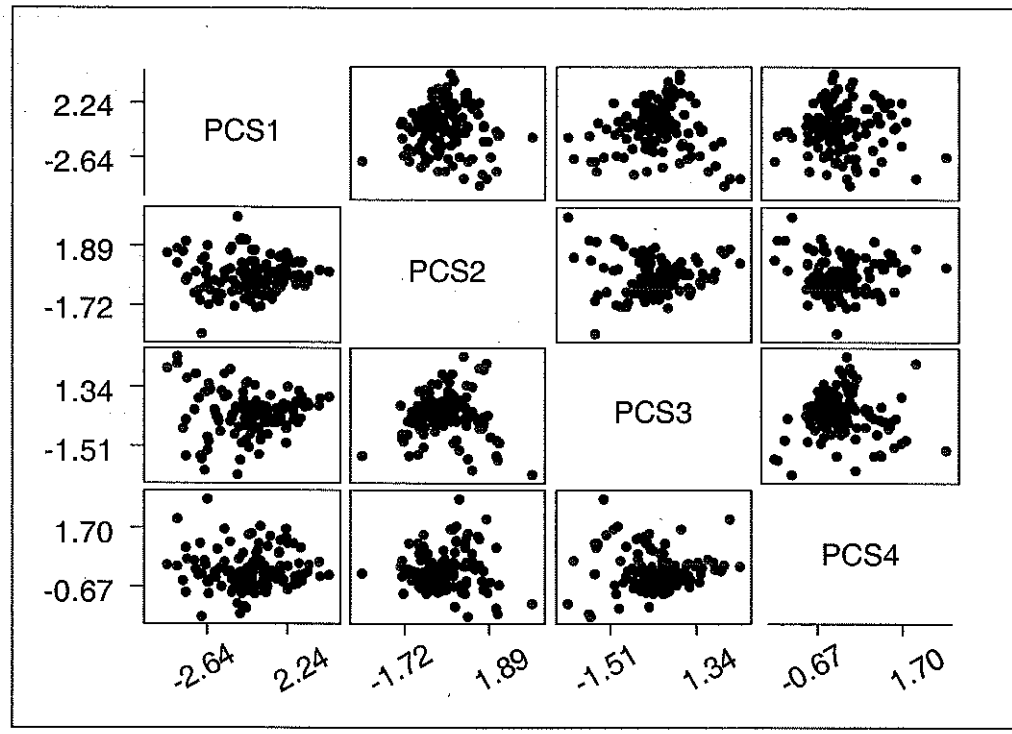


Figure 3.13: Scatter plots of the principal component scores of MM22DEXP

Furthermore, while the scatter plots of the transformed principal component scores (also displayed in Appendix A) do appear to be less correlated and circular in shape, the improvement is not substantial. Given the results of the two principal component analyses are very similar, the curvature in some of the plots in Figure 3.13 is not sufficiently marked and the dangers of exaggeration of estimation errors we feel that the data do not warrant transformation.

The principal component scores were calculated using expression (4) in Chapter 2. The omnidirectional cross correlograms of the scores were then computed at 32 lags with a lag spacing of 12.5 metres. The value of the cross correlogram at the first lag is significantly different from zero for PC1-PC2, PC1-PC4 and PC2-PC4. In fact there is clear structure in the lower lags for PC1-PC2 and PC1-PC4. There is also strong deviation from zero at the ninth and twenty-sixth lags of PC1-PC3. While the cross correlograms for PC2-PC3 and PC3-PC4 show no clear structure there are significant non-zero values at lags two and four. Figure 3.14 displays the most problematic of the cross correlograms. Clearly the original data are not intrinsically correlated.

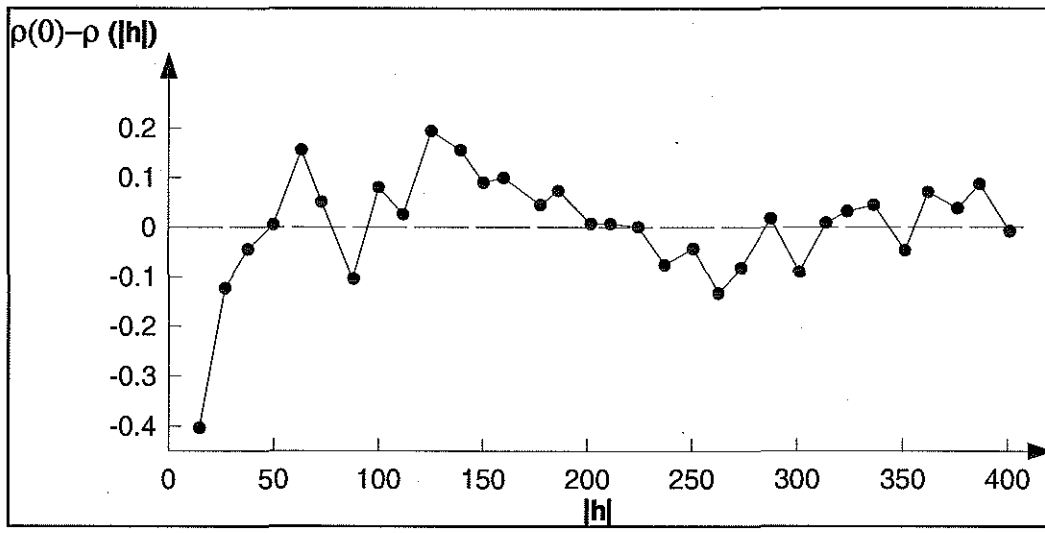


Figure 3.14: Cross correlogram for the first two principal components of MM22DEXP

In order to demonstrate the multivariate geostatistical techniques of cokriging and principal component kriging it was decided that we would investigate the possibility of creating two smaller subsets of MM22DEXP. Several subsets were investigated according to the following criteria. As the data come from a working nickel-cobalt mine these two variables were to be included in both subsets. We also required that one of the subsets consisted of only variables that are highly correlated with nickel and cobalt. We have chosen a highly correlated data set for two reasons. Firstly, as previously stated, a principal component analysis assumes that the relationship between each pair of variables is linear; hence high correlation coefficients between variables are desirable. Secondly, the higher the correlation coefficient between the primary and secondary variables, the greater the contribution of the secondary information to the estimation of the primary variable. More precisely the higher the correlation coefficient, the larger will be the kriging weights of the secondary variables. The highly correlated data set is MM22DHC4 and consists of nickel, cobalt, iron and zinc. The choice of the second subset was based on the fact that the data are from an actual working nickel-cobalt mine; hence we considered it appropriate to include the variables considered to be economically most important to the mining company. This subset is referred to as MM22DTOP4 and consists of nickel, cobalt, magnesium and iron.

3.4.2 Principal Component Analysis of MM22DHC4

The results of the principal component analysis performed on the correlation matrix of nickel, cobalt, iron and zinc are shown in Table 3.7. The principal components extracted 75%, 14%, 7% and 4% of the total variance. Only the first component has eigenvalue greater than one and accounts for 75% of the variability of the original data. The first two components account for an impressive 89% of the variability of the original data. The scree plot displayed in Figure 3.15 shows that the principal components level off at the third and fourth components. If data reduction were the objective here we would be required to retain the first two principal components of this data set. These would then account for 89% of the variability of the original data.

Table 3.7: Eigenanalysis of the correlation matrix of nickel, cobalt, iron and zinc from MM22DHC4

Eigenvalue	3.00	0.55	0.27	0.17
Proportion	0.75	0.14	0.07	0.04
Cumulative	0.75	0.89	0.96	1.000
Variable	PC1	PC2	PC3	PC4
NI	-0.50	-0.42	0.76	0.04
CO	-0.48	-0.60	-0.64	-0.06
FE	-0.51	0.50	-0.02	-0.69
ZN	-0.51	0.47	-0.12	-0.71

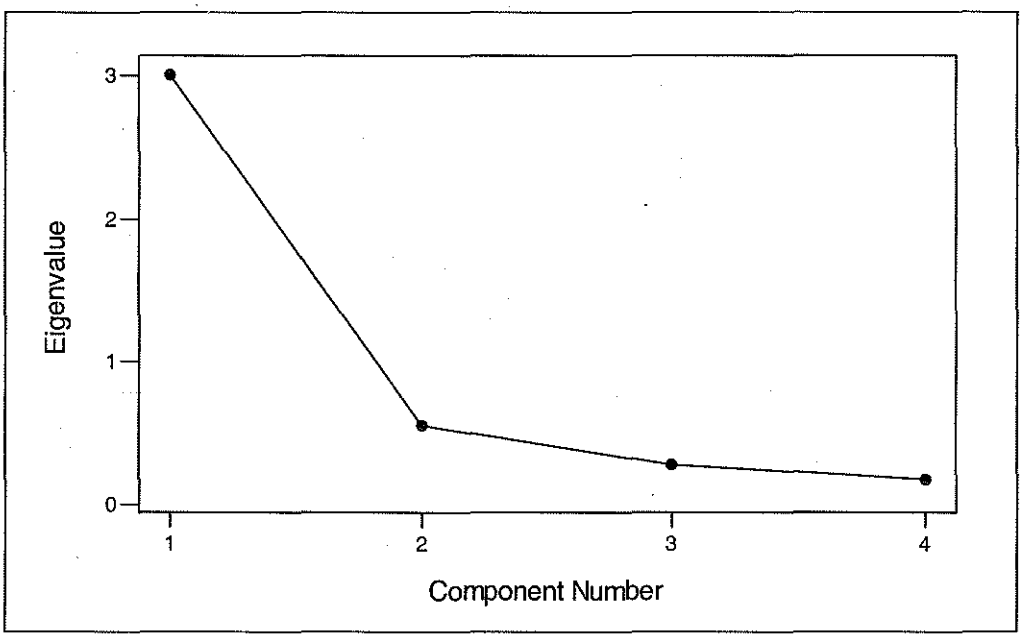


Figure 3.15: Scree plot of eigenvalues for the principal component analysis of MM22HC4

The first principal component has almost equal weighting on nickel, cobalt, iron and zinc. As the weights are all the same sign and of almost equal magnitude it is possible to consider this component to be an overall grade component. The second principal component has its largest negative weight associated with cobalt followed by nickel. Zinc and iron are both positive and are fairly equal in magnitude. There is a clear grouping of nickel-cobalt and iron-zinc. While the actual sign of the coefficients is not relevant, the fact that the groupings display opposite signs is. This suggests that this component is a contrast of nickel-cobalt and iron-zinc. The third principal component is dominated by nickel which in this case is positive. Cobalt, iron and zinc have negative weights however iron barely contributes to this component with a weight of only -0.02 and the weighting on zinc is also small. Hence, as the nickel and cobalt weights are of opposite sign this component appears to be a nickel-cobalt contrast component. The fourth principal component is dominated by negative weights on zinc and iron with cobalt and nickel barely contributing. As the weights are the same sign and of similar magnitude this component appears to be a zinc-iron component. Further interpretation of these principal components requires knowledge of their geological significance. This is

beyond the scope of this research and would require the consultation of a geologist with the relevant expertise.

The scatter plots of the principal components are displayed in Figure 3.16. As outlined earlier it has been decided that the data do not warrant transformation. There is some curvature in the plot of PC1-PC2 however the remainder are sufficiently uncorrelated. There do appear to be some outlying values, particularly in the plots including PC4. The corresponding sample data was checked and it was ascertained that these data were genuine; hence they could not be treated as erroneous and were left in the data set.

Figure 3.17 shows the cross correlograms of the first three principal components from MM22DHC4. Despite an initial high value at the first lag of the cross correlogram of the first and second principal components the remainder of the lags of this and the other graphs appear to be sufficiently close to zero. In addition there is little to no structure in the first plot and clearly no structure in the subsequent plots. The cross correlograms then give an indication that it would be possible to use an intrinsic coregionalisation model for the variables in the MM22DHC4 data set.

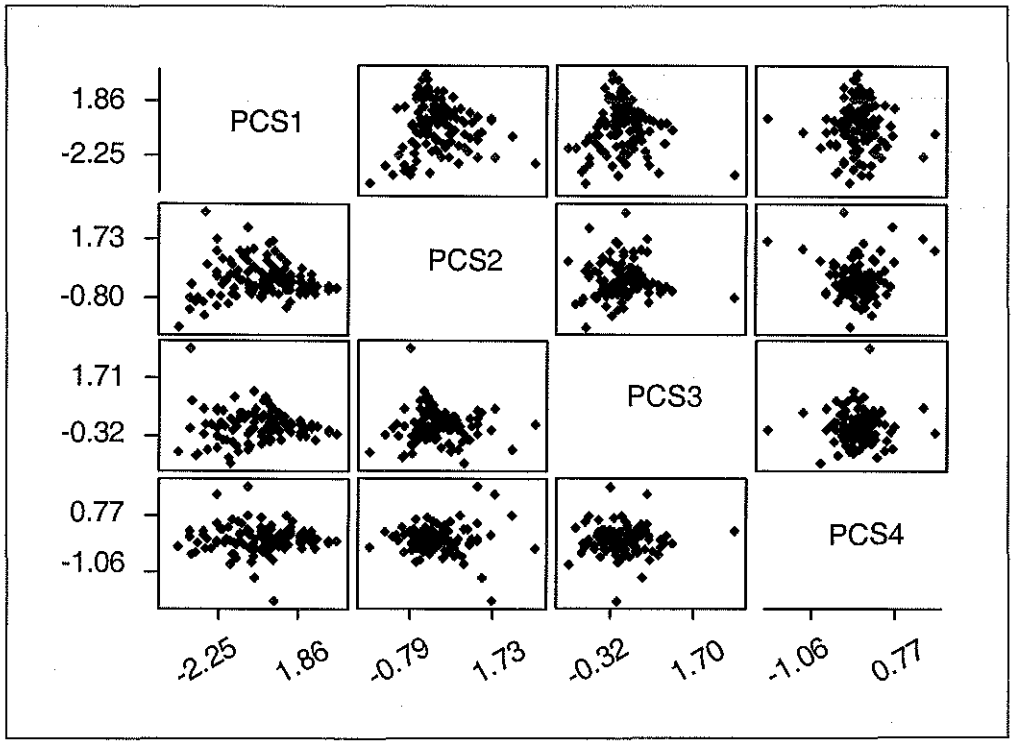


Figure 3.16: Scatter plots of the principal components from MM22DHC4

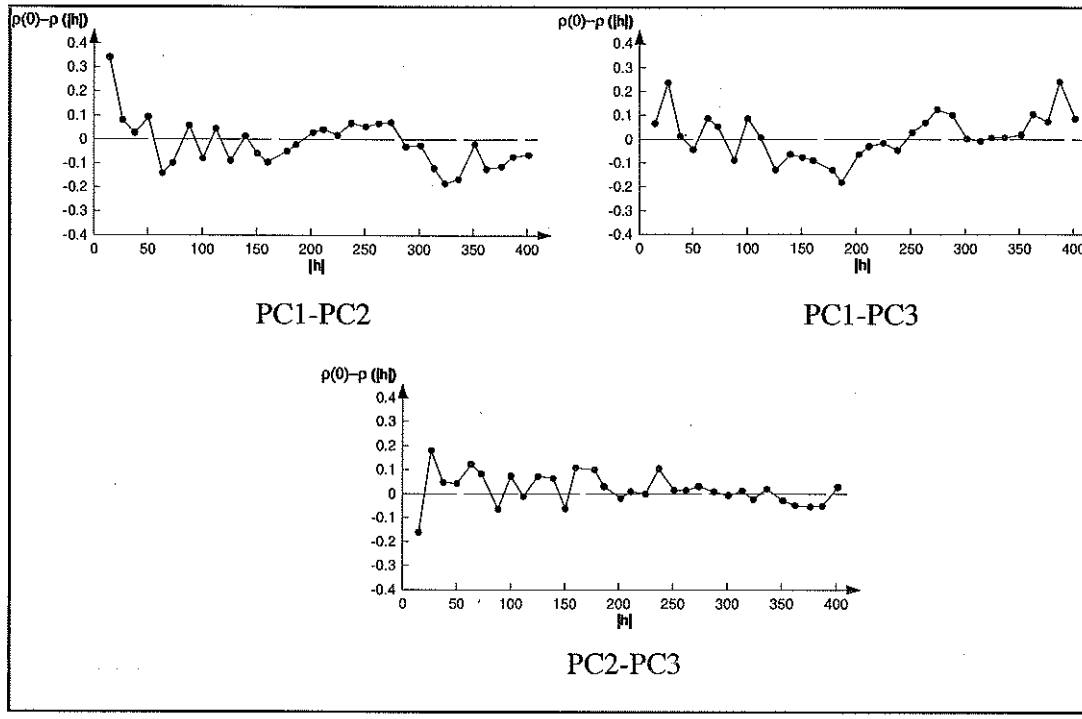


Figure 3.17: Cross correlograms of the first three principal components from MM22DHC4

3.4.3 Principal Component Analysis of MM22DTOP4

The results of the principal component analysis performed on the correlation matrix of nickel, cobalt, iron and magnesium are shown in Table 3.8. The principal components extracted 61%, 22%, 11%and 6% of the total variance. Only the first component has eigenvalue greater than one and accounts for 61% of the variability of the original data. The first two components account for 83% of the variability of the original data. The scree plot displayed in Figure 3.18 shows that the principal components do not level off at all. If data reduction were the objective here we would be required to retain the first three principal components of this data set. These would then account for 94% of the variability of the original data, however one would question the utility of reducing a four variable data set to a three variable set.

Table 3.8: Eigenanalysis of the correlation matrix of nickel, cobalt, magnesium and iron from MM22DTOP4

Eigenvalue	2.43	0.88	0.43	0.26
Proportion	0.61	0.22	0.11	0.06
Cumulative	0.61	0.83	0.94	1.00
Variable	PC1	PC2	PC3	PC4
NI	-0.58	-0.03	-0.20	0.79
CO	-0.55	-0.20	-0.58	-0.56
MG	-0.30	0.94	0.11	-0.16
FE	-0.52	-0.29	0.78	-0.20

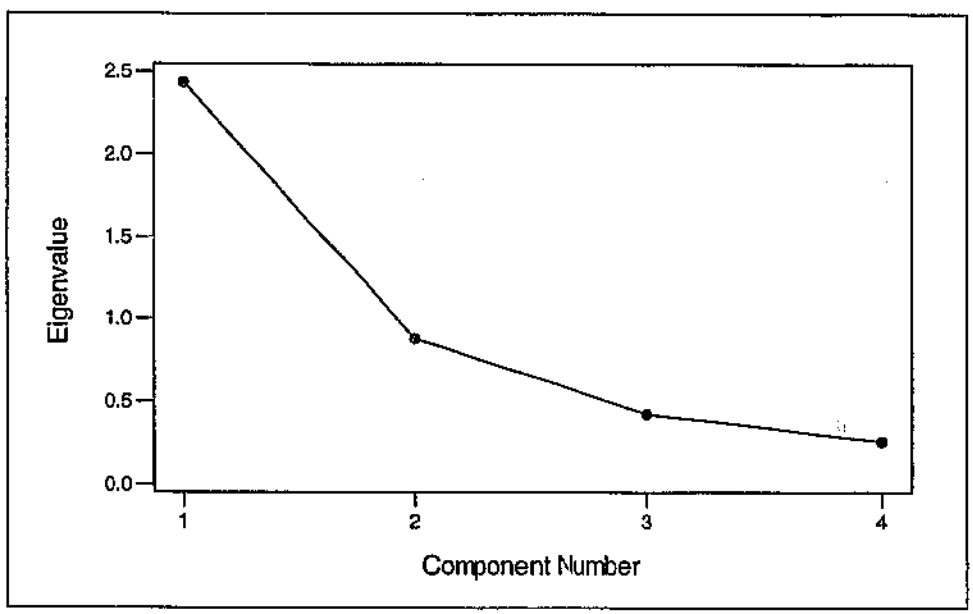


Figure 3.18: Scree plot of eigenvalues for the principal component analysis of MM22DTOP4

The first principal component has almost equal weighting on nickel, cobalt and iron. While the weights are all the same sign magnesium is somewhat smaller in magnitude than the others. Hence this component is not as easily interpreted as for the previous case, although it could still be thought of as an overall grade component. The second principal component is almost totally dominated by magnesium. Nickel barely contributes to this component whereas cobalt and iron have similar weighting. Again this component is not as easily interpreted, other than perhaps being a magnesium component. The third principal component is dominated by iron. There is a positive grouping of magnesium and iron, though the weights are by no means close in value. Nickel and cobalt are both negative for this component, once again the weighting is not equal. However as the weightings on nickel and magnesium are quite small we could consider this a contrast component of cobalt and iron. The fourth principal component is dominated by nickel with all the other weights being negative. Magnesium and iron have fairly equal weighting but are relatively small in magnitude. Hence this could be considered a nickel-cobalt contrast component. Once again further interpretation of these principal components requires knowledge of their geological significance and is beyond the scope of this research.

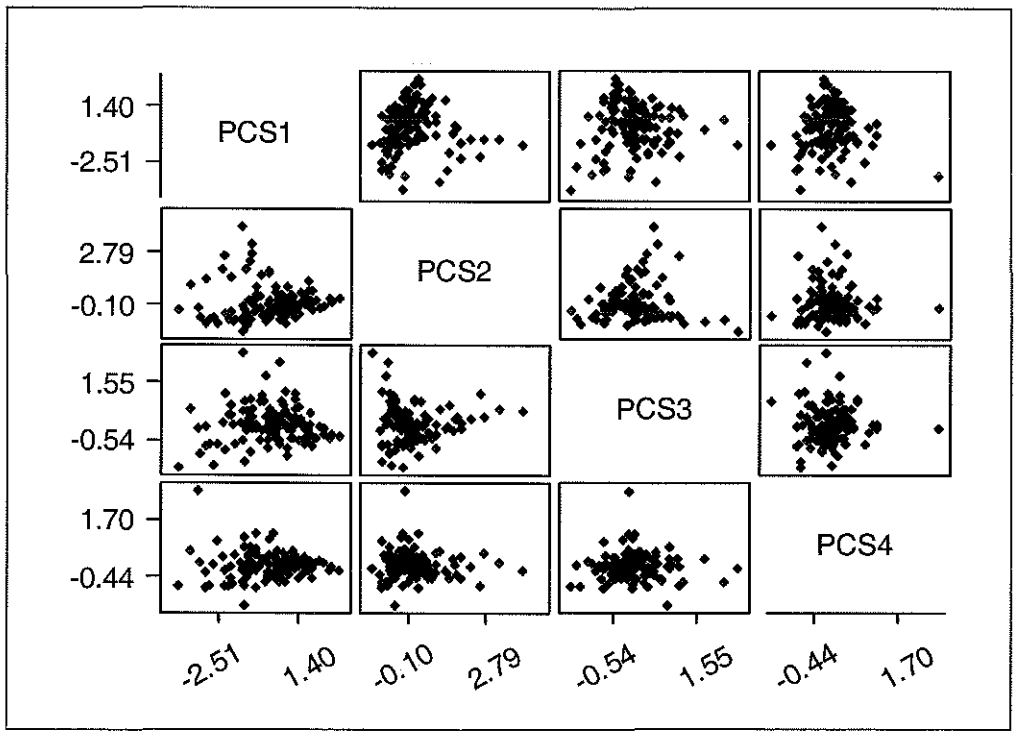


Figure 3.19: Scatter plots of the principal components from MM22DTOP4

The scatter plots of the principal components are displayed in Figure 3.19. As outlined earlier it has been decided that the data do not warrant transformation. As for the previous case there is some curvature in the plot of PC1-PC2 however the remainder are sufficiently uncorrelated. There do appear to be some outlying values, particularly in the plots including PC3 and PC4. The corresponding sample data was checked and it was ascertained that these data were genuine; hence they could not be treated as erroneous and were left in the data set.

Figure 3.20 shows the cross correlogram of the first two principal components from MM22DTOP4. The first two lags have significantly non-zero correlogram values. In addition there is definite structure at the lower lags which indicates that the data are not intrinsically correlated. For this data set we will be required to proceed with the linear model of coregionalisation and perform the estimation using cokriging.

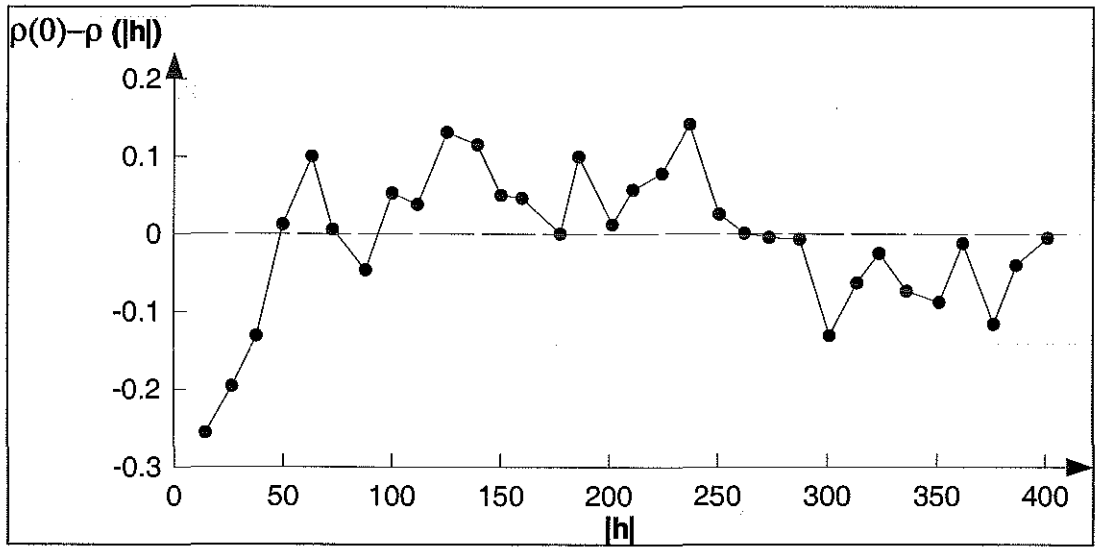


Figure 3.20: Cross correlograms of the first two principal components from MM22DTOP4

4 ORDINARY COKRIGING

In this chapter we demonstrate the estimation technique of ordinary cokriging using two data sets, MM22DHC4 and MM22DTOP4. To begin with we discuss the variography of the MM22DHC4 and MM22DTOP4 data sets. The former was modelled using an intrinsic coregionalisation model; the latter was fitted with a linear model of coregionalisation. All direct and cross semivariograms were calculated with 32 lags at a lag spacing of 12.5 metres and a lag tolerance of 6.25 metres. The directional semivariograms were calculated with an angular tolerance of 45°. All directional models were assessed for fit in the intermediate directions (with angular tolerance set to 22.5°).

All models were cross validated using ordinary cokriging and performed satisfactorily. The cokriging was performed using the software package ISATIS. This package treats each variable nominated for estimation as the primary variable in turn, using the other variables as secondary variables. 1717 estimates were obtained for each variable directly at the locations of the grade control data. The parameter files for all kriging procedures are displayed in Appendix C. Comparisons of the estimates obtained from the cokriging of each data set were made with the MM22DGC grade control data.

4.1 VARIOGRAPHY AND COKRIGING OF MM22DHC4

4.1.1 Variography and the Intrinsic Coregionalisation Model

Recall that in the generic intrinsic correlation model the sills, b_{ij}^l, of any basic structure $g_l(\mathbf{h})$ that constitute the model are proportional to each other. That is $b_{ij}^l = \varphi_{ij} \cdot b_{ij}^l$ for all i, j and l. As we are dealing with standardised data we require that $\sum_{l=0}^{L} b^l = 1$. Under the assumption of second-order stationarity the matrix of coefficients $\Phi = [\varphi_{ij}]$ is the positive semi-definite variance-covariance matrix $\mathbf{C}(0)$, which in our case is the correlation matrix:

$$\mathbf{R} = \begin{bmatrix} 1 & 0.72 & 0.64 & 0.64 \\ 0.72 & 1 & 0.58 & 0.60 \\ 0.64 & 0.58 & 1 & 0.83 \\ 0.64 & 0.60 & 0.83 & 1 \end{bmatrix} = \mathbf{\Phi}$$

The direct and cross semivariogram surfaces for nickel, cobalt, iron and zinc are displayed in Figure 4.1. The semivariogram surface of nickel exhibits the presence of two anisotropic structures, one short range and one long range. The direction of maximum continuity of the short range structure is east-west (azimuth 90°). The direction of maximum continuity of the long range structure is approximately azimuth 70°. Cobalt, iron and zinc also exhibit anisotropy with the direction of maximum continuity being east-west. The east-west anisotropy is also apparent in the cross semivariogram surfaces. Hence the semivariograms were modelled with the direction of maximum continuity being east-west.

The semivariograms were fitted with an intrinsic correlation model consisting of three basic structures, a nugget effect and two spherical structures. The first spherical structure is omnidirectional with a range of 100 metres; the second is modelled with east-west as the major direction with a range of 200 metres and an anisotropy ratio of 0.5. This model can be expressed as:

$$\begin{bmatrix} 1 & 0.72 & 0.64 & 0.64 \\ 0.72 & 1 & 0.58 & 0.60 \\ 0.64 & 0.58 & 1 & 0.83 \\ 0.64 & 0.60 & 0.83 & 1 \end{bmatrix} \left(0.05 g_0(\mathbf{h}) + 0.5 Sph\left(\frac{|\mathbf{h}|}{100}\right) + 0.45 Sph\left(\frac{\mathbf{h'}}{200}\right) \right)$$

$$where \quad \mathbf{h'} = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.5 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figures 4.2 and 4.3 display the direct and cross semivariograms fitted with this model.

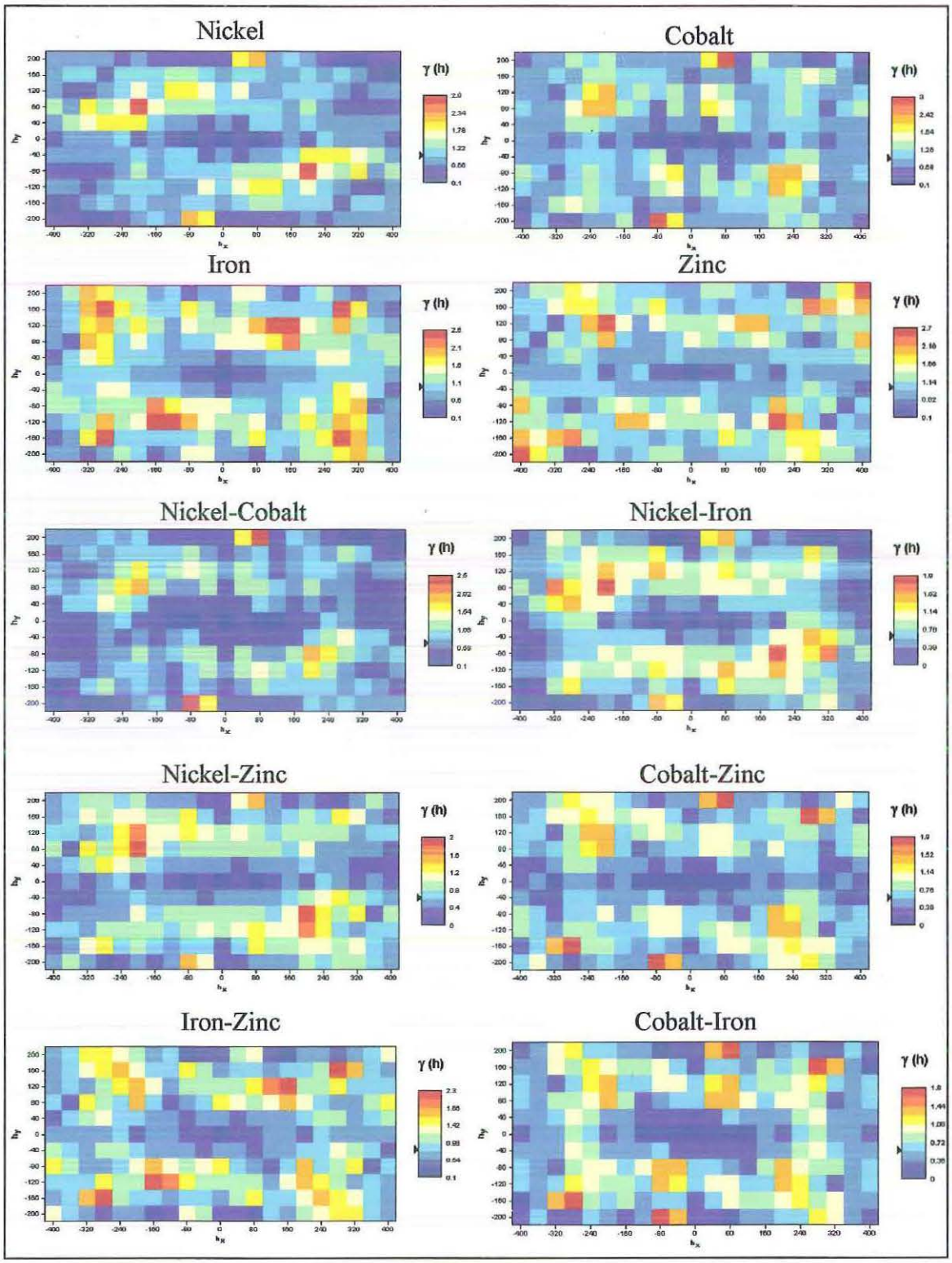


Figure 4.1: Direct and cross variogram surfaces for MM22DHC4

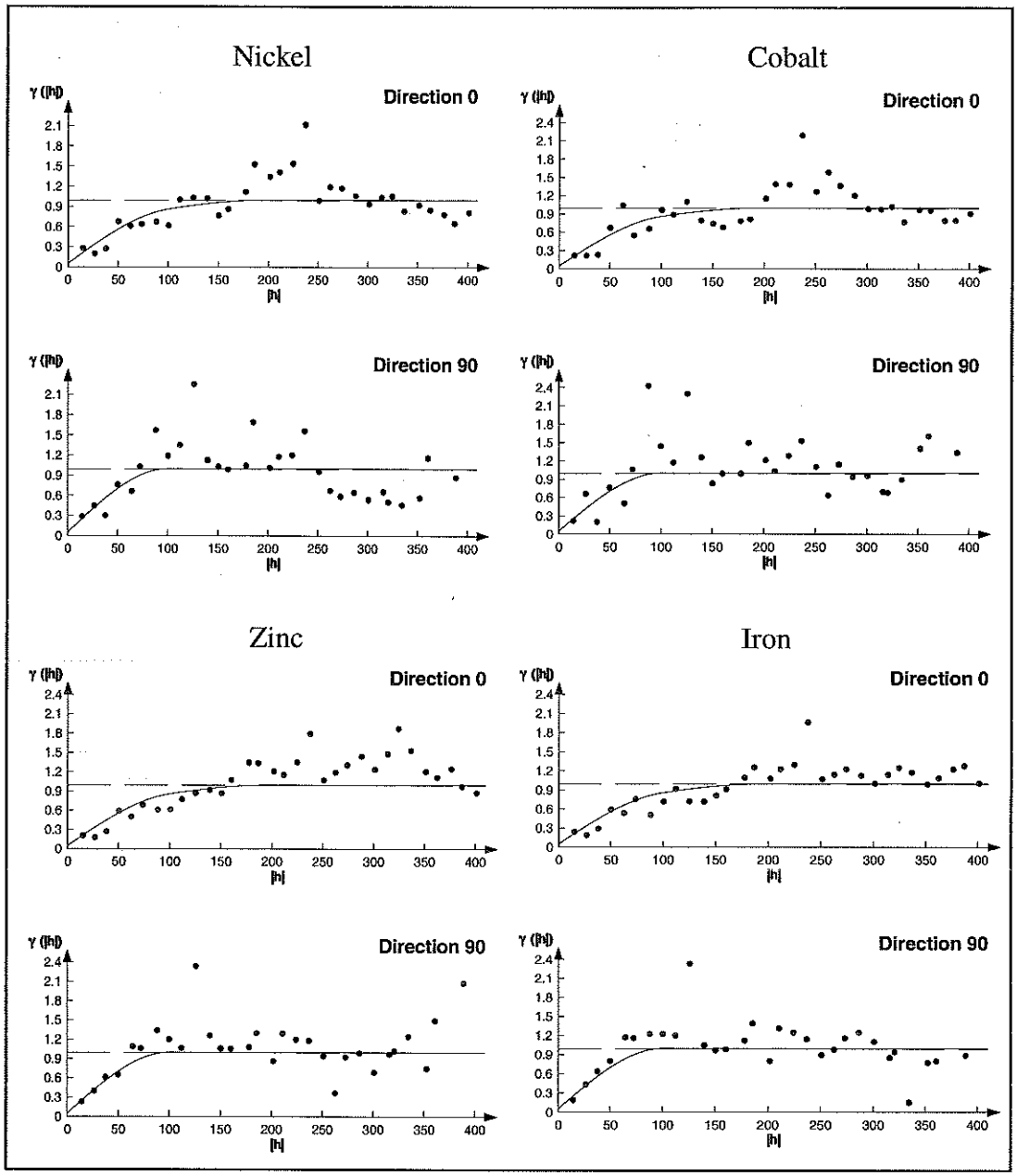


Figure 4.2: Fitted direct semivariograms for MM22DHC4

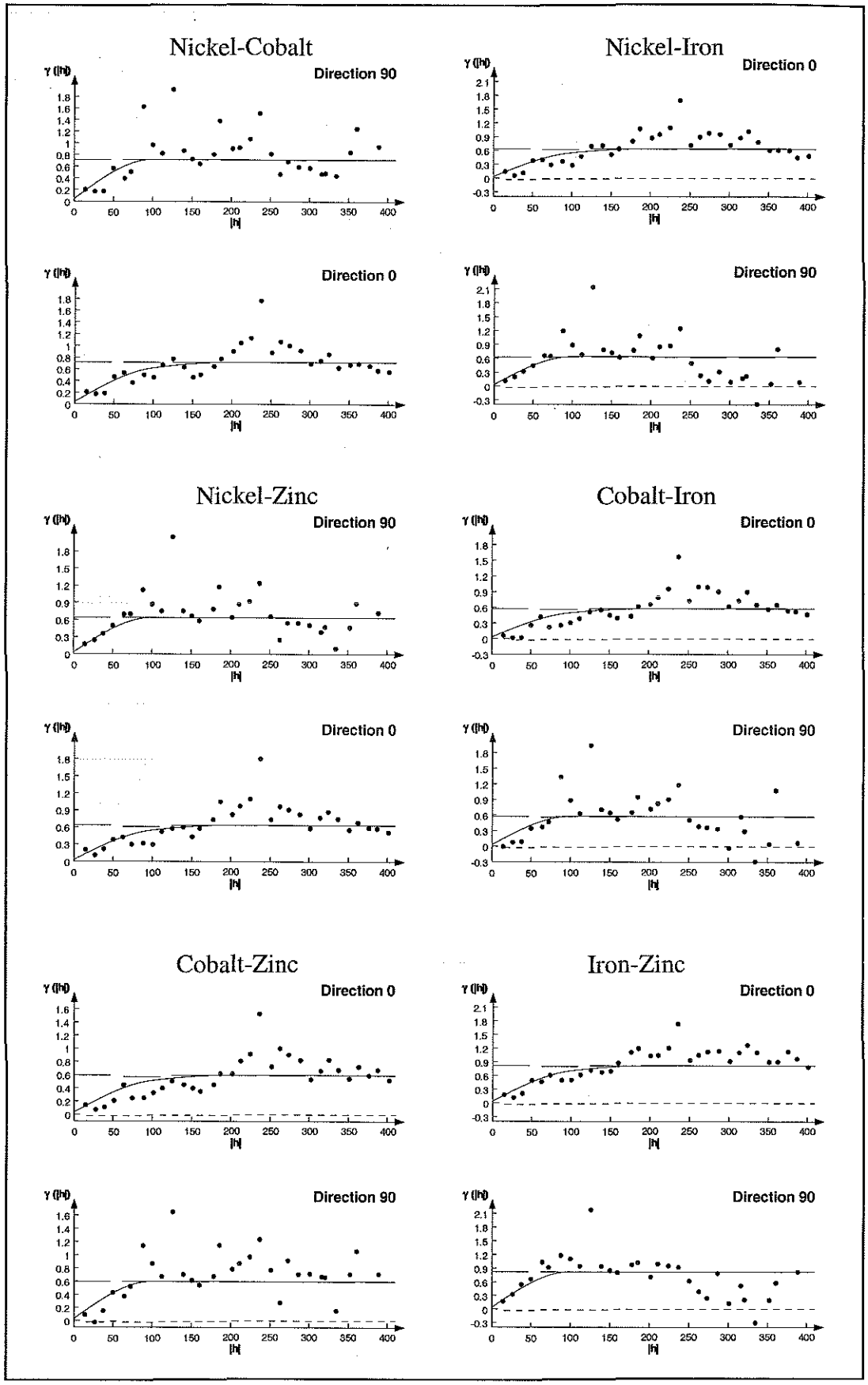


Figure 4.3: Fitted cross semivariograms for MM22DHC4

4.1.2 Cross Validation for Cokriging of MM22DHC4

One method of assessing the suitability of a model is to use cross validation. This technique allows us to compare estimates calculated from our chosen model using the specified kriging algorithm with the actual data at each sample location. The cross validation procedure removes one sample data from the data set, estimates it using the specified model, replaces it then repeats the procedure for all other sample data in turn. An analysis of the residuals is then carried out to check for normality of the errors, presence of outliers and suitability of the neighbourhood parameters.

The cross validation was carried out for all of the variables in this data set using ordinary cokriging to perform the estimation. We present here the results for nickel, those for cobalt, iron and zinc are presented in Appendix B. There was little difference in the cross validation results of several various search neighbourhoods investigated. Figure 4.5 shows the cross validation residual analysis for the search parameters displayed in Table 4.1 and Figure 4.4. The histogram of residuals appears roughly normally distributed with a few extreme high and low values apparent. These extreme values are also apparent in both the plots of the true values versus the fits and the standardised estimates. The location of these values is exhibited in the base map. These values were verified and as no justification could be made for their removal, they were left in the data set. The plot of estimates versus the true values is reasonably well clustered around the 45° bisector, once again with some deviations. Finally the mean error and standardised error variance displayed in Table 4.2 are sufficiently close to the optimal values of zero and one respectively (Bleines et. al, 2001, p. 126). All other neighbourhood search parameters tested failed to further improve any of these results. Overall the model appears to have reproduced the sample (exploration) values successfully for all variables.

Table 4.1: Neighbourhood search parameters

Number of angular sectors	4
Minimum number of samples	4
Optimal number of samples per sector	4
Search radius (major axis of ellipse) (metres)	220
Search radius (minor axis of ellipse) (metres)	120

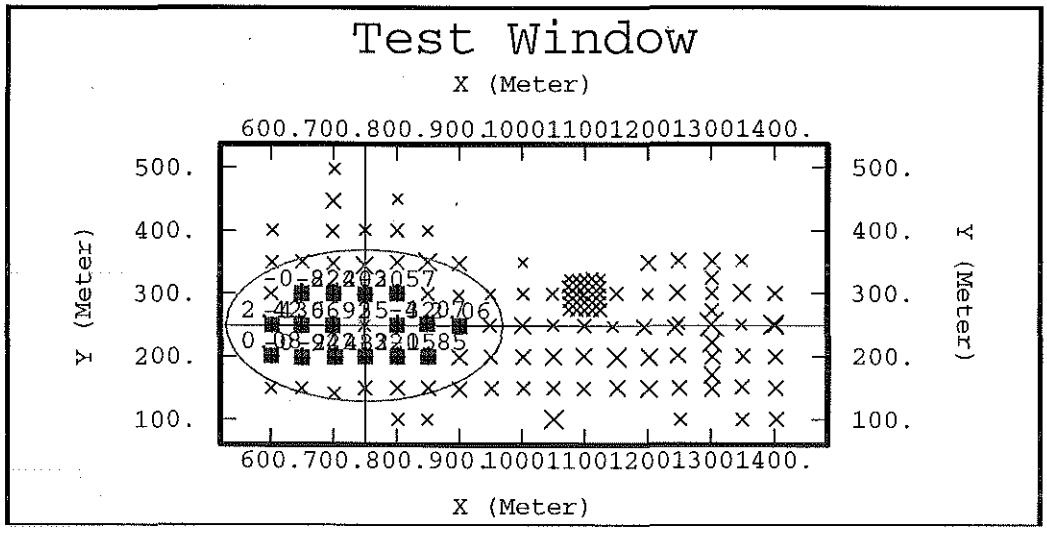


Figure 4.4: Search neighbourhood for MM22DHC4

Table 4.2: Mean and variance of cross validation residuals

	Nickel		Cobalt		Iron		Zinc	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Error	0.02	1.00	0.02	1.24	0.01	0.66	0.02	0.64
St. Error	0.01	2.39	0.02	2.57	0.01	1.55	0.02	1.50

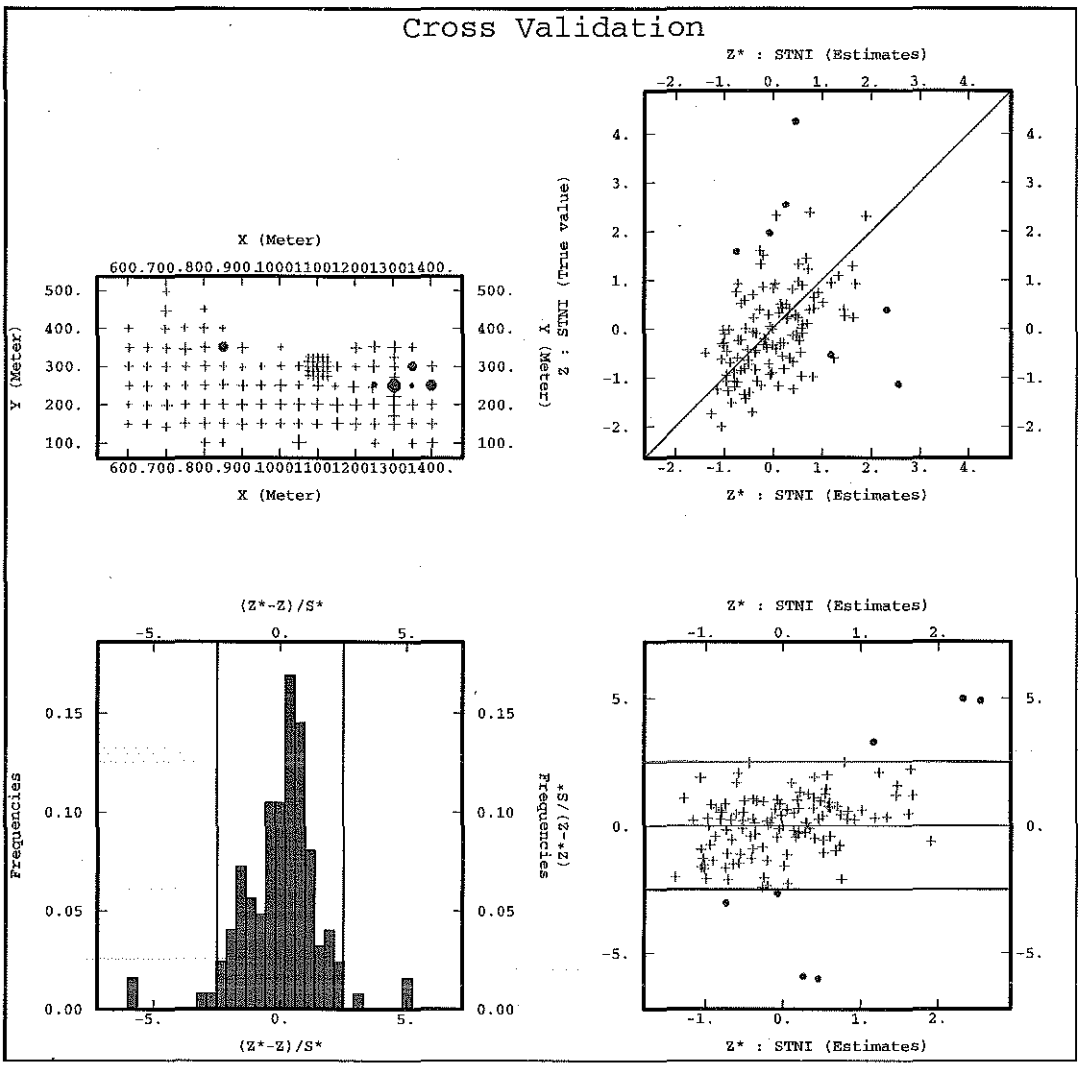


Figure 4.5: Cross validation residual analysis for MM22DHC4 nickel

4.1.3 Cokriging Estimates of MM22DHC4

Ordinary cokriging of the nickel, cobalt, iron and zinc variables was performed directly at the grade control coordinates using the search neighbourhood parameters presented earlier (Table 4.1). As all variography, modelling and estimation was performed on the standardised data it was necessary to transform the estimates by multiplying each value by the sample standard deviation and adding to this the sample mean. Figure 4.6 shows post plots of the estimates for each of the variables along with the post plots of the grade control data. The smoothing nature of the cokriging algorithm is clearly apparent here. The variability of the grade control data has not been reproduced particularly well though overall the regions of high and low values have been well represented. The lack of precise representation of the full variability of the exhaustive data is however a drawback of any regression estimation technique.

For all four variables the overall appearance of the post plots of the estimates is consistent with the grade control post plots. In particular the estimates along the north-western, western and southern perimeters are representative of the grade control values. The values along the eastern perimeter of the study region have been particularly poorly estimated. However most of these values lie outside the region defined by the locations of the exploration data hence have been estimated using values that are in some cases at least 75 metres away. Nickel seems to have been the most affected by this as the last line of exploration drill holes consists of high values only. Zinc is the least affected with pockets of high and low values still well represented in this region.

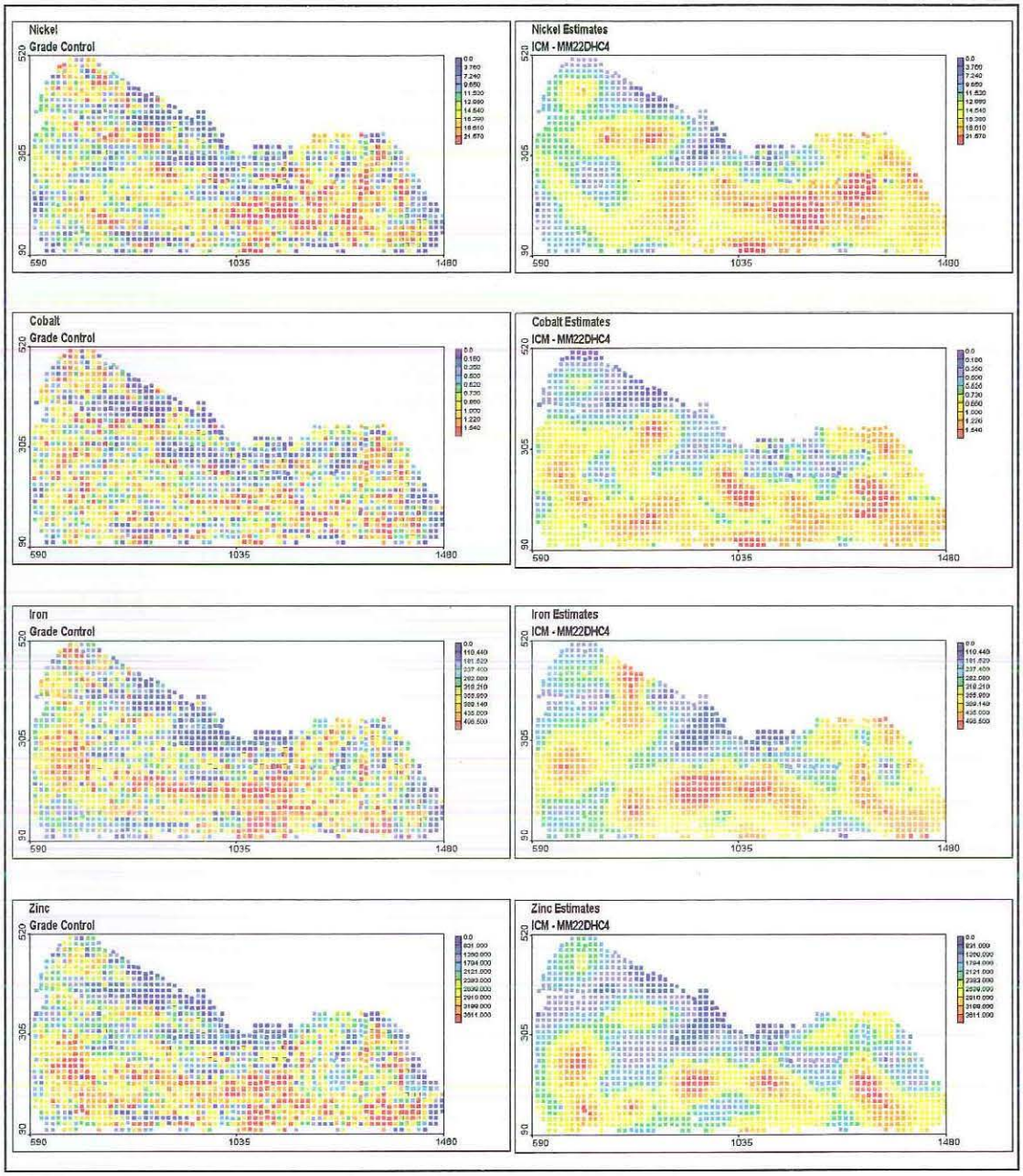


Figure 4.6: Cokriged estimates – Intrinsic Coregionalisation Model MM22DHC4

The residual post plots displayed in Figure 4.7 reflect the poor estimation along the eastern perimeter with some of the highest residuals occurring here. In general the area of greatest underestimation was the far north-western corner of the study region. Iron appears to have been affected the most in this area with the greatest underestimation occurring at the far north-western perimeter and the greatest overestimation occurring just to the east of this region. In most cases the largest errors were associated with the areas containing extreme high or low values. Figure 4.8 displays the normal probability plots and histograms of the residuals for nickel, cobalt, iron and zinc. The histograms of residuals for nickel, iron and zinc are reasonably symmetric and appear to be approximately normally distributed which is supported by the linear shape of the normal score plots. The histogram of cobalt residuals however is negatively skewed. Despite one clear outlying residual, the normal plot shows an approximately linear shape that is consistent with a normal distribution.

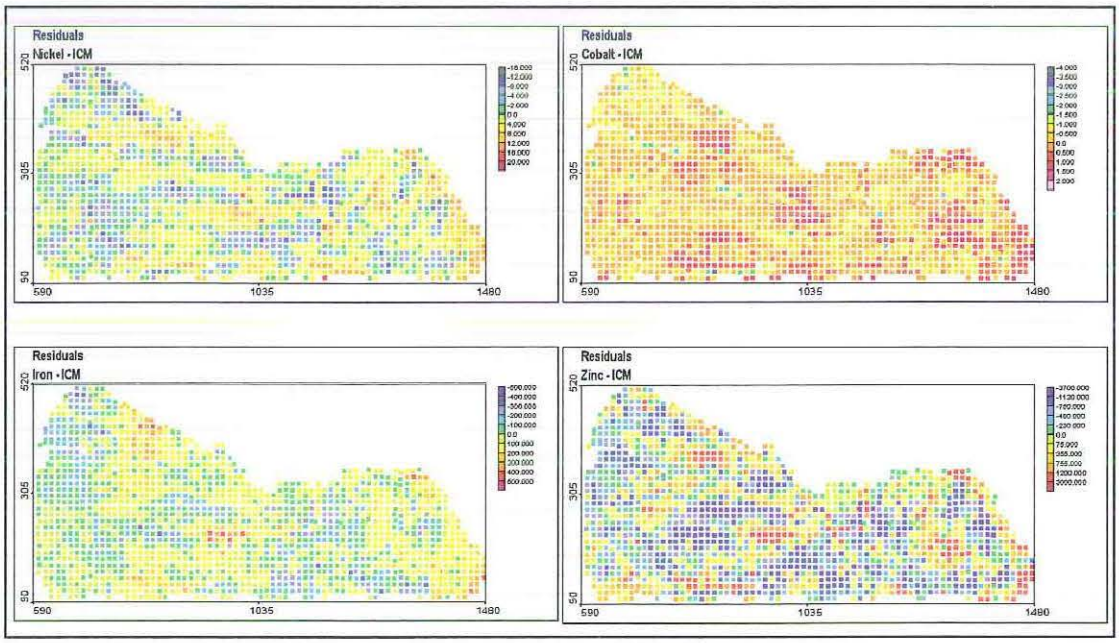


Figure 4.7: Residual post plots for MM22DHC4

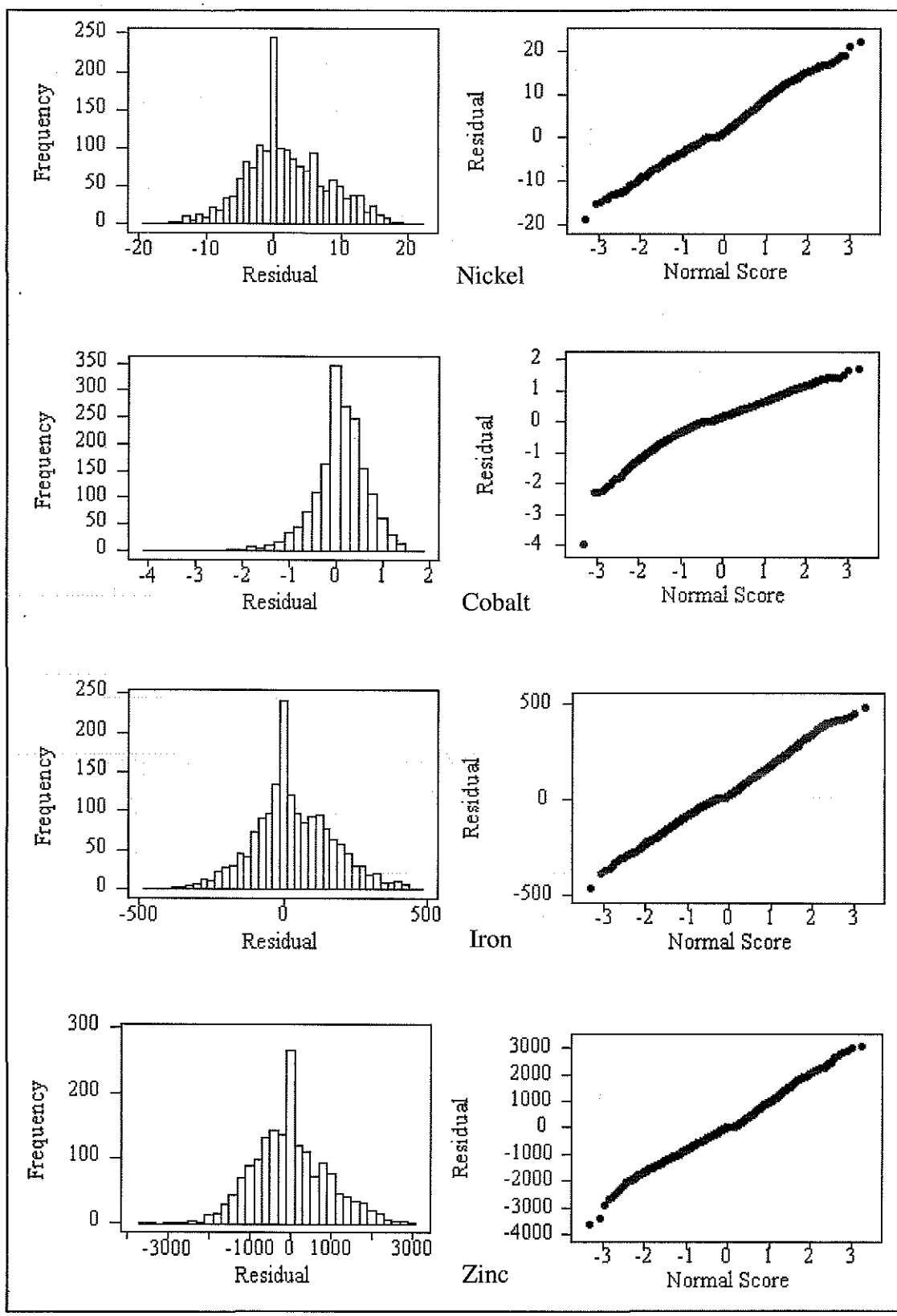


Figure 4.8: Residual histograms and normal probability plots for MM22DHC4

4.2 VARIOGRAPHY AND COKRIGING OF MM22DTOP4

4.2.1 Variography and the Linear Model of Coregionalisation

As the variables in the MM22DTOP4 data set are not intrinsically correlated it is necessary for us to jointly model the direct and cross semivariograms and build a linear model of coregionalisation. Goovaerts (1997, p. 114) identifies some practical guidelines for selecting the basic structures in a linear model of coregionalisation. These include that every basic structure incorporated in the cross semivariogram model $\gamma_{ij}(\mathbf{h})$ must be present in both direct semivariogram models $\gamma_{ii}(\mathbf{h})$ and $\gamma_{jj}(\mathbf{h})$. Furthermore if a basic structure is not present on a direct semivariogram model it cannot be incorporated on any cross semivariogram model involving this variable. However it is not necessary for a cross semivariogram model $\gamma_{ij}(\mathbf{h})$ to include every structure that appears in both direct semivariogram models $\gamma_{ii}(\mathbf{h})$ and $\gamma_{jj}(\mathbf{h})$.

The procedure used for constructing our linear model of coregionalisation was to model the direct semivariograms for each of the five variables then use these basic structures to obtain the best possible fit for the cross semivariograms. In addition to the direct and cross variogram surfaces for nickel, cobalt and iron displayed previously in Figure 4.1 those for magnesium are displayed in Figure 4.9. The direct surface for magnesium is isotropic as is the surface for iron-magnesium. Nickel-magnesium and cobalt-magnesium are anisotropic with the direction of maximum continuity being east-west.

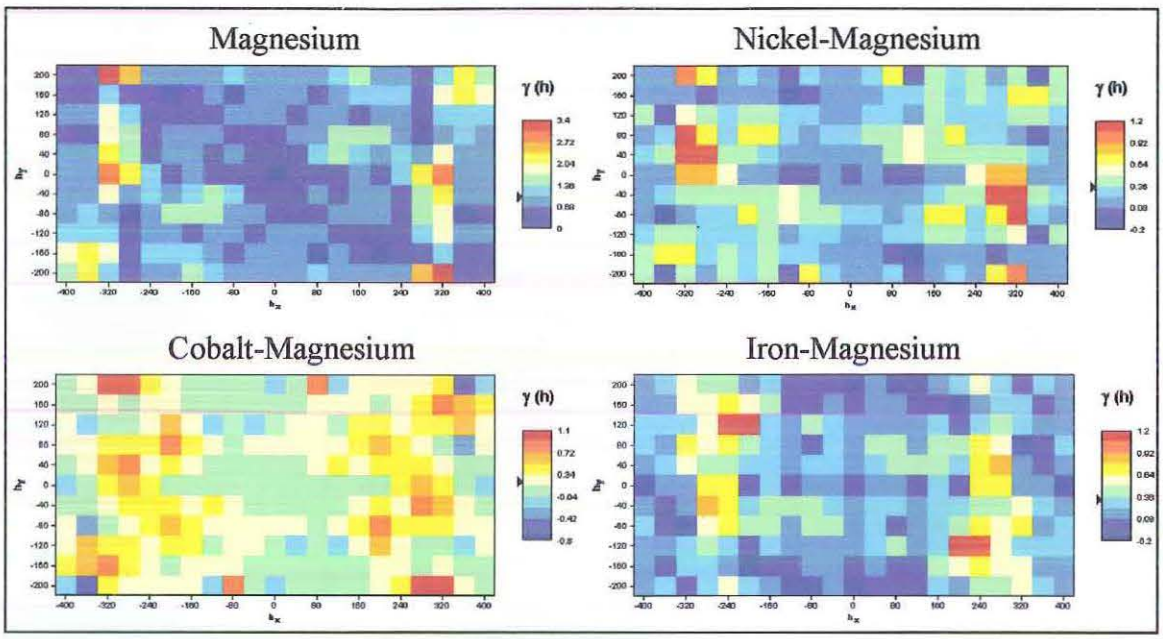


Figure 4.9: Magnesium direct and cross semivariogram surfaces

Having obtained a suitable fit on all the direct and cross semivariograms it was necessary to check the coregionalisation matrices for positive semi-definiteness. The criterion used to check for this condition is that a matrix whose eigenvalues are greater than or equal to zero is a positive semi-definite matrix (Datta, 1995, p. 22). Furthermore a symmetric diagonally dominant matrix with positive diagonal entries is positive definite. This second characteristic allowed us to modify the coregionalisation matrices in turn to ensure the positive semi-definiteness condition was met. This meant a combination of increasing the values on the main diagonals and decreasing the values on the off diagonals until a positive semi-definite matrix was achieved. Having deduced one matrix, the others were modified such that the sum of the coefficients for each basic model equalled the total sill for that semivariogram. The process of creating a diagonally dominant matrix was then repeated for the next coregionalisation matrix and so on until all three matrices were positive semi-definite. While it is recognised here that these matrices may not be optimally positive semi-definite (as for the minimised weighted sum of squares algorithm outlined in Goovearts (1997)) with only four variables it was possible to deduce the positive semi-definite matrices manually in order to provide an adequate fit.

The final model selected consists of three basic structures, a nugget effect and two spherical structures. The eigenvalues corresponding to each coregionalisation matrix are displayed in Table 4.3 confirming that the model is permissible.

Table 4.3: Eigenvalues corresponding to each of the coregionalisation matrices

Nugget	Spherical 1	Spherical 2
0.02	0.09	0.01
0.05	0.13	0.13
0.07	0.30	0.70
0.18	0.76	1.60

The first spherical structure is isotropic and has a range of 75 metres. The second spherical structure is modelled with the direction of maximum continuity in the east-west direction and has a range of 200 metres and an anisotropy ratio of 0.7. This model can be expressed as:

$$\begin{bmatrix} 0.1 & 0.05 & 0.03 & 0 \\ 0.05 & 0.09 & 0.05 & 0 \\ 0.03 & 0.05 & 0.08 & 0.02 \\ 0 & 0 & 0.02 & 0.05 \end{bmatrix} g_0(\mathbf{h})$$

$$\begin{bmatrix} 0.27 & 0.17 & 0.03 & 0.19 \\ 0.17 & 0.51 & 0 & 0.17 \\ 0.03 & 0 & 0.1 & 0.04 \\ 0.19 & 0.17 & 0.04 & 0.4 \end{bmatrix} Sph\left(\frac{|\mathbf{h}|}{75}\right)$$

$$\begin{bmatrix} 0.63 & 0.5 & 0.3 & 0.45 \\ 0.5 & 0.42 & 0.18 & 0.4 \\ 0.3 & 0.18 & 0.08 & 0.12 \\ 0.45 & 0.4 & 0.12 & 0.55 \end{bmatrix} Sph\left(\frac{\mathbf{h}'}{200}\right)$$

$$where\ \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.7 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.7 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figures 4.10 and 4.11 display the direct and cross semivariograms fitted with this (permissible) linear model of coregionalisation.

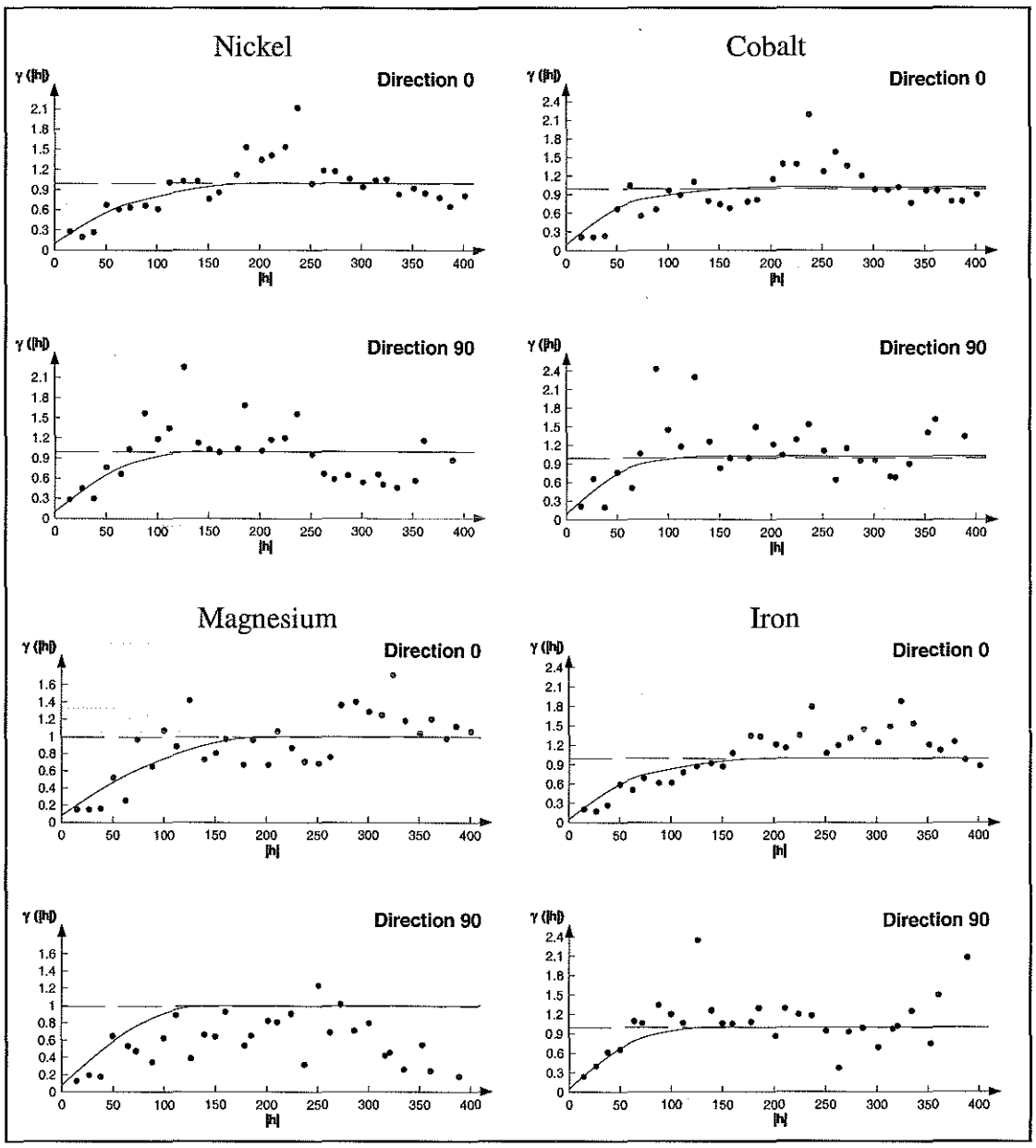


Figure 4.10: Fitted direct semivariograms for MM22DTOP4

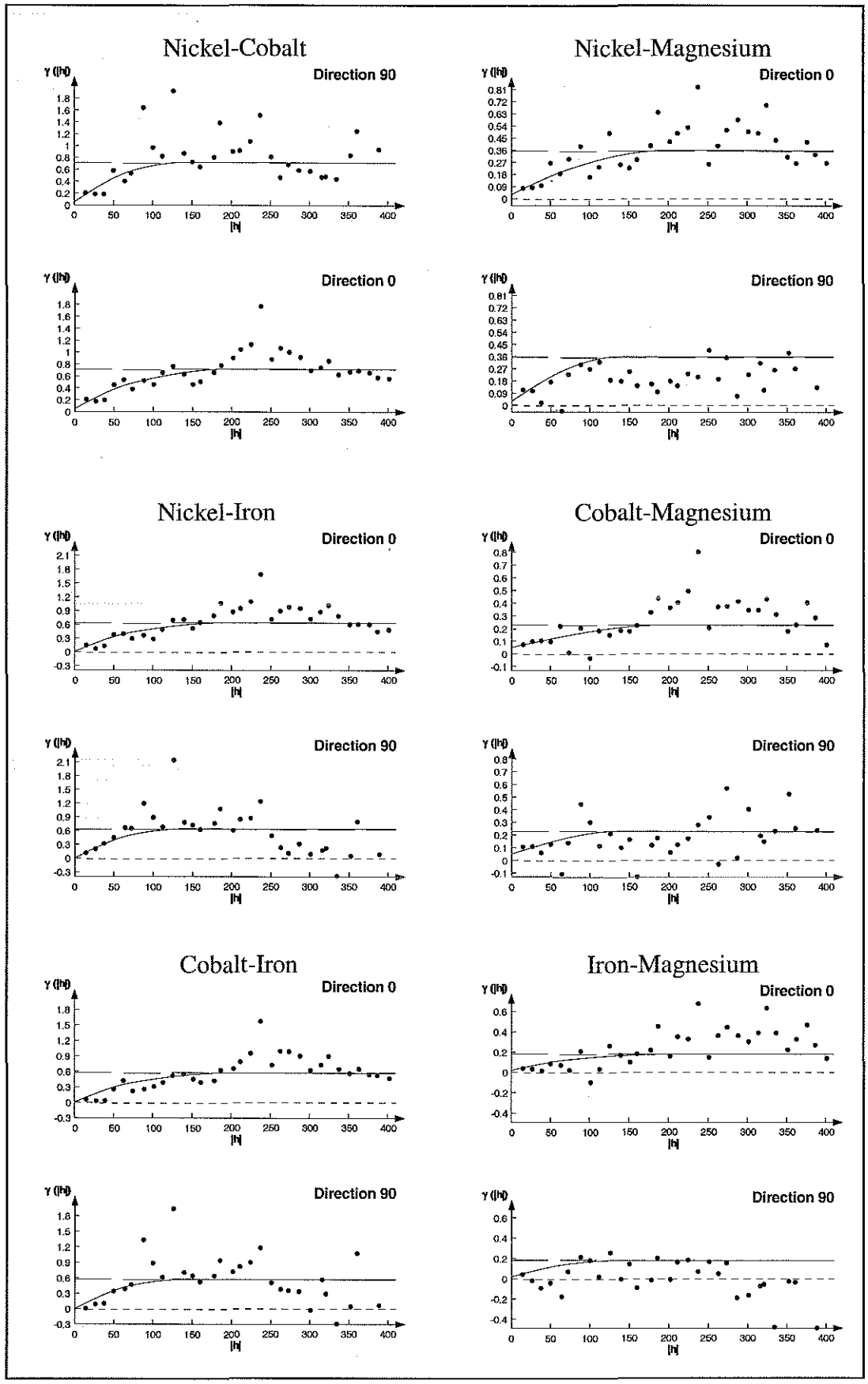


Figure 4.11: Fitted cross semivariograms for MM22DTOP4

4.2.2 Cross Validation for Cokriging of MM22DTOP4

The cross validation was carried out for all of the variables in this data set using ordinary cokriging for the estimation. We present here the results for nickel, those for cobalt, magnesium and iron are presented in Appendix B. Several neighbourhood parameters were investigated with little difference in results. Figure 4.13 shows the cross validation residual analysis for the search parameters displayed in Table 4.4 and Figure 4.12. Note that the minimum number of points in this case was set to three. While this may appear to be too small, however, there was a group of nine locations at the extreme northern perimeter of the study region that could not be estimated using a minimum of four points. An increase in the size of the search neighbourhood was considered. However as the affected points were outside the region of sample data, hence extrapolated, it was deemed more appropriate to reduce the number of points to three. In this way the extreme locations could be estimated without affecting the satisfactory performance of the search parameters for the remainder of the region.

The cross validation residuals appear to be roughly normally distributed and the plot of estimates versus the true values is well clustered around the 45° bisector, with only two obvious outliers. The mean errors displayed in Table 4.5 are sufficiently close to zero and the standardised error variance is sufficiently close to the ideal value of one respectively. Overall the model appears to have reproduced the sample values successfully for all variables.

Table 4.4: Neighbourhood search parameters

Number of angular sectors	4
Minimum number of samples	3
Optimal number of samples per sector	4
Search radius (major axis of ellipse) (metres)	220
Search radius (minor axis of ellipse) (metres)	100

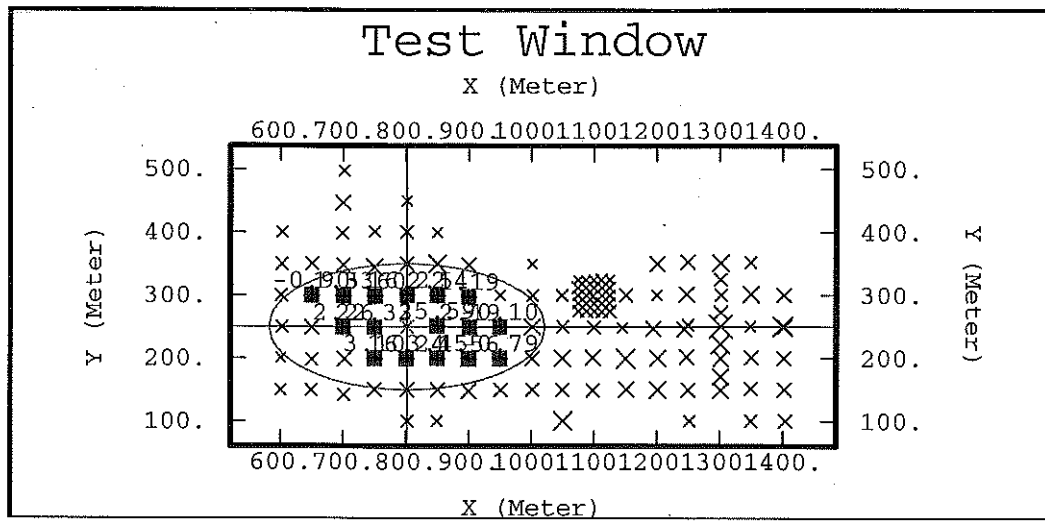


Figure 4.12: Search neighbourhood for MM22DTOP4

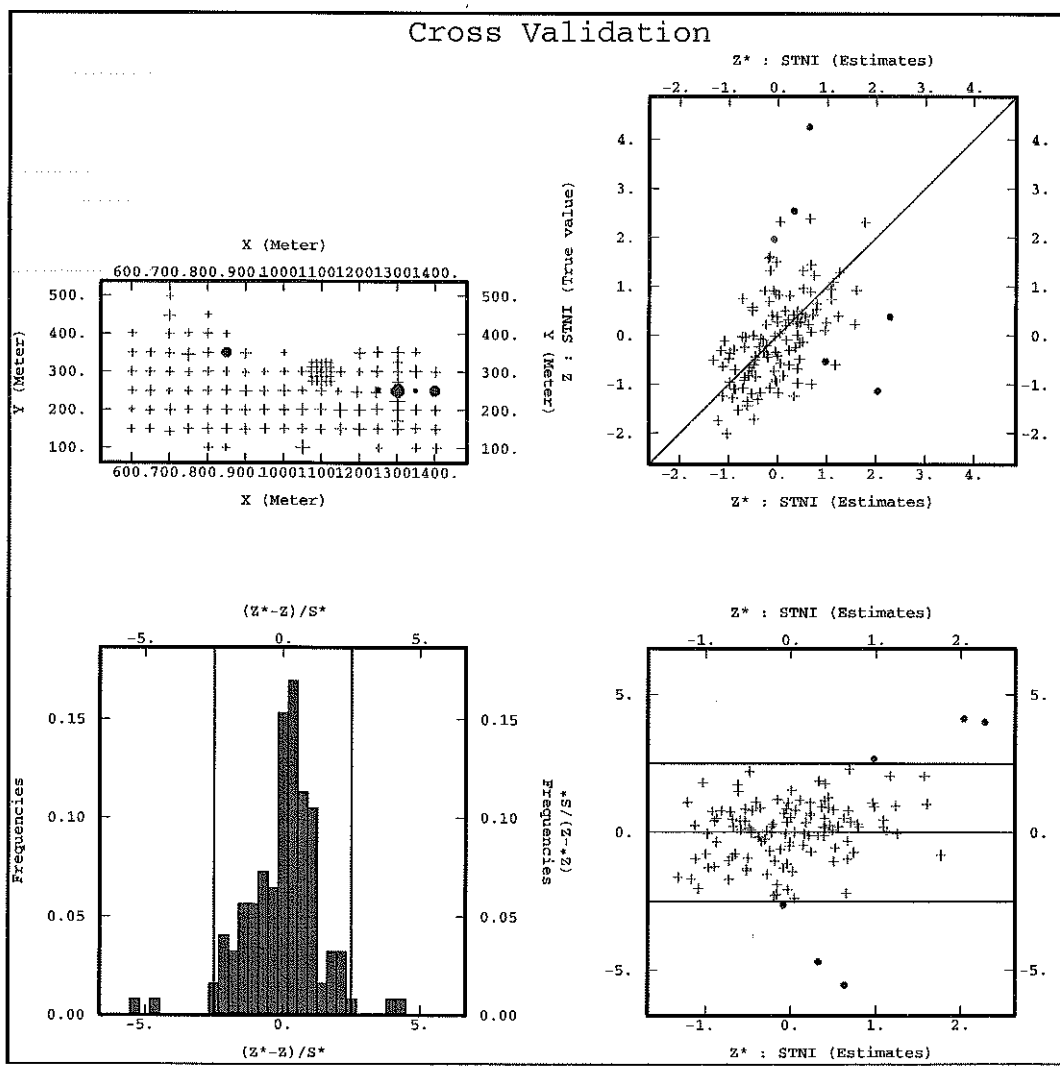


Figure 4.13: Cross validation residual analysis for MM22DTOP4

Table 4.5: Mean and variance of cross validation residuals

	Nickel		Cobalt		Magnesium		Iron	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Error	0.02	0.87	0.02	1.15	-0.02	0.76	0.01	0.64
St. Error	0.02	1.85	0.02	1.96	-0.02	1.48	0.01	1.45

4.2.3 Cokriging Estimates of MM22DTOP4

The ordinary cokriging of the nickel, cobalt, magnesium and iron variables was performed using the linear model of coregionalisation discussed earlier. The estimates were calculated at each of the 1717 grade control sample locations using the search neighbourhood parameters displayed in Table 4.4. The parameter files for this cokriging procedure are located in Appendix C. As previously, in section 4.1.3, it was necessary to transform the estimates into their unstandardised form by multiplying each value by the sample standard deviation and adding to this the sample mean. Figure 4.14 shows post plots of the estimates for each of the variables along with the post plots of the grade control data. Once again the smoothing nature of the cokriging algorithm is clearly apparent. The estimates of nickel, cobalt and iron appear to be similar to those obtained from the cokriging using the intrinsic coregionalisation model. As the discussion of these variables is analogous to that in section 4.2.3 we omit it here.

The overall appearance of the post plots of the magnesium estimates is consistent with the grade control values. The estimates along the north-western, western and eastern perimeters are representative of the grade control values. The values along the southern perimeter of the study region have been poorly estimated. While the areas of high magnesium concentrations have been well represented the concentrations of lower values have not. This is a result of an under representation of low values in the southern most line of exploration drill holes.

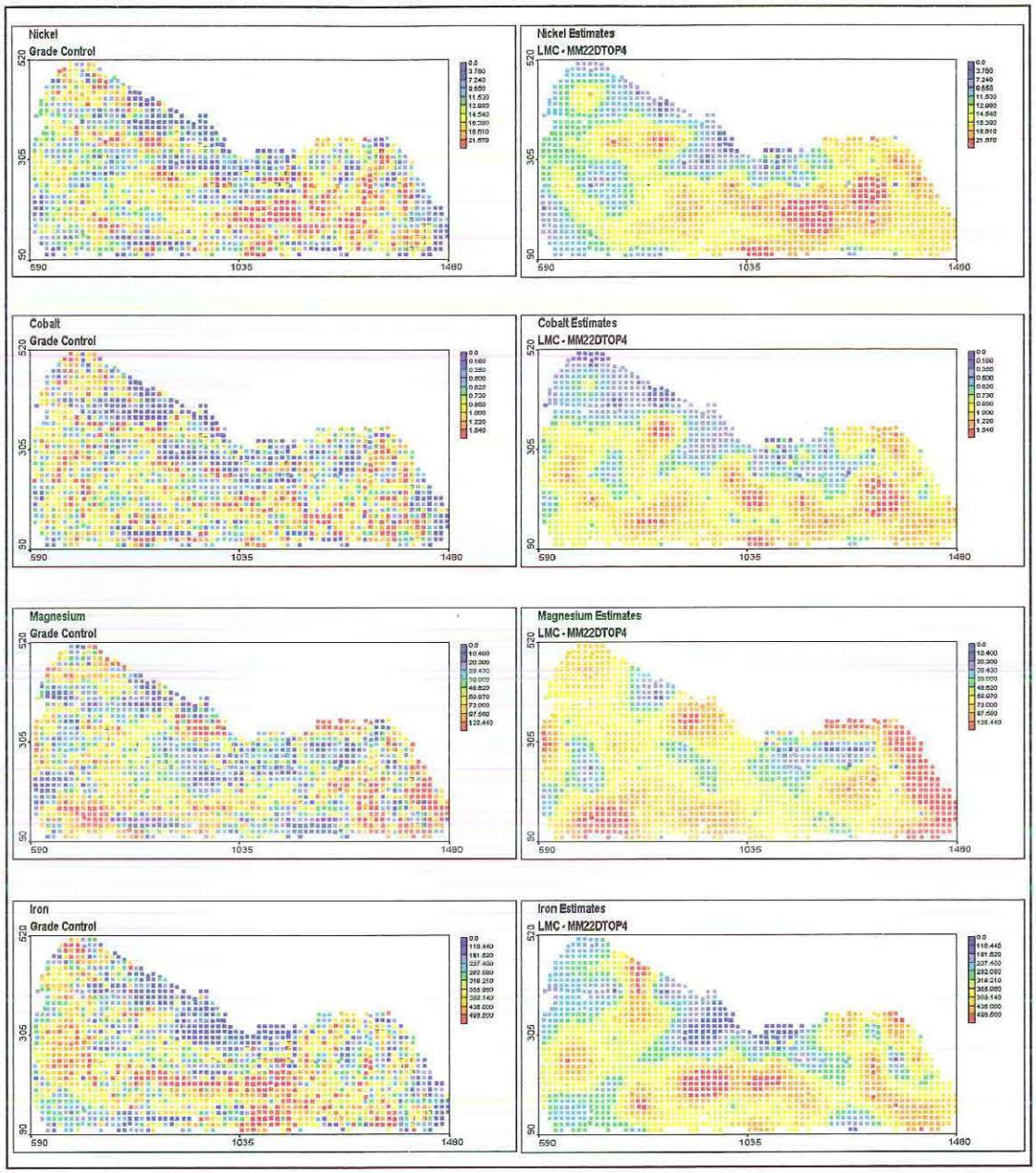


Figure 4.14: Cokriged estimates – Linear model of Coregionalisation MM22DTOP4

The residual post plots for nickel, cobalt and iron are displayed in Figure 4.15 and are similar to those obtained in the previous section. The residual plot of magnesium shows a small pocket of overestimated values at the eastern edge of the study region and a small pocket of underestimated values at the south-eastern perimeter. Otherwise the residuals are of fairly small magnitude across the entire study region. The histograms and normal score plots of residuals for nickel, iron and zinc are similar to those presented in the previous section. Those for magnesium are displayed in Figure 4.16. The histogram of residuals is reasonably symmetric and

looks approximately normally distributed. The normal scores plot however deviates from a straight line along the middle section.

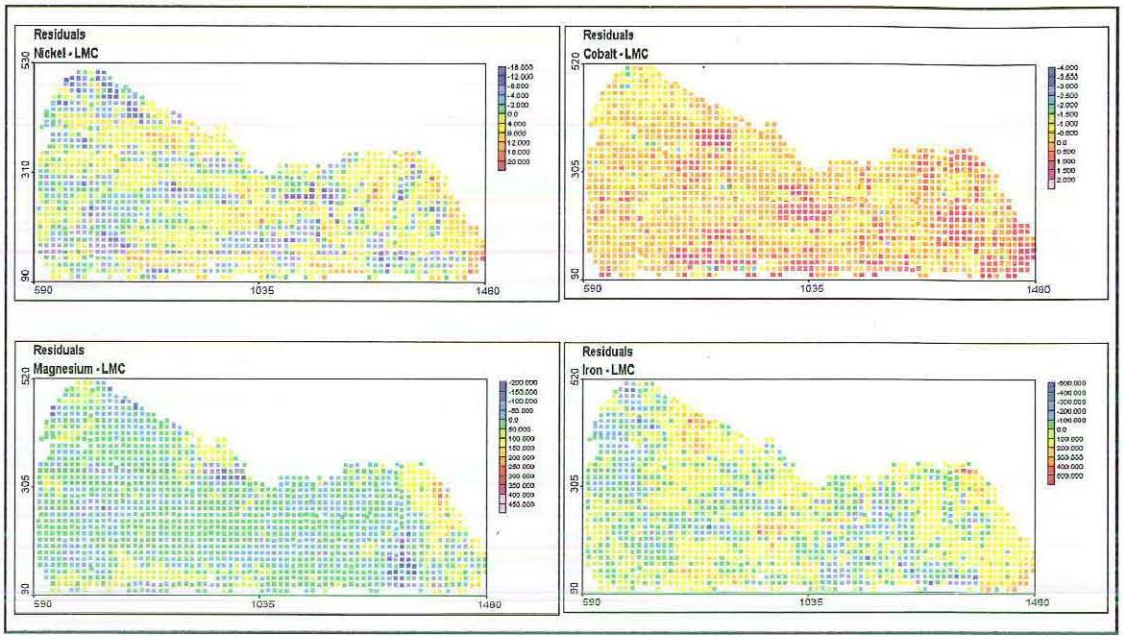


Figure 4.15: Residual post plots for MM22DTOP4

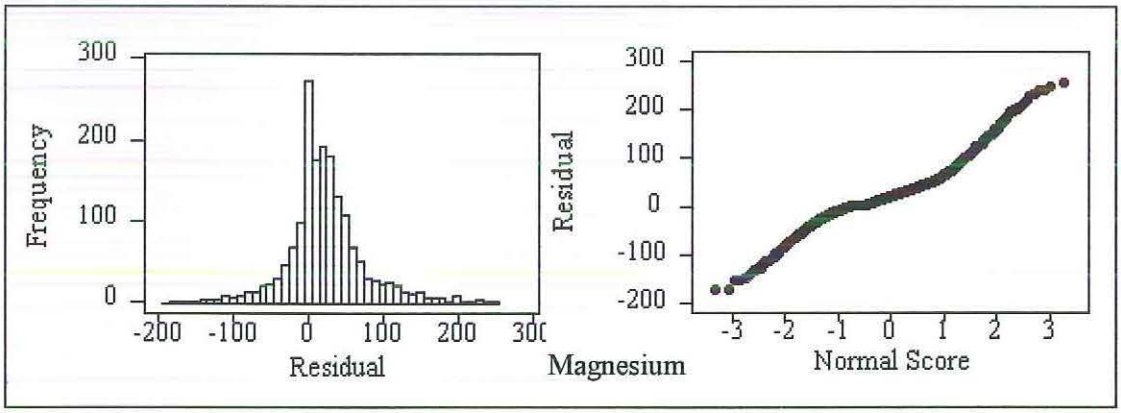


Figure 4.16: Residual histogram and normal score plot for magnesium

5 PRINCIPAL COMPONENT KRIGING

In this chapter we demonstrate the estimation technique of principal component kriging using the principal components extracted in the eigenanalysis of the MM22DHC4 data set. As the orthogonality of the principal components extends to any separation vector **h** we were able to individually model them. All of the semivariograms were calculated with 32 lags at a lag spacing of 12.5 metres and a lag tolerance of 6.25 metres. The directional semivariograms were calculated with an angular tolerance of 45°. All directional models were assessed for fit in the intermediate directions (with lag tolerance set to 22.5°).

Once the spatial variability of the principal component scores had been modelled the estimates at unsampled locations were calculated using ordinary kriging. All models were cross validated using ordinary kriging and performed satisfactorily. Once again the software package used to perform the kriging was ISATIS. 1717 estimates were obtained for each principal component directly at the locations of the grade control data. The parameter files for all kriging procedures are displayed in Appendix C. Having obtained the estimates of the principal component scores it was necessary to recalculate the nickel, cobalt, iron and zinc estimates using the coefficients obtained in the eigenanalysis of the MM22DHC4 data set. These estimates were then compared to the MM22DGC grade control data.

5.1 Variography of the Principal Components from MM22DHC4

For the first principal component (PC1) the semivariogram surface exhibits anisotropy with the direction of maximum continuity being east-west. This is further supported by the standardised semivariogram surface also displayed in Figure 5.1. The final model chosen consists of three structures, a nugget effect and two spherical structures. The first spherical structure is isotropic with a range of 90 metres. The second spherical structure is anisotropic and is modelled with east-west as the major direction with a range of 200 metres and with an anisotropy ratio of 0.45. This model can be expressed as:

$$\gamma(\mathbf{h}) = 0.1 g_0(\mathbf{h}) + 0.47 Sph\left(\frac{|\mathbf{h}|}{90}\right) + 2.47 Sph\left(\frac{\mathbf{h}'}{200}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.45 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.45 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 5.2 shows the directional experimental semivariograms fitted with this linear model of regionalisation.

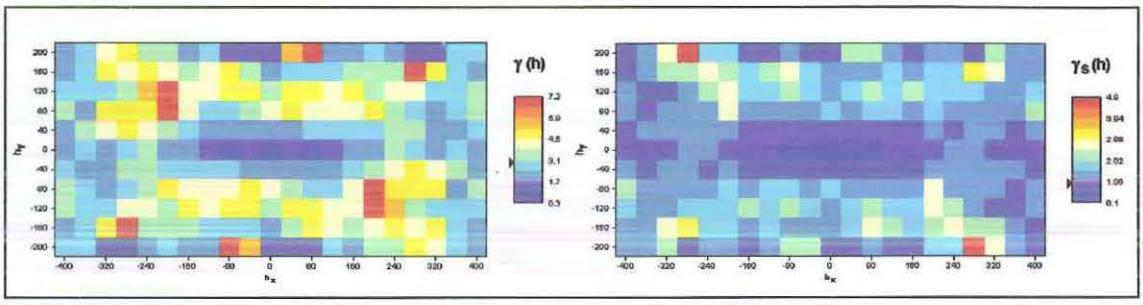


Figure 5.1: Semivariogram and standardised semivariogram surfaces for PC1

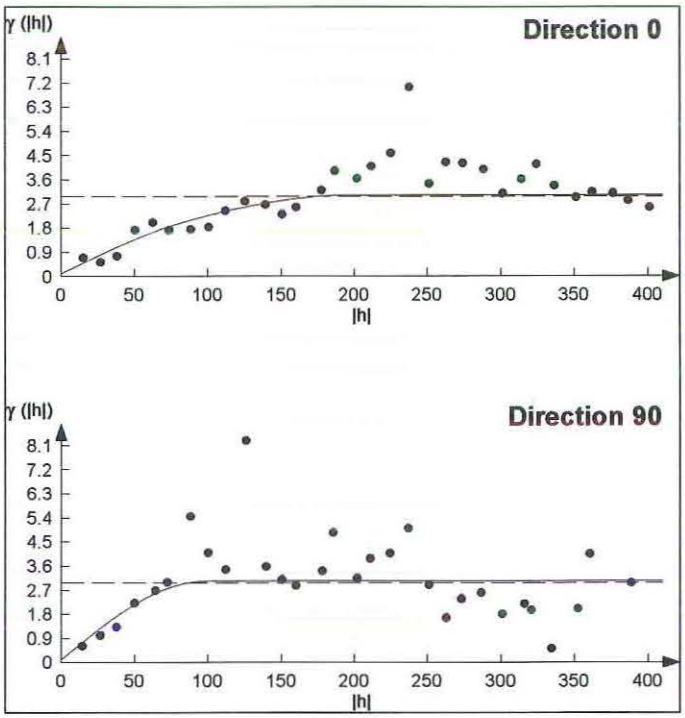


Figure 5.2: Fitted directional semivariograms for PC1

Figure 5.3 displays the semivariogram surface for the second principal component (PC2). The surface does not exhibit any anisotropy which is further supported by the standardised semivariogram surface also displayed in Figure 5.3. The final model chosen consists of two structures, a nugget effect and an isotropic spherical structure with a range of 85 metres. This model can be expressed as:

$$\gamma(\mathbf{h}) = 0.08 g_0(\mathbf{h}) + 0.47 Sph\left(\frac{|\mathbf{h}|}{85}\right)$$

Figure 5.4 shows the omnidirectional experimental semivariogram fitted with this linear model of regionalisation.

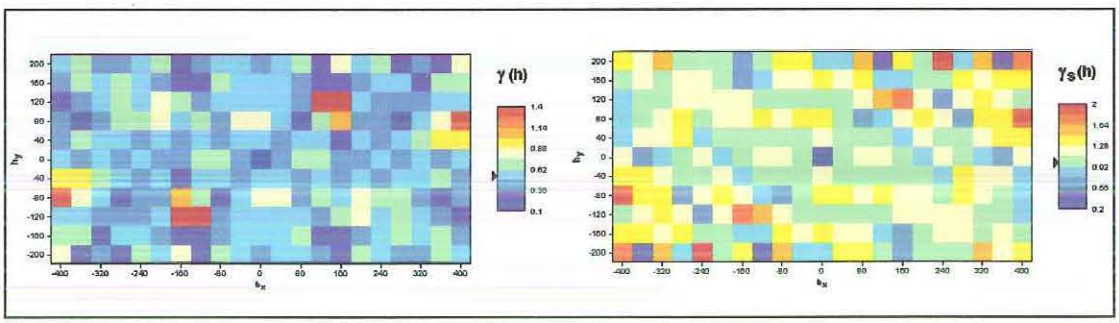


Figure 5.3: Semivariogram surface for PC2

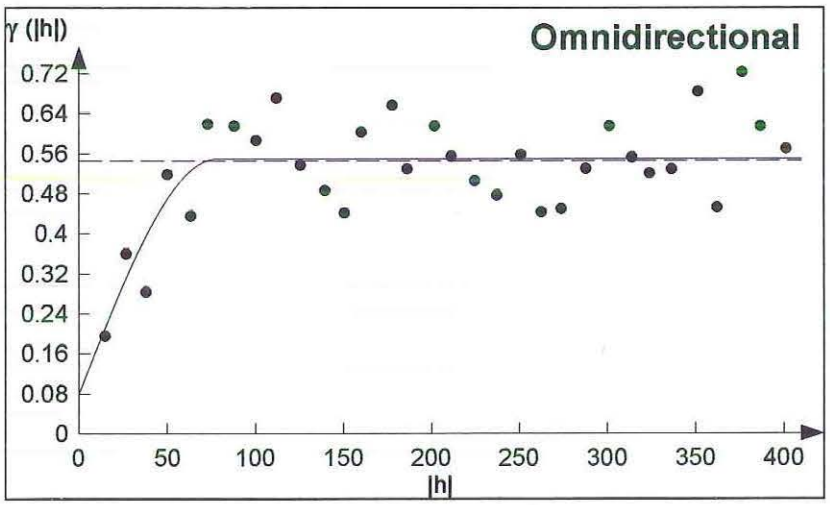


Figure 5.4: Fitted omnidirectional semivariogram for PC2

Figure 5.5 displays the semivariogram surface for the third principal component (PC3). This component exhibits anisotropy with the direction of maximum continuity at azimuth angle 65°.

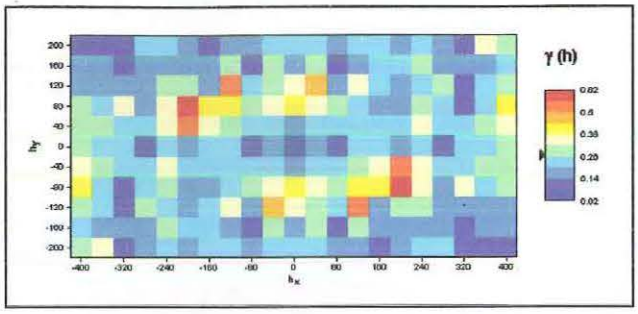


Figure 5.5: Semivariogram surface for PC3

Figure 5.6 displays the directional semivariograms fitted with the final model. It consists of three structures, a nugget effect and 2 spherical structures. The first spherical structure is omnidirectional with a range of 50 metres; the second is modelled in the direction of maximum continuity (azimuth 65°) with a range of 140 metres and an anisotropy ratio of 0.7. This model can be expressed as:

$$\gamma(\mathbf{h}) = 0.01 g_0(\mathbf{h}) + 0.05 Sph\left(\frac{|\mathbf{h}|}{50}\right) + 0.215 Sph\left(\frac{\mathbf{h}'}{140}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.7 \end{bmatrix} \begin{bmatrix} cos65° & sin65° \\ -sin65° & cos65° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0.42 & 0.91 \\ -0.63 & 0.30 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

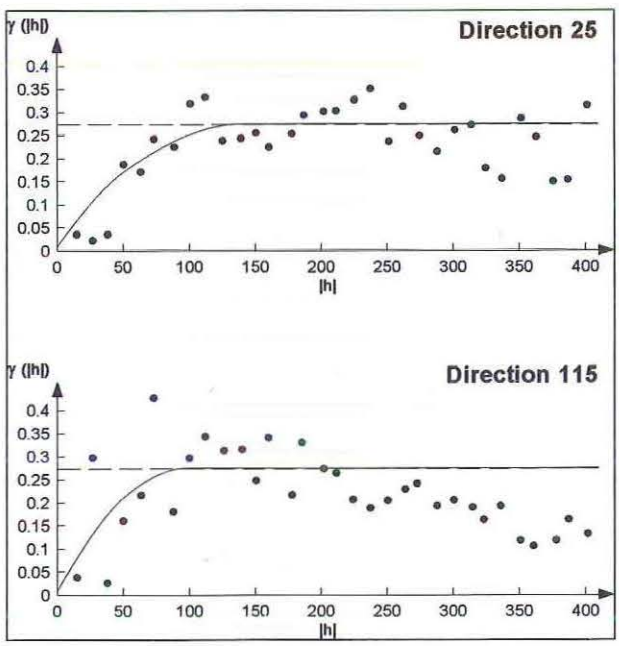


Figure 5.6: Fitted directional semivariograms for PC3

The semivariogram surface of the fourth principal component (PC4) displayed in Figure 5.7 exhibits both a short range isotropy and a long range anisotropy. From the semivariogram surface the direction of maximum continuity appears to be at azimuth angle 65°, whereas the standardised semivariogram surface suggests north-south (azimuth 0°). Investigation of several directional semivariograms, such as those displayed in Figure 5.8, revealed that the anisotropy in various directions was not sufficiently marked, hence it was possible to treat the data as being isotropic.

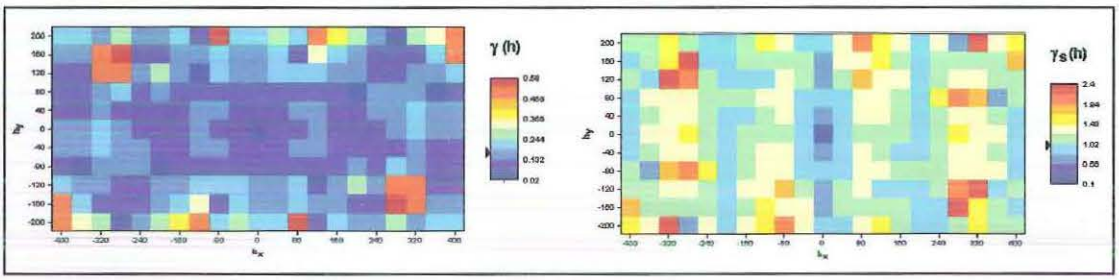


Figure 5.7: Semivariogram and standardised semivariogram surfaces for PC4

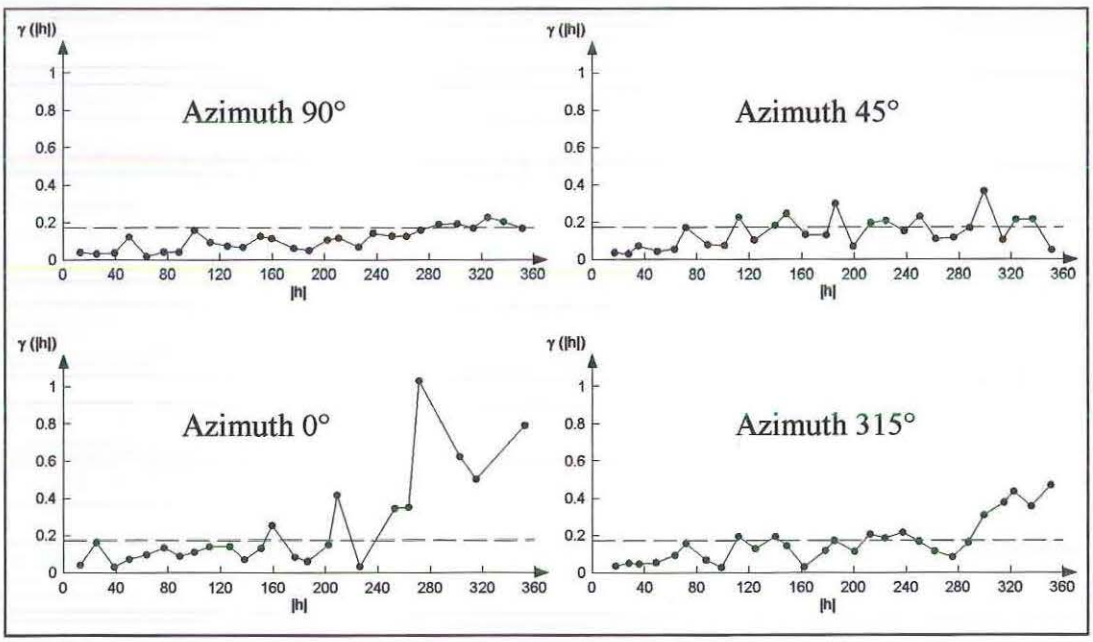


Figure 5.8: Directional semivariograms for PC4

The omnidirectional model chosen is displayed in Figure 5.9 and consists of three structures, a nugget effect and two isotropic spherical structures. The first spherical

structure has a range of 100 metres and the second has a range of 250 metres. This model can be expressed as:

$$\gamma(\mathbf{h}) = 0.018 g_0(\mathbf{h}) + 0.065 Sph\left(\frac{|\mathbf{h}|}{100}\right) + 0.092 Sph\left(\frac{|\mathbf{h}|}{250}\right)$$

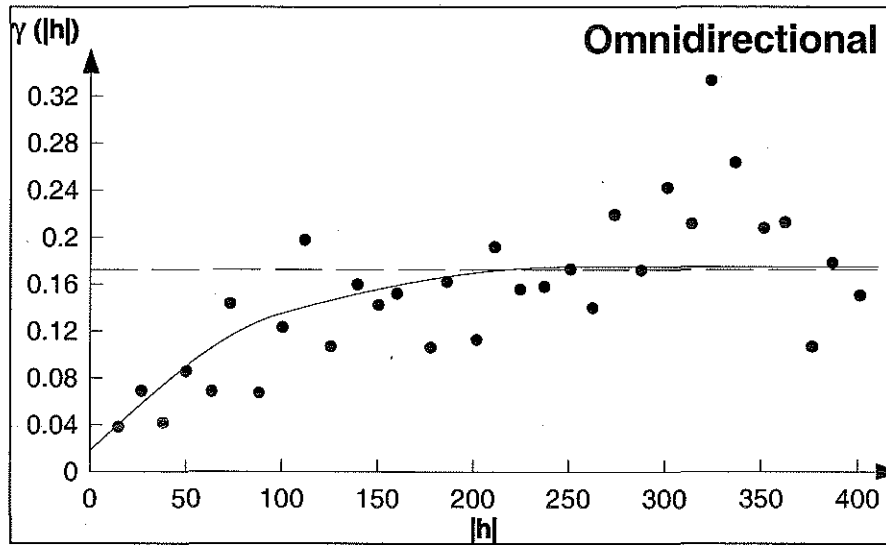


Figure 5.9: Fitted omnidirectional semivariogram for PC4

5.2 Cross validation of the Principal Components

Cross validation was carried out individually using ordinary kriging for each of the principal components. Various search neighbourhoods were investigated with little difference in the results. The final search neighbourhoods used are displayed in Table 5.1. We present here the results for the first principal component, the remaining cross validation results for PC2, PC3 and PC4 are displayed in Appendix B. Figure 5.10 displays the search neighbourhood for the first principal component and Figure 5.11 displays the residual analysis of the cross validation. The residual histogram shows that the errors are roughly normally distributed and the estimates versus true value plot is reasonably close to the 45° line with the exception of a few outliers. In all cases the model provided successful estimates. Table 5.2 shows the mean and variance of the errors and the standardised errors. The mean errors are sufficiently close to zero and the standardised error variances are sufficiently close to one.

Table 5.1: Neighbourhood search parameters

	PC1	PC2	PC3	PC4
Number of angular sectors	4	4	4	4
Minimum number of samples	3	3	3	3
Optimal number of samples per sector	4	4	4	4
Search radius (major axis of ellipse) (metres)	220	120	190	250
Search radius (minor axis of ellipse) (metres)	110	120	100	250
Rotation (azimuth)	90°	NA	65°	NA

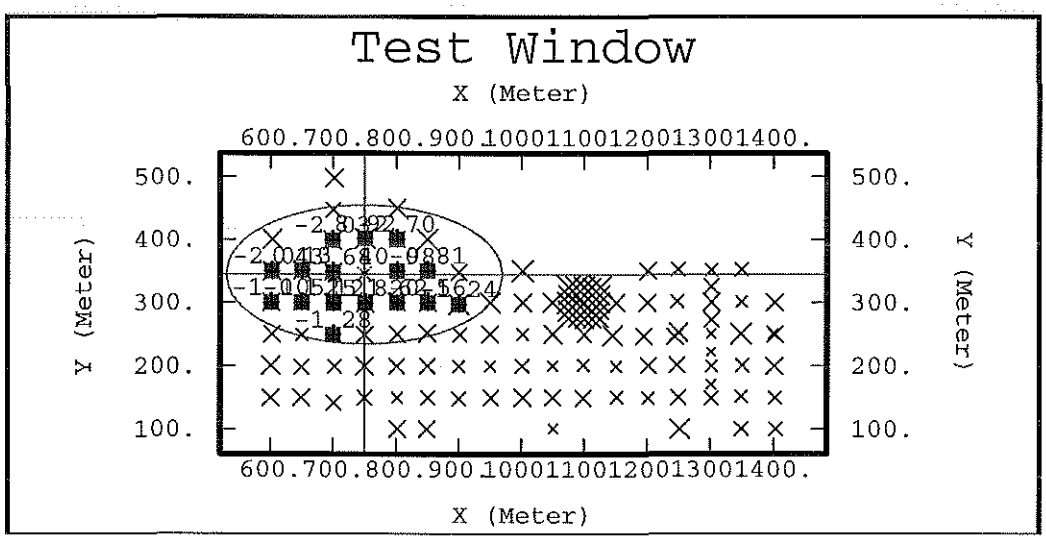


Figure5.10: Search neighbourhood for PC1

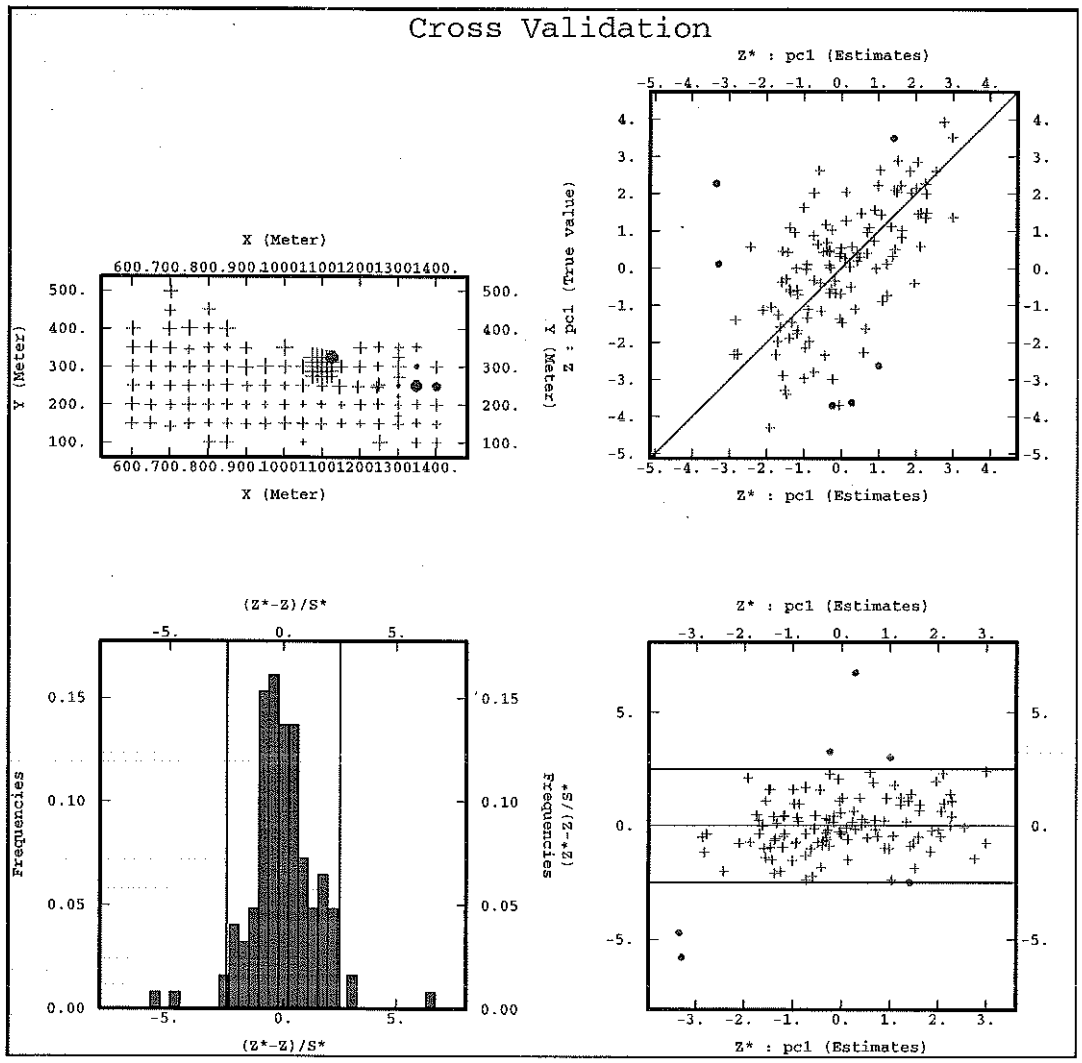


Figure 5.11: Cross validation residual analysis for PC1

Table 5.2: Mean and variance of cross validation residuals

	PC1		PC2		PC3		PC4	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Error	-0.04	2.27	0.00	0.70	0.01	0.34	0.00	0.15
St. Error	-0.02	2.20	0.01	1.74	0.02	2.35	0.01	1.85

5.3 Principal Component Kriging Estimates

The ordinary kriging of each of the principal components was performed at the 1717 grade control data locations using the neighbourhood search parameters displayed in Table 5.1. The parameter files used are displayed in Appendix C. Having obtained the estimates of the principal component scores it was necessary to calculate the individual variable estimates using equation (60) which we recall is

$$z_{PCK}^{(i)*}(\mathbf{u}) = \sum_{k=1}^{K} a_{ki} y_{OK}^{(k)*}(\mathbf{u})\sigma_i + m_i$$

where σ_i and m_i are the sample standard deviation and mean respectively of the ith variable. The coefficients a_{ki} are obtained from the transpose of the matrix of eigenvectors calculated in the principal component analysis of MM22DHC4.

$$\mathbf{A}=[a_{ki}]=\mathbf{Q}^{\mathrm{T}}=\begin{bmatrix} -0.50 & -0.48 & -0.51 & -0.51 \\ -0.42 & -0.60 & 0.50 & 0.47 \\ 0.76 & -0.64 & -0.02 & -0.12 \\ 0.04 & -0.06 & -0.70 & 0.71 \end{bmatrix}$$

These calculations were performed using an EXCEL spreadsheet.

Recall from section 3.4.2 we determined that if data reduction were the aim of performing a principal component analysis on the MM22DHC4 data set that we would retain only the first two principal components. These two components accounted for 89% of the total variability of the original data. In addition to calculating the estimates of nickel, cobalt, iron and zinc using all four principal components we also estimated these variables by retaining only the first two principal components. The estimates of the variables then were obtained by discarding the estimates of the scores of both the third and fourth principal components and the last two columns of the matrix of eigenvectors **Q**. The estimates from only the retained principal components are then calculated using equation (61) which we recall is

$$z_{PCKP}^{(i)*} = \sum_{p=1}^{P} a_{pi} y_{OK}^{(p)*}(\mathbf{u})\sigma_i + m$$

where σ_i and m_i are the sample standard deviation and mean respectively of the ith variable and P is the number of principal components retained. The coefficients a_{pi} for the truncated estimates are:

$$\mathbf{A}_P = \left[a_{pi} \right] = \begin{bmatrix} -0.50 & -0.51 & -0.51 & -0.51 \\ -0.42 & -0.60 & 0.50 & 0.47 \end{bmatrix}$$

Figures 5.12 shows the post plots of the nickel, cobalt, iron and zinc grade control data and principal component kriging estimates (all four principal components). Overall the estimates have reproduced the grade control samples well, once again though the smoothing nature of the kriging process is apparent. The same areas of over and underestimation are apparent as for the previous kriging methods. In fact the post plots are almost identical to those we have already seen. Figure 5.13 shows the post plots of the grade control data and the estimates obtained using only the first two principal components. Clearly the grade control data have been reasonably well reproduced with very little difference to those estimates obtained using all four principal components. The histograms and normal score plots for the residuals from both types of principal component kriging are very similar to those presented earlier, hence are omitted here. Figures 5.14 and 5.15 display the residual post plots for both types of principal component kriging. Both sets of plots are once again very similar to those of the other kriging methods.

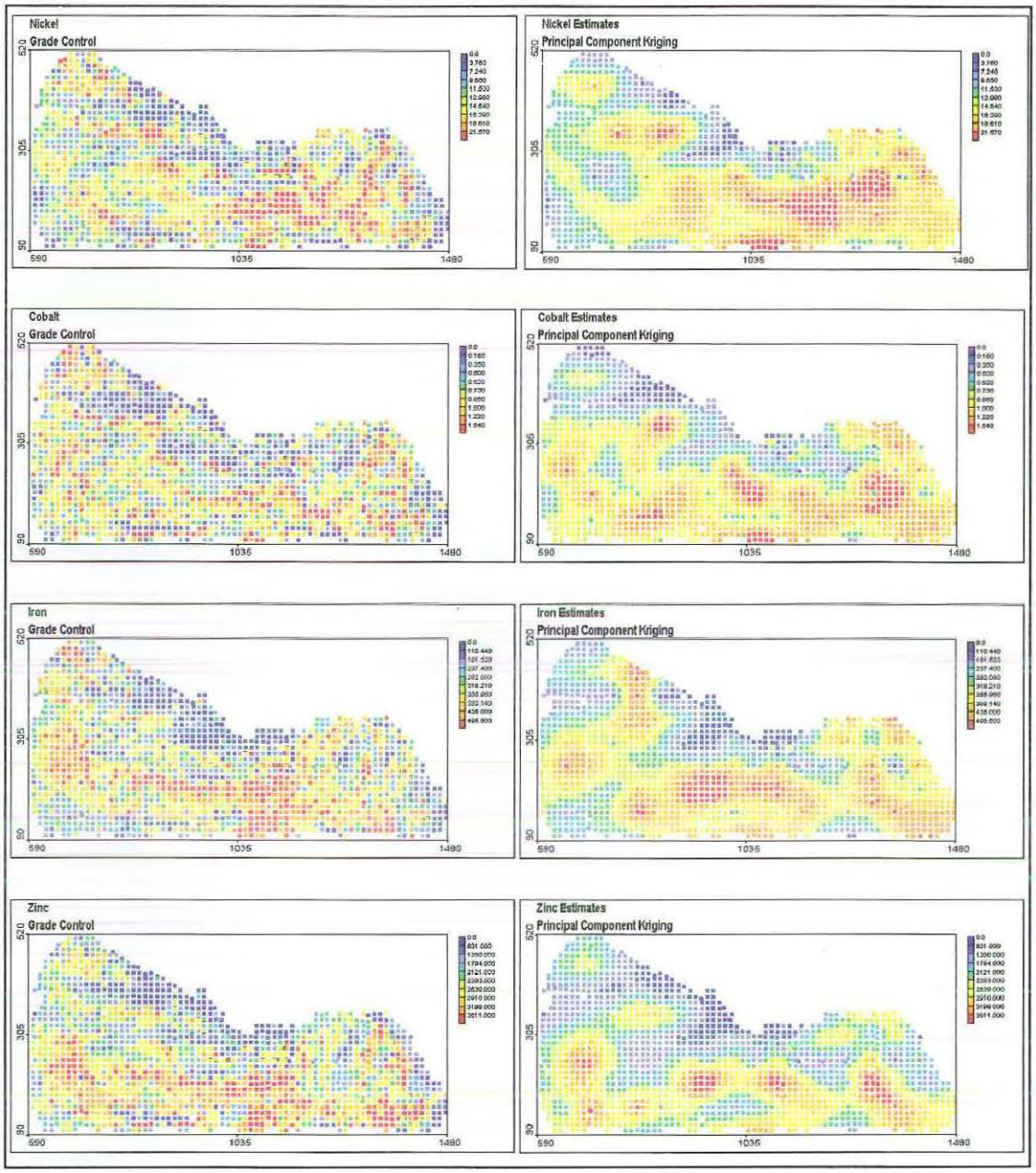


Figure 5.12: Principal component kriging estimates

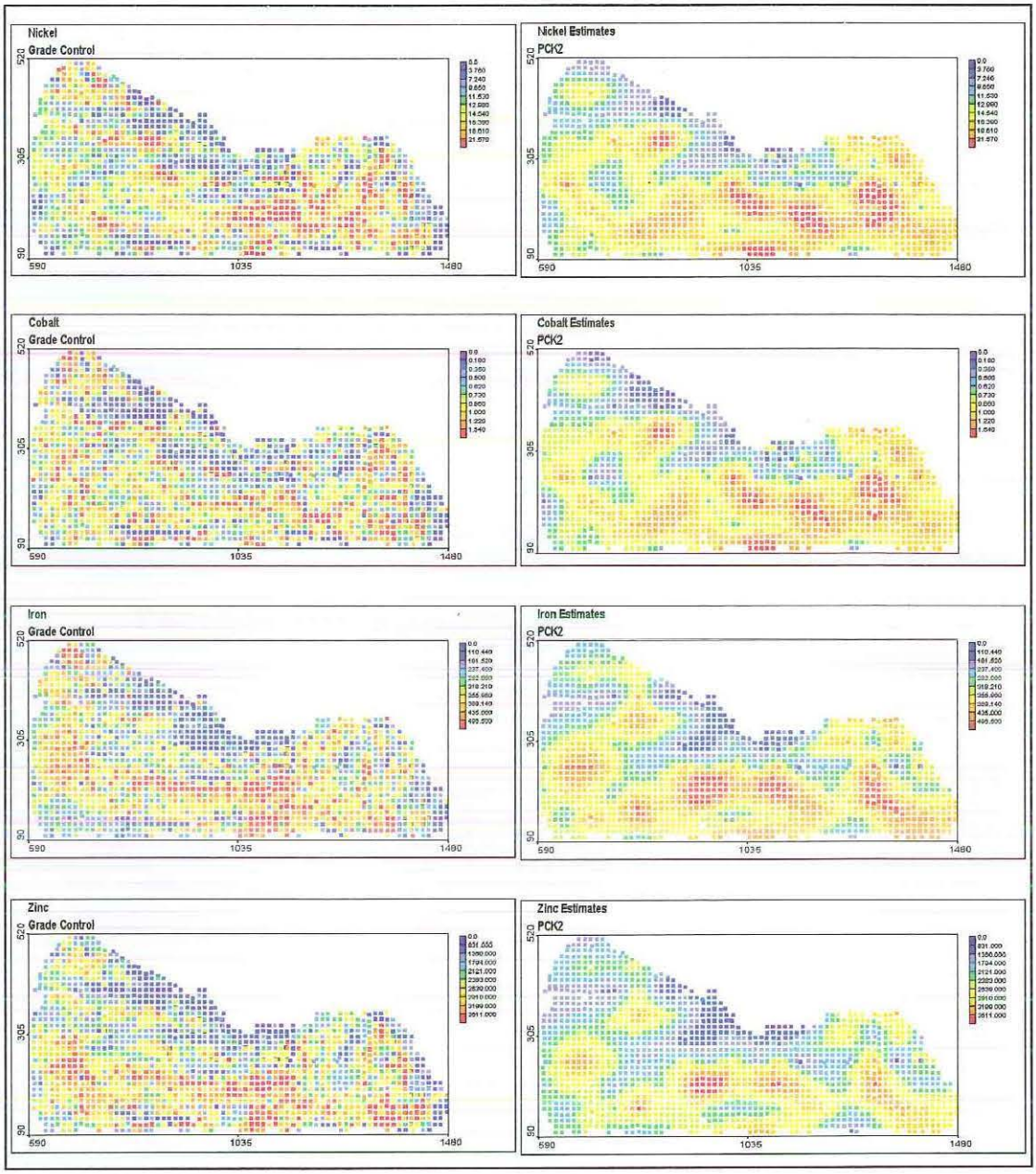


Figure 5.13: Principal component kriging estimates using two retained principal components

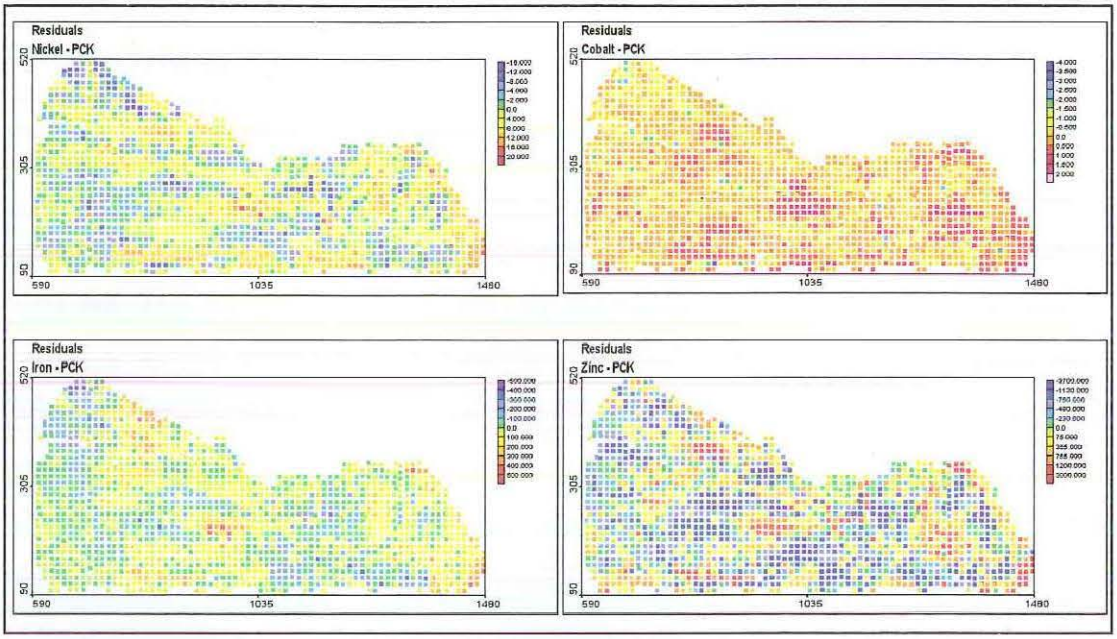


Figure 5.14: Residual post plots for principal component kriging

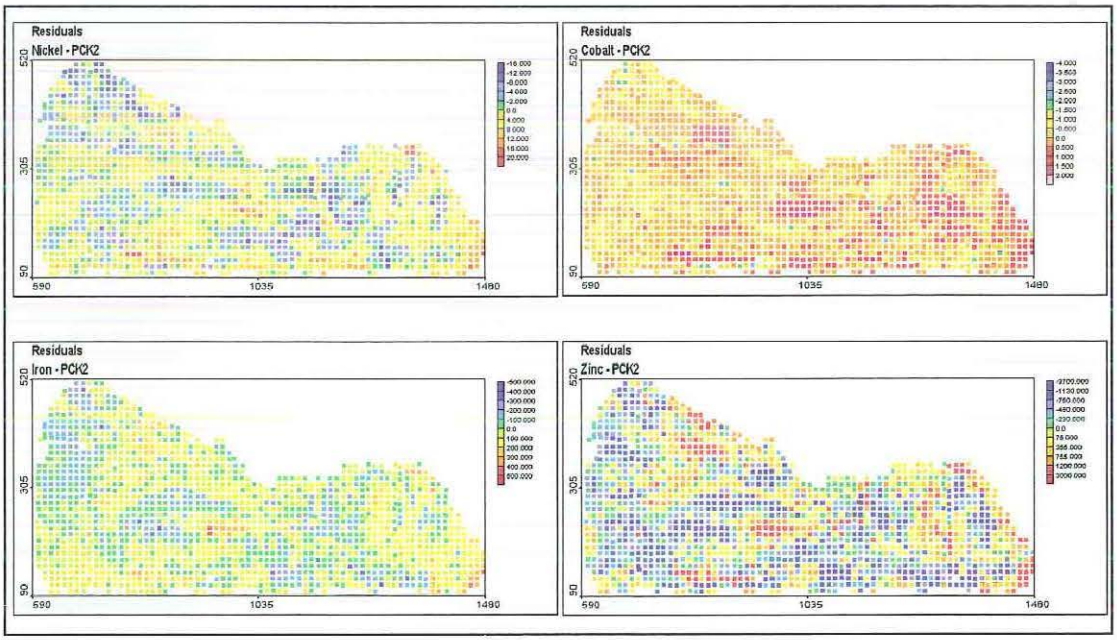


Figure 5.15: Residual post plots for principal component kriging using two retained principal components

6 Ordinary Kriging

In this chapter we demonstrate the estimation technique of ordinary kriging using the variables nickel, cobalt, magnesium, iron and zinc from the MM22DEXP data set. The data were left in their raw (unstandardised) form for this section. In addition to the grade control data for comparison the estimates obtained from ordinary kriging were used to assess whether the multivariate techniques enhanced the estimation, the discussion of which we defer until the next chapter. For the variography and modelling the same parameters were used as for the previous methods, that is, all semivariograms were calculated with 32 lags at a lag spacing of 12.5 metres and a lag tolerance of 6.25 metres. The directional semivariograms were calculated with an angular tolerance of 45.

The cross validation using ordinary kriging for each of the variables showed that the models all performed satisfactorily. As previously ISATIS was used for the estimation. Again 1717 estimates were obtained for each variable directly at the locations of the grade control data. The ordinary kriging parameter files are also displayed in Appendix C. The accuracy of the ordinary kriging estimates was assessed via comparison with the MM22DGC grade control data.

6.1 Variography of Nickel, Cobalt, Magnesium, Iron and Zinc

The semivariogram surface for nickel is displayed in Figure 6.1. As expected it displays the same anisotropies identified earlier in Chapter 4, that is the presence of two structures, one short range and one long range. The direction of maximum continuity of the short range structure is east-west (azimuth 90°), which is confirmed by the standardised semivariogram surface, also displayed in Figure 6.1. The direction of maximum continuity of the long range structure is approximately azimuth 70°. Detailed examination of the directional experimental semivariograms, including those shown if Figure 6.2, revealed that the anisotropy is most pronounced in the east-west (major) and north-south (minor) directions.

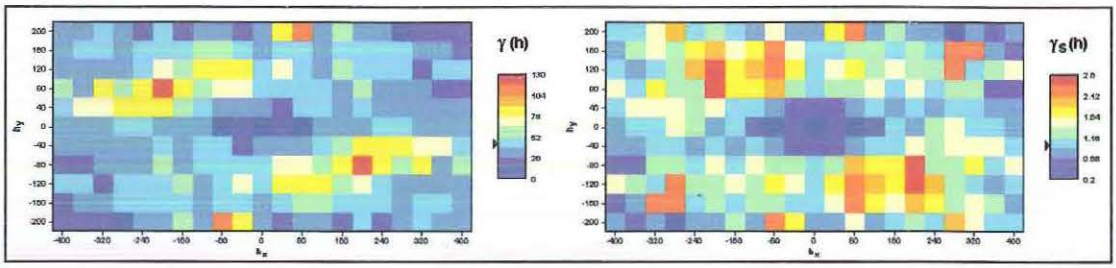


Figure 6.1: Semivariogram and standardised semivariogram surfaces for nickel

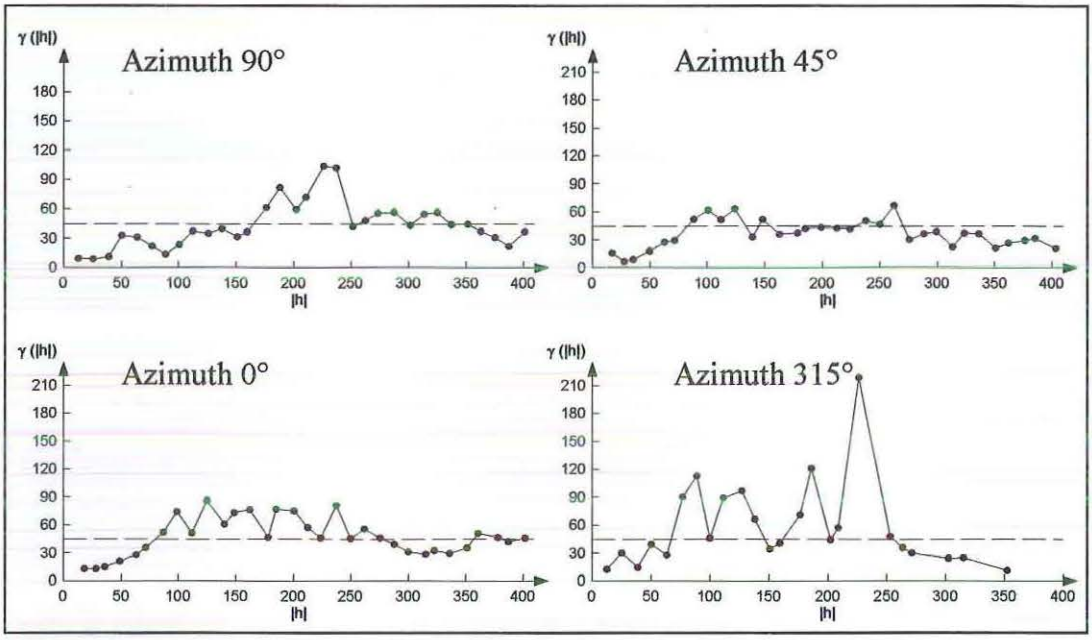


Figure 6.2: Directional semivariograms for nickel

The model selected consists of a nugget effect and two spherical structures. The first spherical structure is isotropic with a range of 90 metres. The second spherical structure is modelled with east-west as the major direction with a range of 160 metres and an anisotropy ratio of 0.56. This model can be expressed as:

$$\gamma(\mathbf{h}) = 5g_0(\mathbf{h}) + 12.15Sph\left(\frac{|\mathbf{h}|}{90}\right) + 28.35Sph\left(\frac{\mathbf{h}'}{160}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.56 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \cdot \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.56 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 6.3 shows the directional semivariograms fitted with this model.

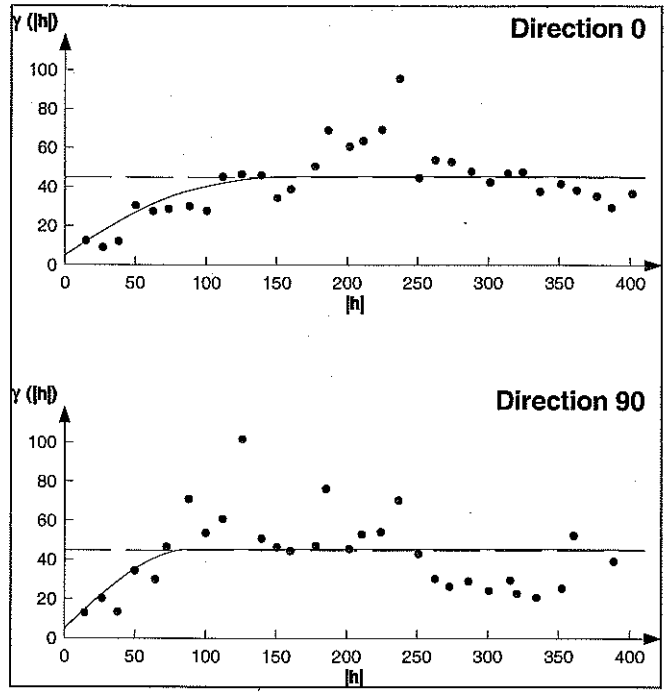


Figure 6.3: Fitted nickel directional semivariograms

The semivariogram and standardised semivariogram surfaces for cobalt are displayed in Figure 6.4. The direction of maximum continuity of the anisotropy clearly looks to be in the east-west direction. The model for cobalt consists of a nugget effect and two spherical structures. The first spherical structure is isotropic with a range of 100 metres. The second spherical structure is modelled with east-west as the major direction with a range of 200 metres and an anisotropy ratio of 0.5. This model can be expressed as:

$$\gamma(\mathbf{h}) = 0.012 g_0(\mathbf{h}) + 0.16 Sph\left(\frac{|\mathbf{h}|}{100}\right) + 0.13 Sph\left(\frac{\mathbf{h}'}{200}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.5 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 6.5 shows the directional semivariograms fitted with this model.

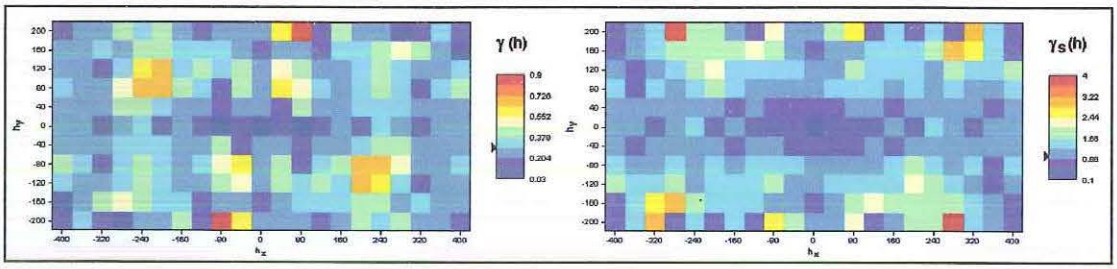


Figure 6.4: Semivariogram and standardised semivariogram surfaces for cobalt

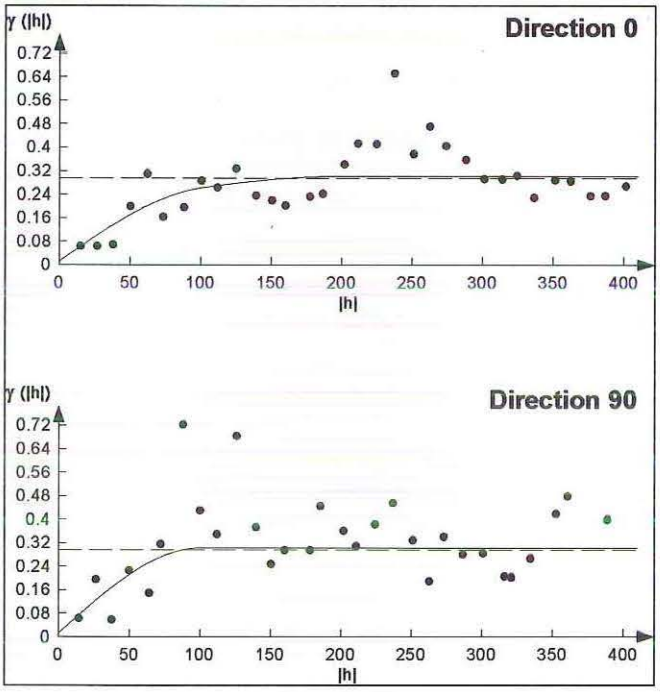


Figure 6.5: Fitted cobalt directional semivariograms

The semivariogram surface for magnesium, displayed in Figure 6.6, exhibits a short range isotropy and a long range anisotropy. The direction of maximum continuity is azimuth 330°. This anisotropy is more pronounced in the madogram surface, also displayed in Figure 6.6.

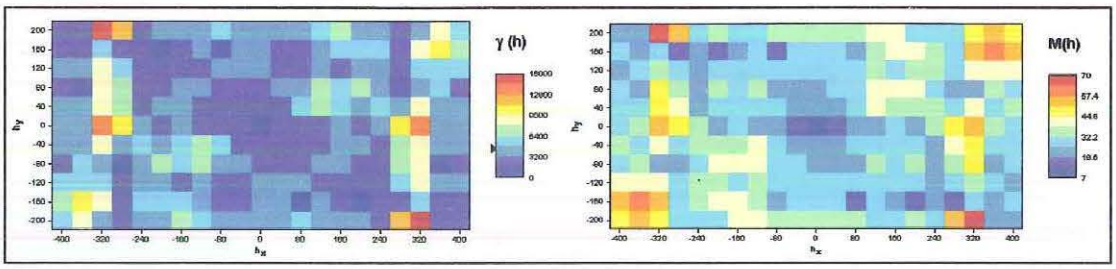


Figure 6.6: Semivariogram and madogram surfaces for magnesium

The model for magnesium consists of three structures, a nugget effect and two spherical structures. The first spherical structure is isotropic with a range of 140 metres. The second spherical structure is modelled in the direction of maximum continuity azimuth 330° with a range of 300 metres and an anisotropy ratio of 0.5. This model can be expressed as:

$$\gamma(\mathbf{h}) = 135 g_0(\mathbf{h}) + 2300 Sph\left(\frac{|\mathbf{h}|}{140}\right) + 2162 Sph\left(\frac{\mathbf{h}'}{300}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} cos330° & sin330° \\ -sin330° & cos330° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0.87 & -0.5 \\ 0.25 & 0.43 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 6.7 shows the directional semivariograms fitted with this model.

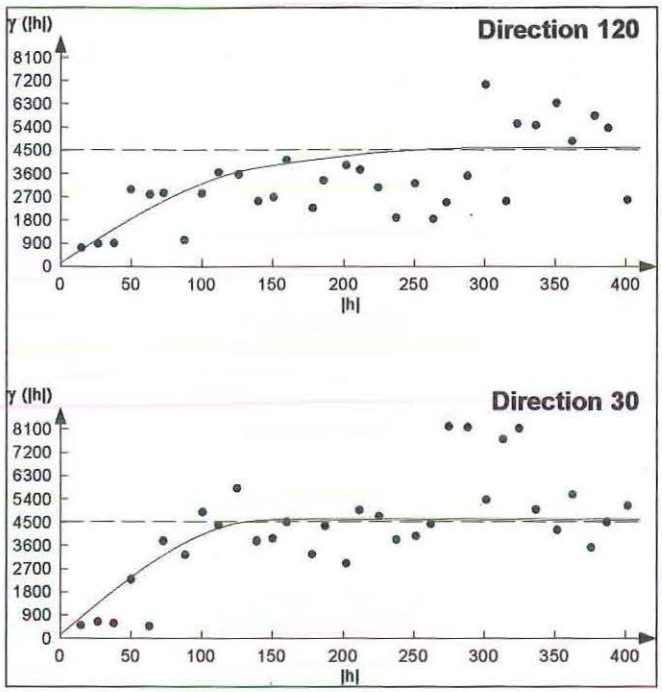


Figure 6.7: Fitted semivariograms for magnesium

The semivariogram surface for iron is displayed in Figure 6.8. The data are clearly anisotropic with the direction of maximum continuity being east-west.

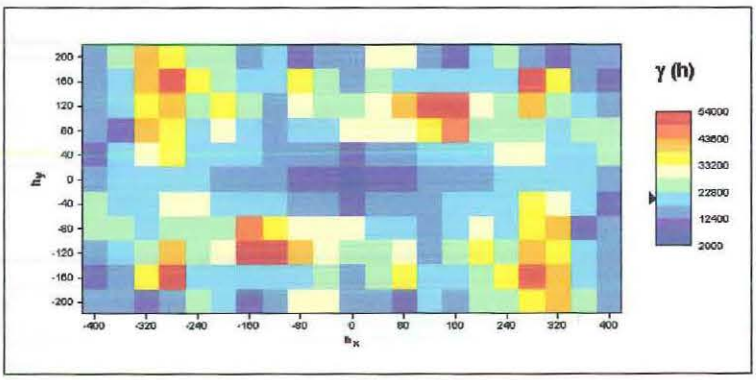


Figure 6.8: Semivariogram surface for iron

The model for iron consists of three structures, a nugget effect and two spherical structures. The first spherical structure is isotropic with a range of 75 metres. The second spherical structure is modelled with east-west as the direction with a range of 185 metres and an anisotropy ratio of 0.54. This model can be expressed as:

$$\gamma(\mathbf{h}) = 1030 g_0(\mathbf{h}) + 3100 Sph\left(\frac{|\mathbf{h}|}{75}\right) + 16790 Sph\left(\frac{\mathbf{h}'}{185}\right)$$

$$where \quad \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.54 \end{bmatrix} \begin{bmatrix} cos90° & sin90° \\ -sin90° & cos90° \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.54 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 6.9 displays the directional semivariograms fitted with this model.

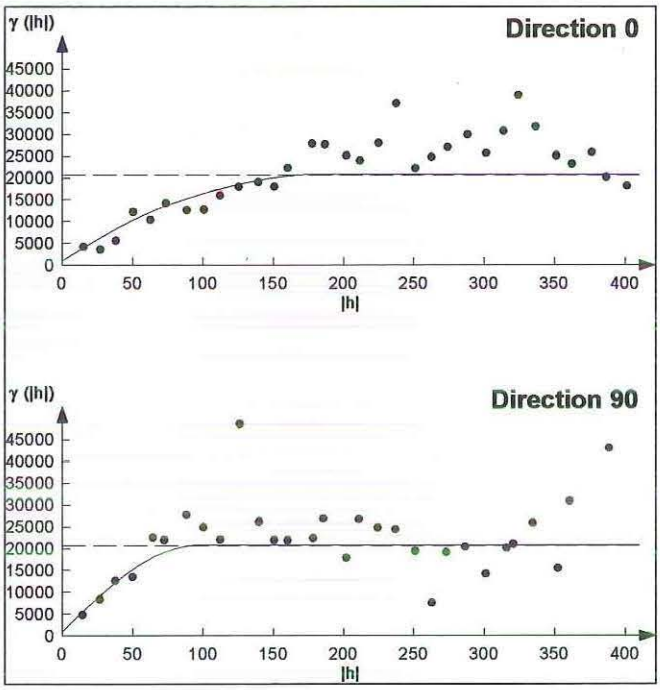


Figure 6.9: Fitted semivariograms for iron

The semivariogram and standardised semivariogram surfaces for zinc are displayed in Figure 6.10. The anisotropy is clear here with the direction of maximum continuity being east-west.

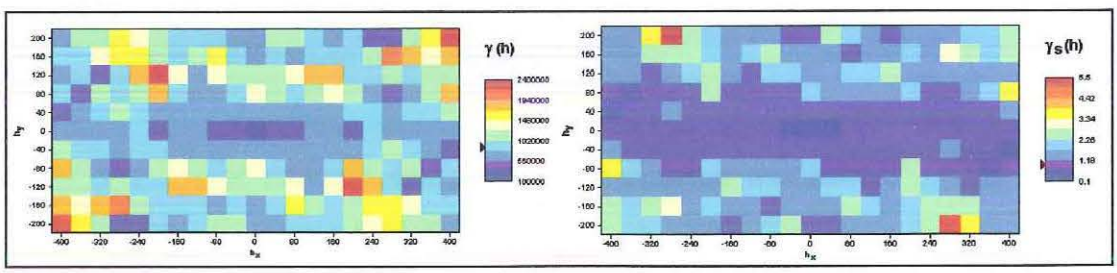


Figure 6.10: Semivariogram and standardised semivariogram surfaces for zinc

The final model selected consists of a nugget effect, a spherical isotropic short range structure of 85 metres and an anisotropic long range structure of 200 metres modelled with east-west as the major direction and an anisotropy ratio of 0.4. This model can be expressed as:

$$\gamma(\mathbf{h}) = 17700 g_0(\mathbf{h}) + 274239 Sph\left(\frac{|\mathbf{h}|}{85}\right) + 601500 Sph\left(\frac{\mathbf{h}'}{200}\right)$$

$$where\ \mathbf{h}' = \begin{bmatrix} 1 & 0 \\ 0 & 0.4 \end{bmatrix} \begin{bmatrix} cos90^\circ & sin90^\circ \\ -sin90^\circ & cos90^\circ \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.4 & 0 \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Figure 6.11 shows the directional semivariograms fitted with this model.

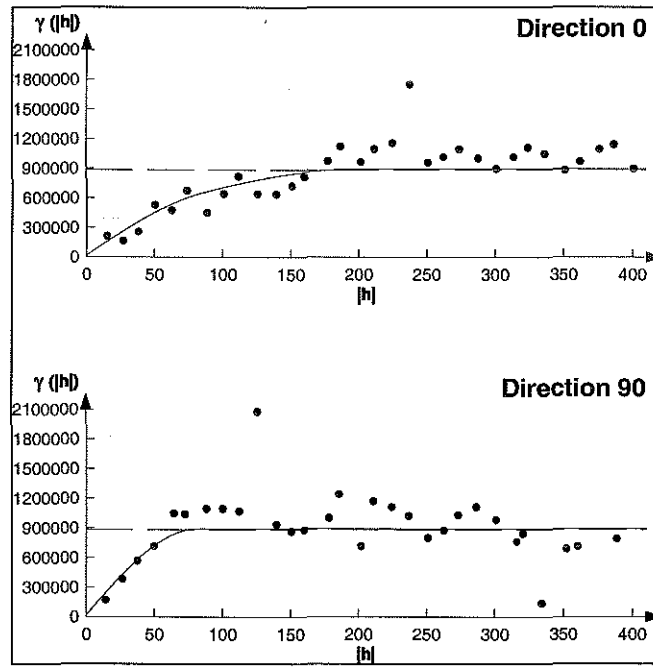


Figure 6.11: Fitted directional semivariograms for zinc

6.2 Cross Validation of Ordinary Kriging

As for the previous data sets cross validation was carried out for all the variables for various search neighbourhoods with little difference in the results. The final search neighbourhoods used in the cross validation and subsequent kriging are displayed in Table 6.1 with those for nickel displayed in Figure 6.12. We present

here the cross validation results for nickel, those for cobalt, magnesium, iron and zinc are displayed in Appendix B. Overall, in all cases, the models have estimated the sample data values well. In particular for nickel, as displayed in Figure 6.13, the residual histogram appears to be roughly normally distributed with the exception of a few values at each tail and the estimates versus the true values plot is reasonably well clustered around the 45° bisector. The standard error of the mean is close to zero and although the variance of the standardised errors is relatively high, as displayed in Table 6.2, it is not high enough to be of concern.

Table 6.1: Neighbourhood search parameters

	Nickel	Cobalt	Magnesium	Iron	Zinc
Number of angular sectors	4	4	4	4	4
Minimum number of samples	3	3	3	3	3
Optimal number of samples per sector	4	4	4	4	4
Search radius (major axis) (metres)	180	220	300	210	220
Search radius (minor axis) (metres)	110	120	160	110	110
Rotation (azimuth)	90°	90°	330°	90°	90°

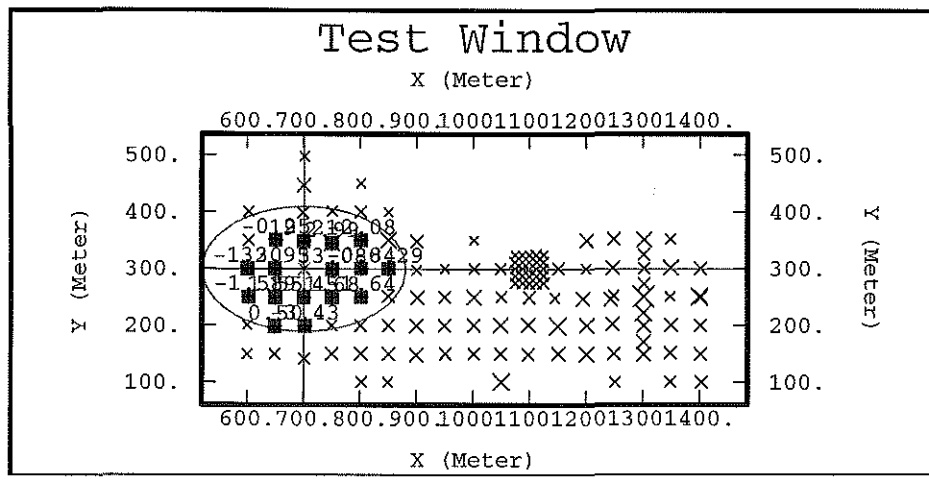


Figure 6.12: Neighbourhood search parameters for nickel

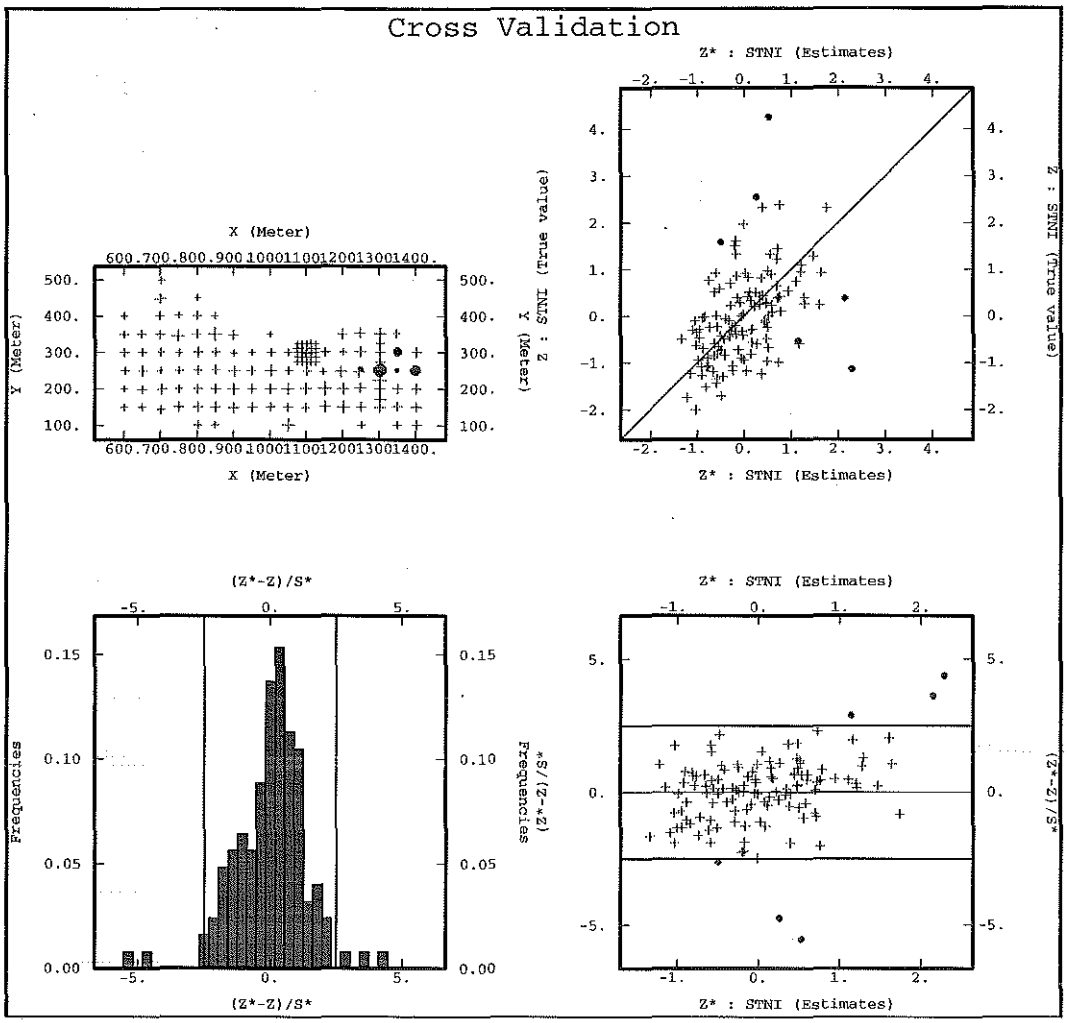


Figure 6.13: Cross validation residual analysis for nickel

Table 6.2: Mean and variance of cross validation residuals

	Nickel		Cobalt		Magnesium		Iron		Zinc	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Error	0.02	0.92	0.02	1.24	-0.02	0.77	0.01	0.69	0.02	0.65
St. E	0.02	1.85	0.02	2.54	-0.02	1.85	0.01	1.96	0.02	1.72

6.3 Ordinary Kriging Estimates of Nickel, Cobalt, Magnesium, Iron and Zinc

Ordinary kriging of the nickel, cobalt, magnesium, iron and zinc variables was performed directly at the 1717 grade control coordinates using the search neighbourhood parameters presented earlier (Table 6.1). Figure 6.15 shows post plots of the estimates for each of the variables along with the post plots of the grade control data and the residuals. As previously the smoothing nature of the kriging algorithm is clearly apparent here. Overall the estimates have reproduced the grade control values well. The same areas of over and underestimation are apparent as for the previous kriging methods. In fact the post plots are once again almost identical to those we have already seen. Figure 6.14 displays the residual post plots of each of the variables. Once again we omit the residual histograms and normal score plots as they are almost identical to those presented earlier.

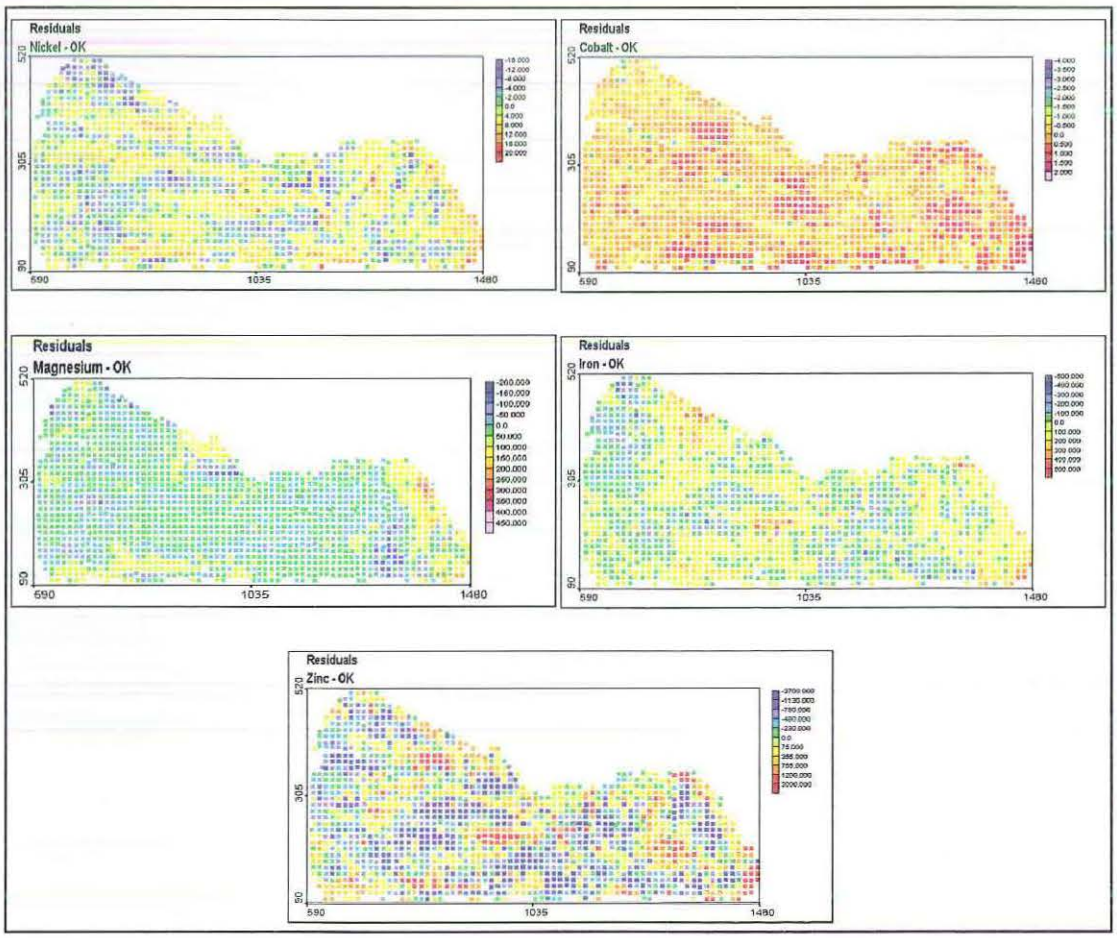


Figure 6.14: Residual post plots for ordinary kriging

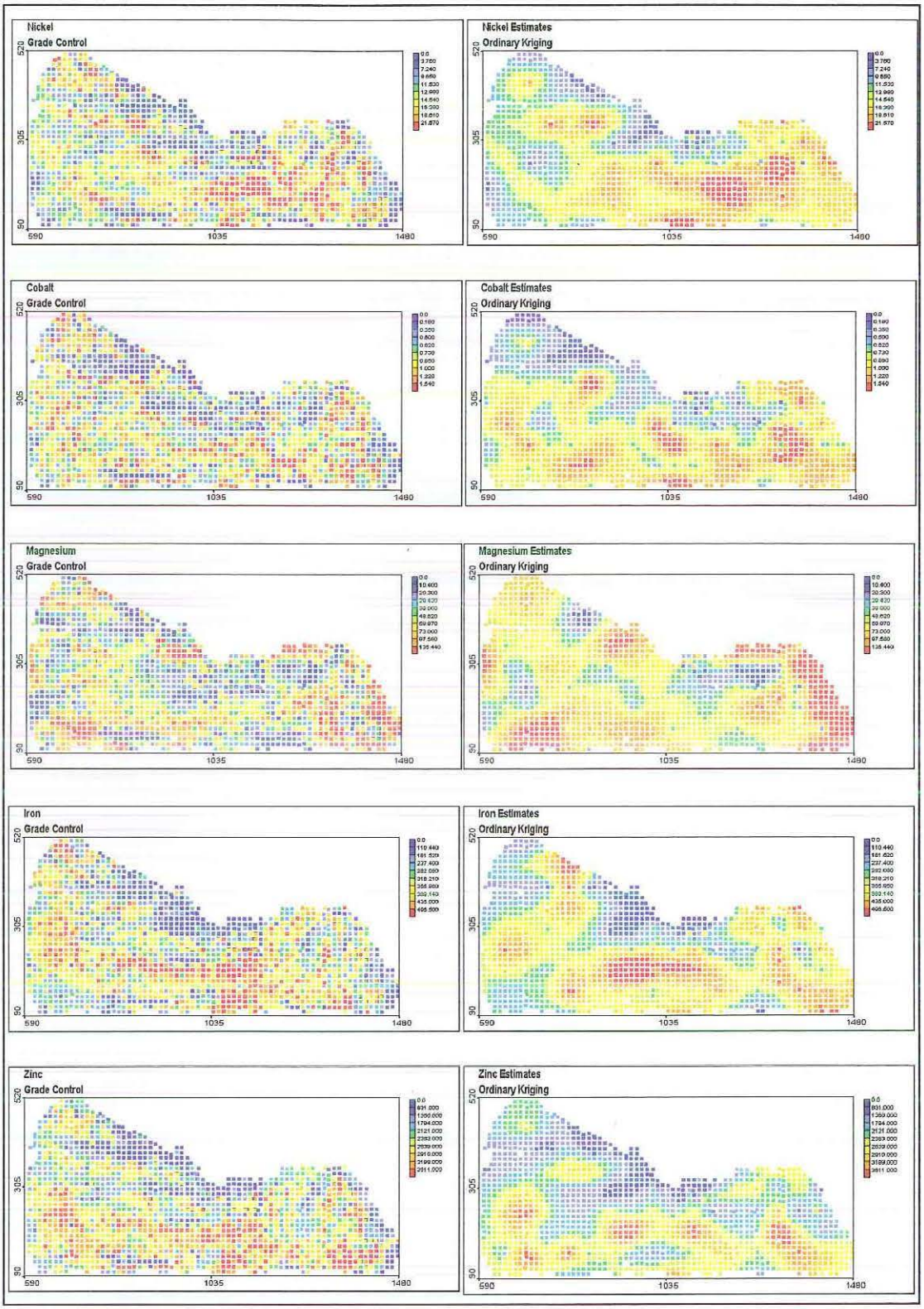


Figure 6.15: Ordinary kriging estimates

7 Comparison of Estimates

The post plots shown earlier in Figures 4.6, 4.14, 5.12, 5.13 and 6.15 show that, as expected, the ordinary cokriging and ordinary kriging techniques used have resulted in very similar estimates. The principal component kriging estimates are also very similar to the other estimates. In addition the principal component estimates using only the retained principal components are very similar to all other estimates. This indicates that very little information has been lost by retaining only two principal components. Figures 7.1 to 7.5 and Tables 7.2 to 7.6 display the descriptive statistics for each of the variables. Overall all of the kriging techniques used have reproduced the grade control and exploration descriptive statistics reasonably well. In all cases the mean and median of the estimates has over estimated that of the exploration and grade control data. The only exception is zinc where the mean and median of the estimates are close in value to those of the grade control data. The values obtained for the mean and median for each variable are very close in value for each kriging method.

The standard deviation of the estimates in all cases is considerably lower than that of the grade control and exploration data, yet another indication of the smoothing of the variability due to the kriging process. This is further supported by the coefficient of variation which in all cases is lower for the estimates than that of the grade control and exploration values. The standard deviation and coefficient of variation obtained for each very is very similar for each kriging technique used. The skewness of the histograms of the grade control data for each variable has been well reproduced for all variables except cobalt. While the grade control cobalt is strongly positively skewed the estimates are not, in fact the skewness of the estimates is around one third of that of the grade control data.

Table 7.1 displays the mean square errors and mean absolute errors for each variable and each kriging method used. The lowest mean square error for each variable is coloured red and the lowest mean absolute error is coloured blue. Overall for the kriging methods used for each variable the values for these measures are quite similar. For nickel the lowest mean square error and mean absolute error were

associated with the principal component kriging estimates. The lowest mean square error for cobalt is associated with the principal component kriging estimates using only two principal components. Correct to two decimal places cobalt had identical values for the mean absolute error for all methods. However correct to four decimal places the principal component kriging estimates had the lowest value for this measure. Both the mean square error and mean absolute error for magnesium were lowest for the estimates obtained from cokriging. The iron mean square error and mean absolute error were the lowest for the principal component kriging estimates obtained using only two principal components. For zinc the lowest values for both error measures were associated with principal component kriging.

Table 7.1: Mean square error and mean absolute error for each method

		Nickel	Cobalt	Magnesium	Iron	Zinc
OCK - ICM	MSE	41.71	0.34	NA	20489.95	847999.31
	MAE	4.88	0.43	NA	107.91	705.32
OCK - LMC	MSE	41.46	0.33	3269.59	20003.96	NA
	MAE	4.86	0.43	39.51	106.59	NA
PCK	MSE	40.50	0.33	NA	20497.71	838231.21
	MAE	4.81	0.43	NA	107.92	699.99
PCK2	MSE	42.51	0.32	NA	18568.06	884728.72
	MAE	5.05	0.43	NA	105.29	733.77
OK	MSE	41.56	0.34	3612.69	20365.44	843825.91
	MAE	4.87	0.43	41.09	107.32	701.78

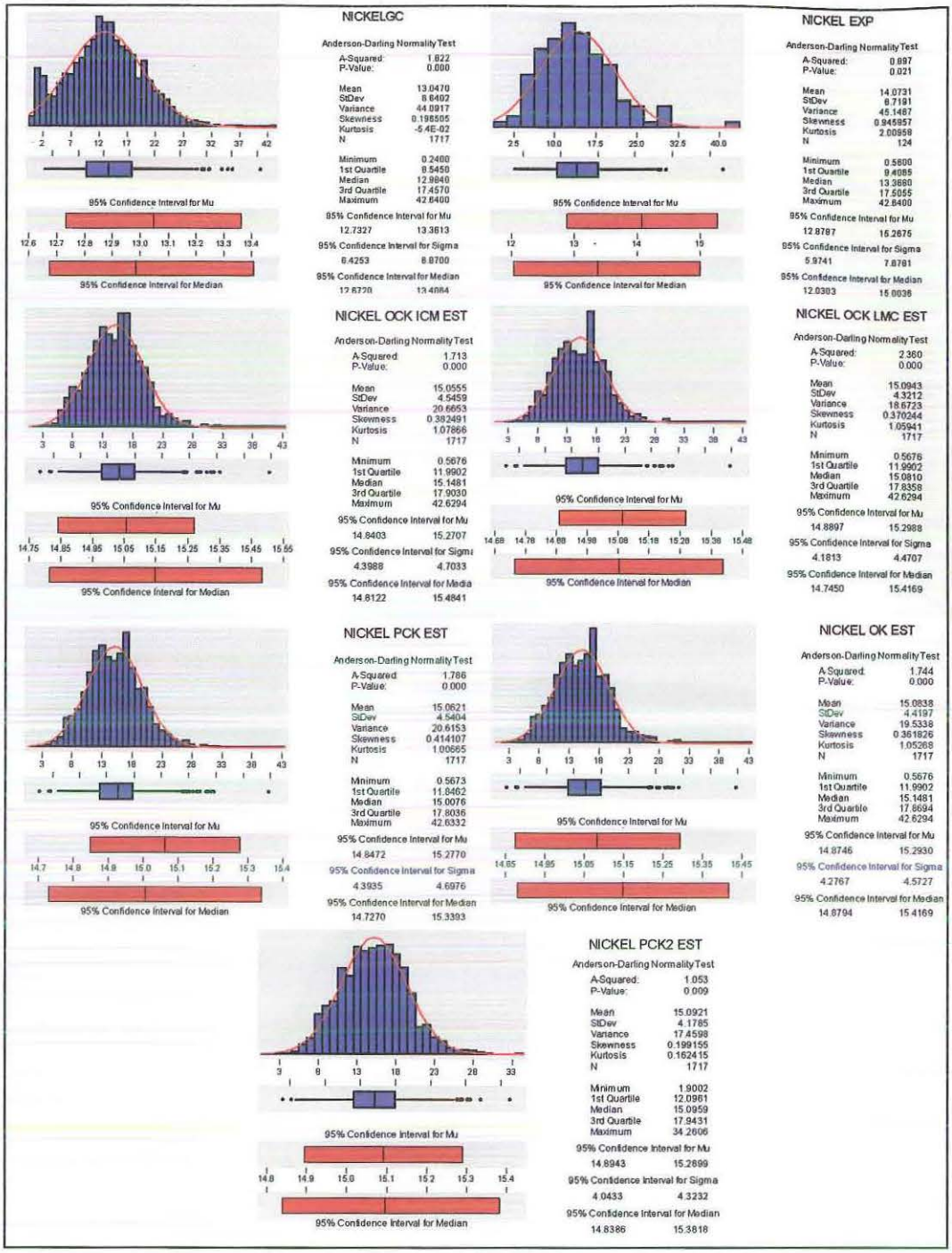


Figure 7.1 and Table 7.2: Nickel descriptive statistics

	GC	EXP	OCK ICM EST	OCK LMC EST	PCK EST	PCK2 EST	OK EST
n	1717	124	1717	1717	1717	1717	1717
Mean	13.05	14.07	15.06	15.09	15.06	15.09	15.08
Median	12.98	13.37	15.15	15.08	15.01	15.10	15.15
SD	6.64	6.72	4.55	4.32	4.54	4.18	4.42
Skewness	0.20	0.95	0.38	0.37	0.41	0.20	0.36
CV	0.51	0.48	0.30	0.29	0.30	0.28	0.29

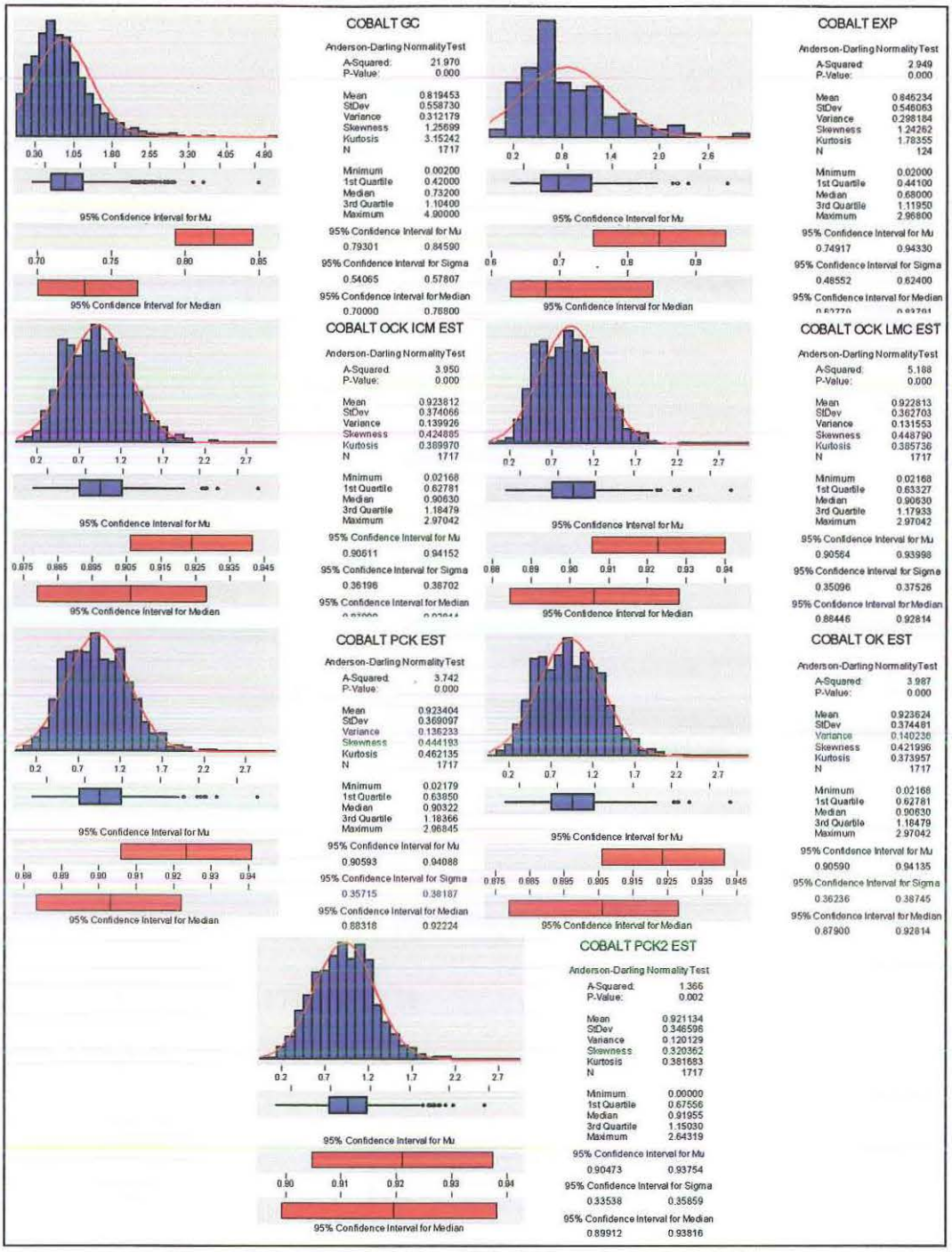


Figure7.2 and Table 7.3: Cobalt descriptive statistics

	GC	EXP	OCK ICM EST	OCK LMC EST	PCK EST	PCK2 EST	OK EST
n	1717	124	1717	1717	1717	1717	1717
Mean	0.82	0.85	0.92	0.92	0.92	0.92	0.92
Median	0.78	0.68	0.91	0.91	0.90	0.92	0.91
SD	0.56	0.55	0.37	0.36	0.37	0.35	0.37
Skewness	1.26	1.24	0.42	0.45	0.44	0.32	0.42
CV	0.68	0.65	0.40	0.49	0.40	0.38	0.46

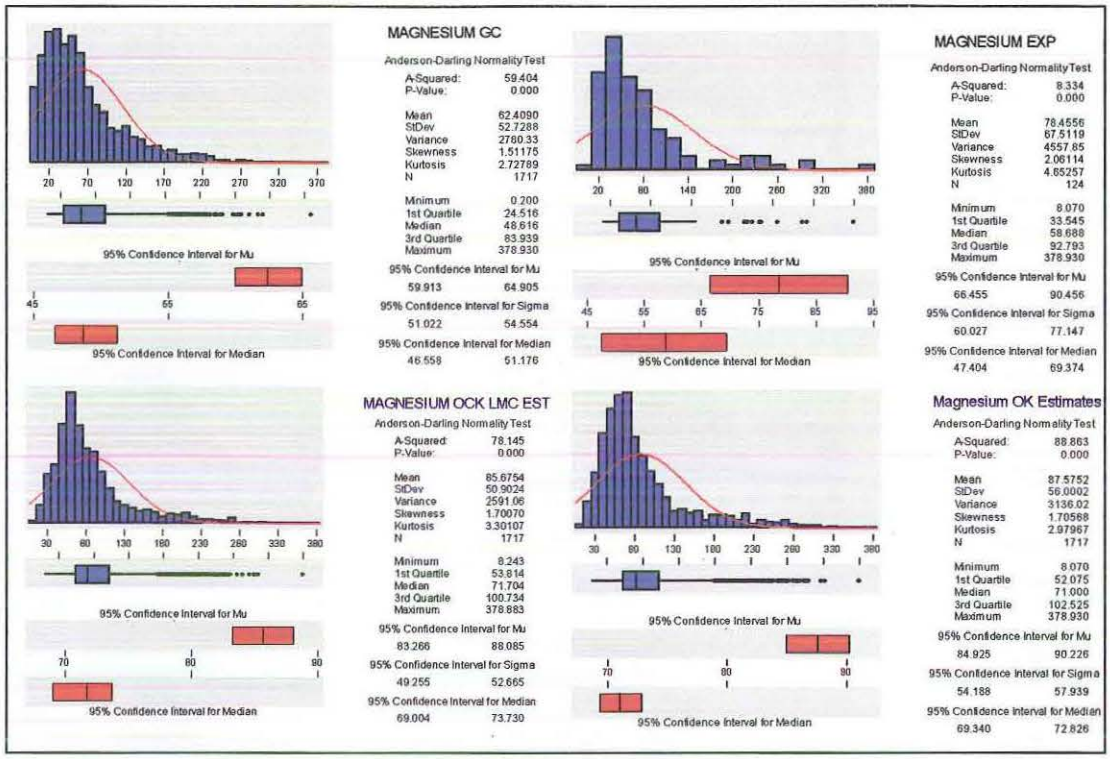


Figure 7.3: Magnesium descriptive statistics

Table 7.4: Magnesium descriptive statistics

	GC	EXP	OCK LMC EST	OK EST
n	1717	124	1717	1717
Mean	62.41	78.46	85.68	87.58
Median	48.62	58.69	71.70	71.00
SD	52.73	67.51	50.90	56.00
Skewness	1.51	2.06	1.70	1.71
CV	0.84	0.86	0.59	0.64

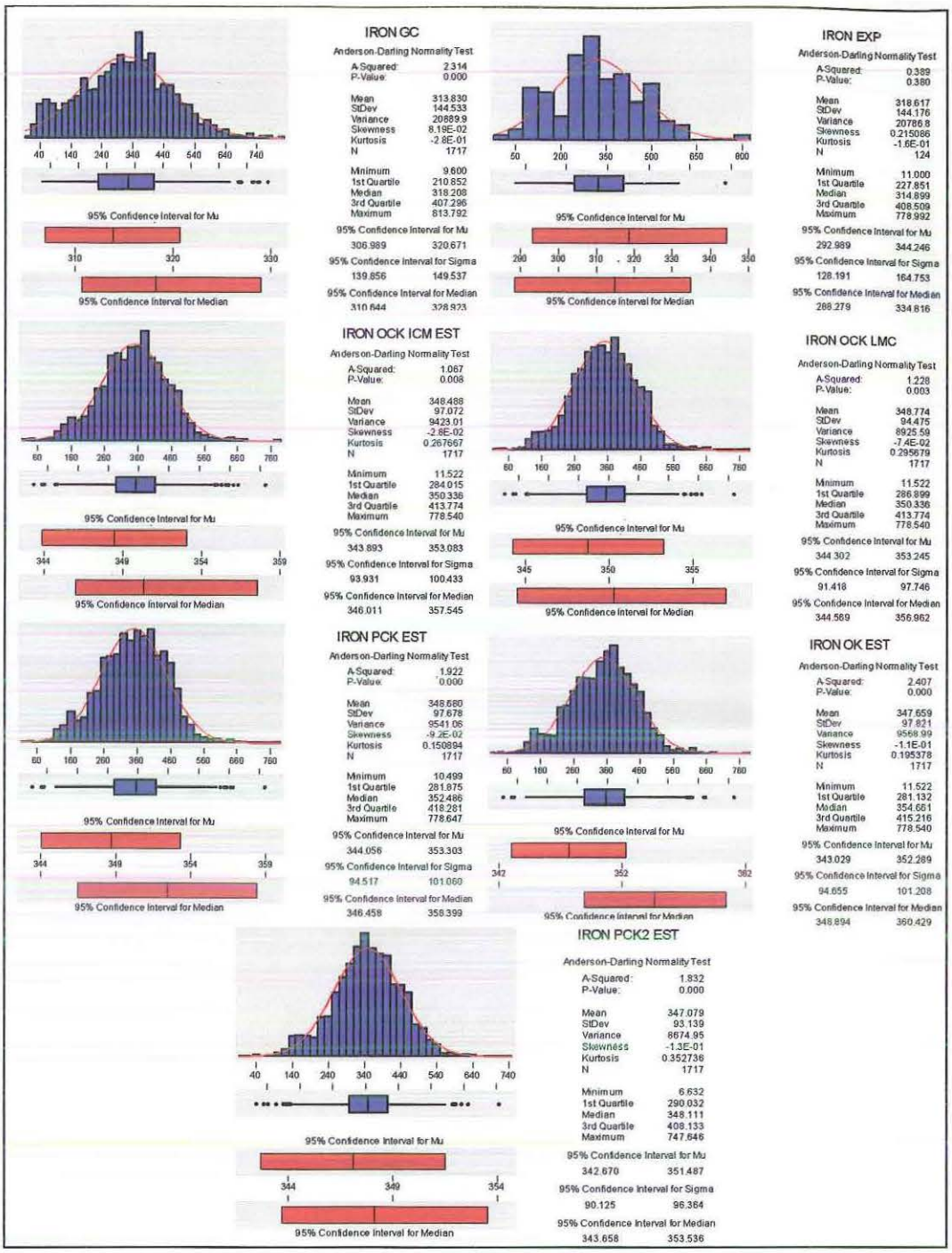


Figure 7.4 and Table 7.5: Iron descriptive statistics

	GC	EXP	OCK ICM EST	OCK LMC EST	PCK EST	PCK2 EST	OK EST
n	1717	124	1717	1717	1717	1717	1717
Mean	313.83	318.62	348.49	348.77	348.68	347.08	347.66
Median	312.19	315.00	350.34	350.34	352.49	348.11	354.66
SD	144.53	144.18	97.07	94.48	97.68	93.14	97.82
Skewness	0.08	0.22	-0.03	-0.07	-0.09	-0.01	-0.01
CV	0.46	0.45	0.28	0.27	0.28	0.27	0.28

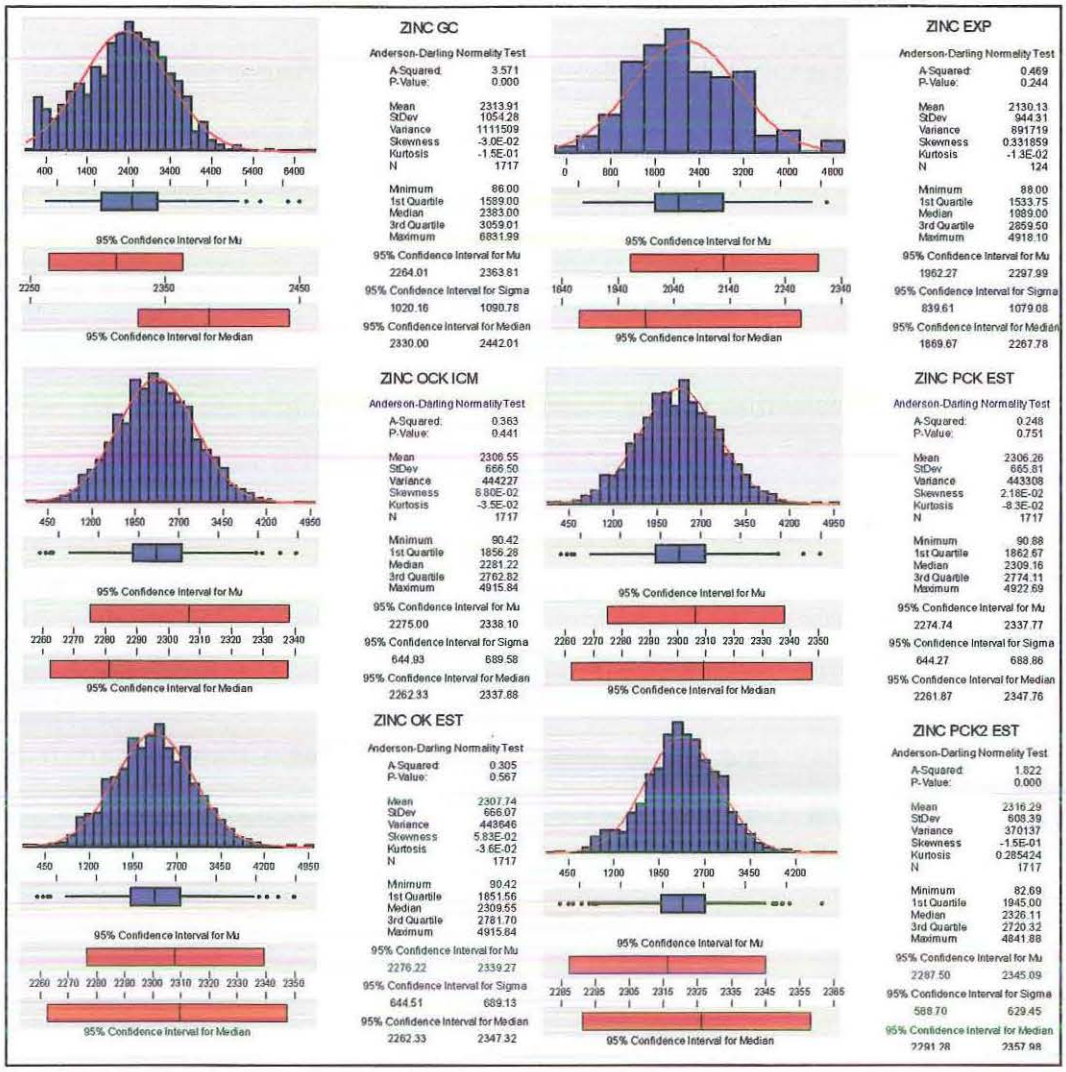


Figure 7.5: Zinc descriptive statistics

Table 7.6: Zinc descriptive statistics

	GC	EXP	OCK LMC EST	PCK EST	PCK2 EST	OK EST
n	1717	124	1717	1717	1717	1717
Mean	2312.91	2130.13	2306.55	2306.26	2316.29	2307.74
Median	2383.00	1989.00	2281.22	2309.16	2326.11	2309.55
SD	1054.28	944.31	666.50	665.81	608.39	666.07
Skewness	-0.03	0.33	0.09	0.02	-0.02	0.06
CV	0.46	0.44	0.29	0.29	0.26	0.29

Next we consider the question of conditional bias by assessing the Q-Q plots of the grade control data and estimates of each variable for each of the kriging techniques used. These plots are displayed in Figures 7.6 to 7.10. It is clear from these that the overestimation of the mean for the nickel, cobalt, and iron variables has resulted from significant overestimation of the low values and underestimation of the high values. Overall nickel and iron appear to be the least affected by conditional bias. As all of the kriging techniques produced similar estimates for each variable, as was expected, there is little difference in the Q-Q plots.

The underestimation of high values is not too marked for nickel for all techniques except the principal component kriging where only two principal components were retained. This technique has grossly underestimated the particularly high values. The full principal component kriging has the best performance with regard to underestimation of the high values. There has been significant overestimation of the low values, particularly around the 5^{th} to 20^{th} percentiles. There is little difference between the techniques in the overestimation of low values although after the 25^{th} percentile principal component kriging performs marginally better than the others.

For cobalt the underestimation of the high values is significant regardless of the kriging technique. In this case the principal component kriging (two principal components) was the worst offender. There has been some overestimation of low values for this variable, again particularly around the 5^{th} to 20^{th} percentiles. Overall principal component kriging is marginally better than the other techniques here, although all perform similarly.

Magnesium has suffered overestimation across all values although the cokriging estimates perform somewhat better than the ordinary kriging estimates from the 80^{th} percentile. Magnesium was particularly difficult to model with very erratic semivariograms. Little benefit was gained from examining the other measures such as the madogram and pairwise relative semivariograms, though these measures did assist in better defining the ranges of the model. The poor appearance of the Q-Q plot for this variable indicates that the model has not adequately captured the erratic behaviour of this variable.

The underestimation of high iron values is not as marked as for the other variables discussed so far. In this case the principal component kriging method

obtained using only two principal components is the worst offender, though certainly not as significant as seen previously. Ordinary kriging has the best performance from the 95th percentile. The overestimation of lower values for iron follows a similar pattern to nickel and cobalt with little difference in techniques.

For zinc the underestimation of the high values is significant regardless of the kriging technique. Again the principal component kriging (two principal components) was the worst offender, though only marginally. The best performer for this variable was the ordinary cokriging using the intrinsic correlation model. There has been some overestimation of low values for this variable, particularly around the 5th to 15th percentiles. The only technique that differs from the others at the lower end of the plot is principal component kriging (two principal components) with marginally higher overestimation from the 10th to 35th percentiles.

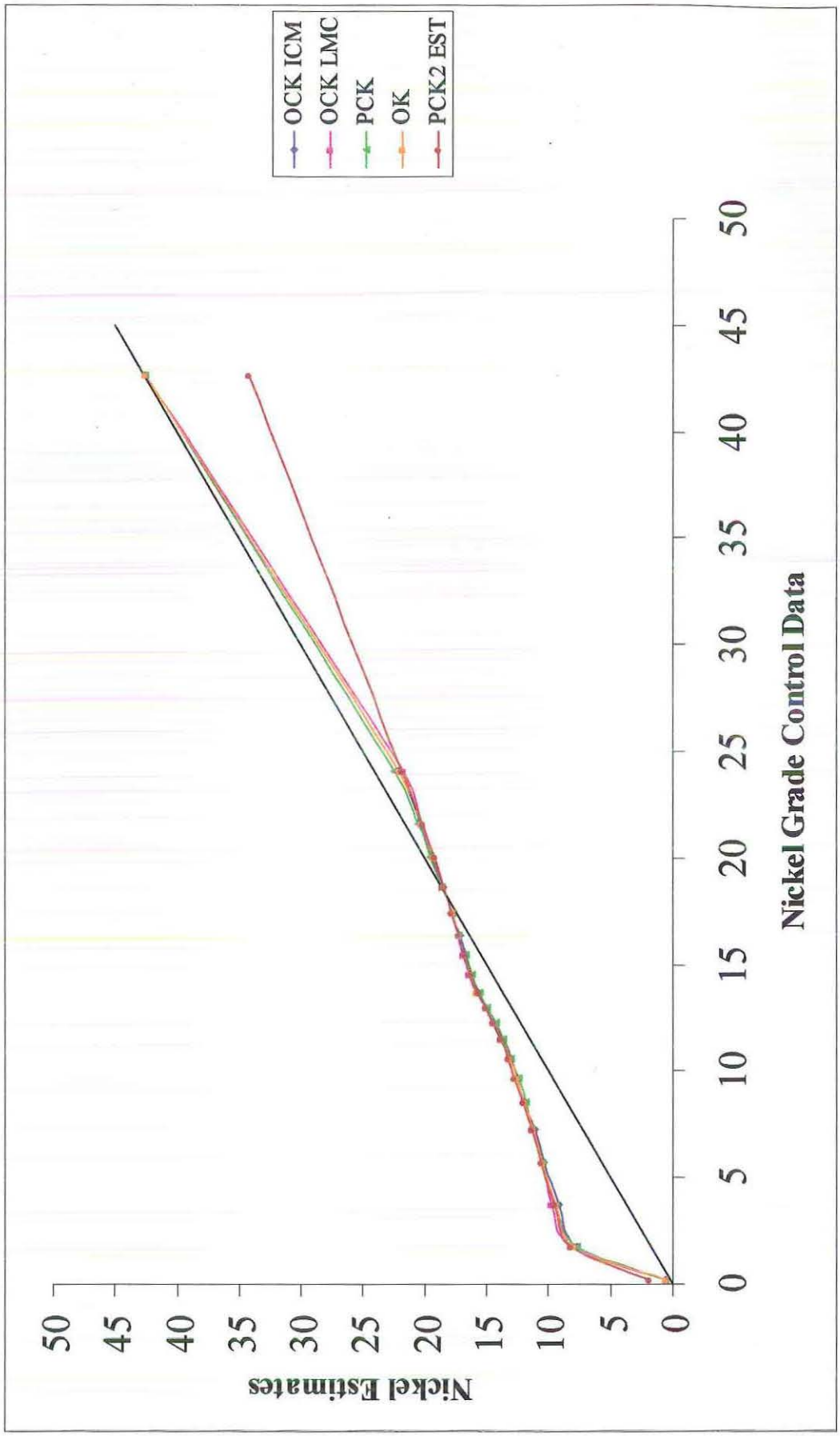


Figure 7.6: Q-Q plots for nickel

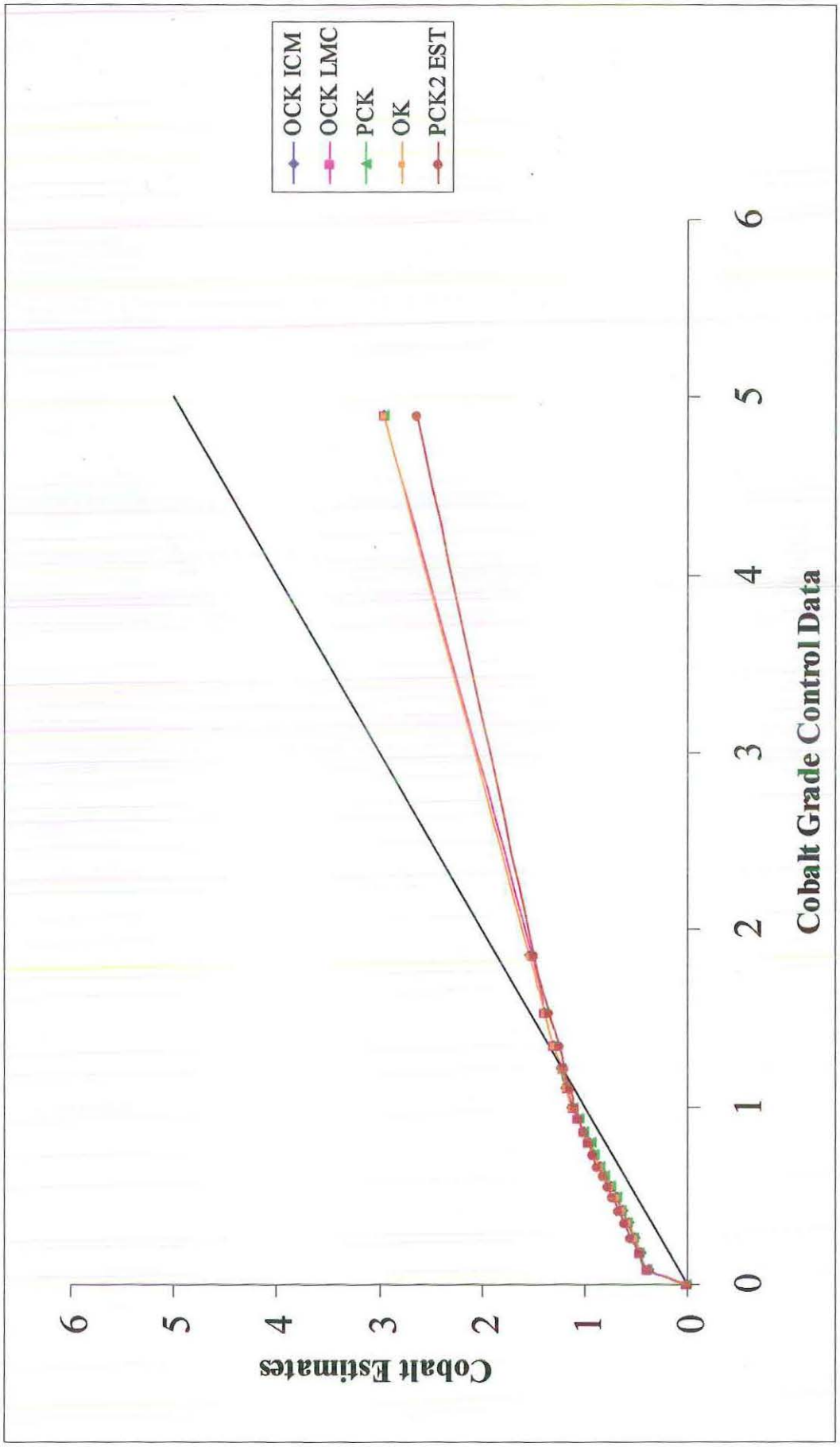


Figure 7.7: Q-Q plots for cobalt

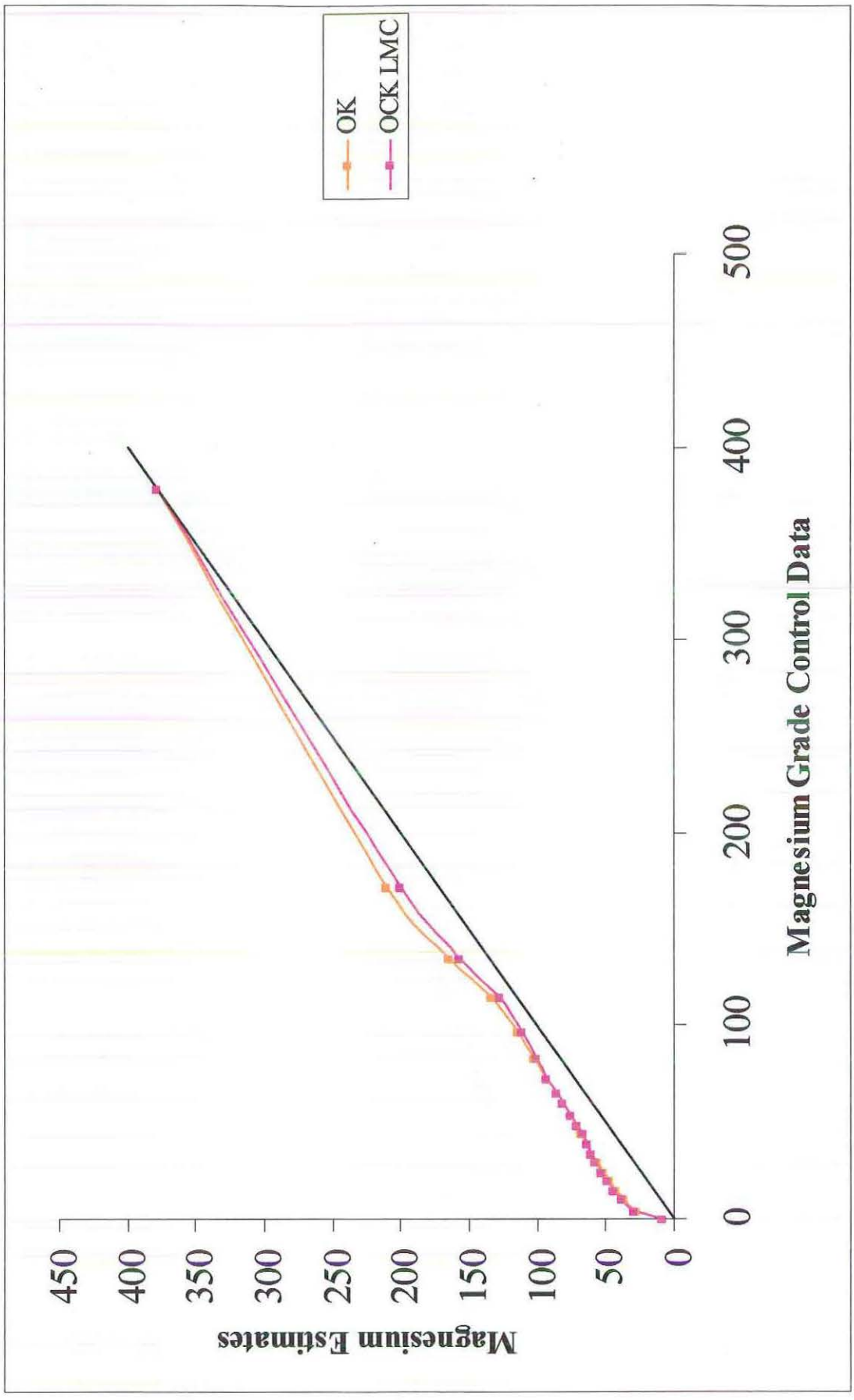


Figure 7.8: Q-Q plots for magnesium

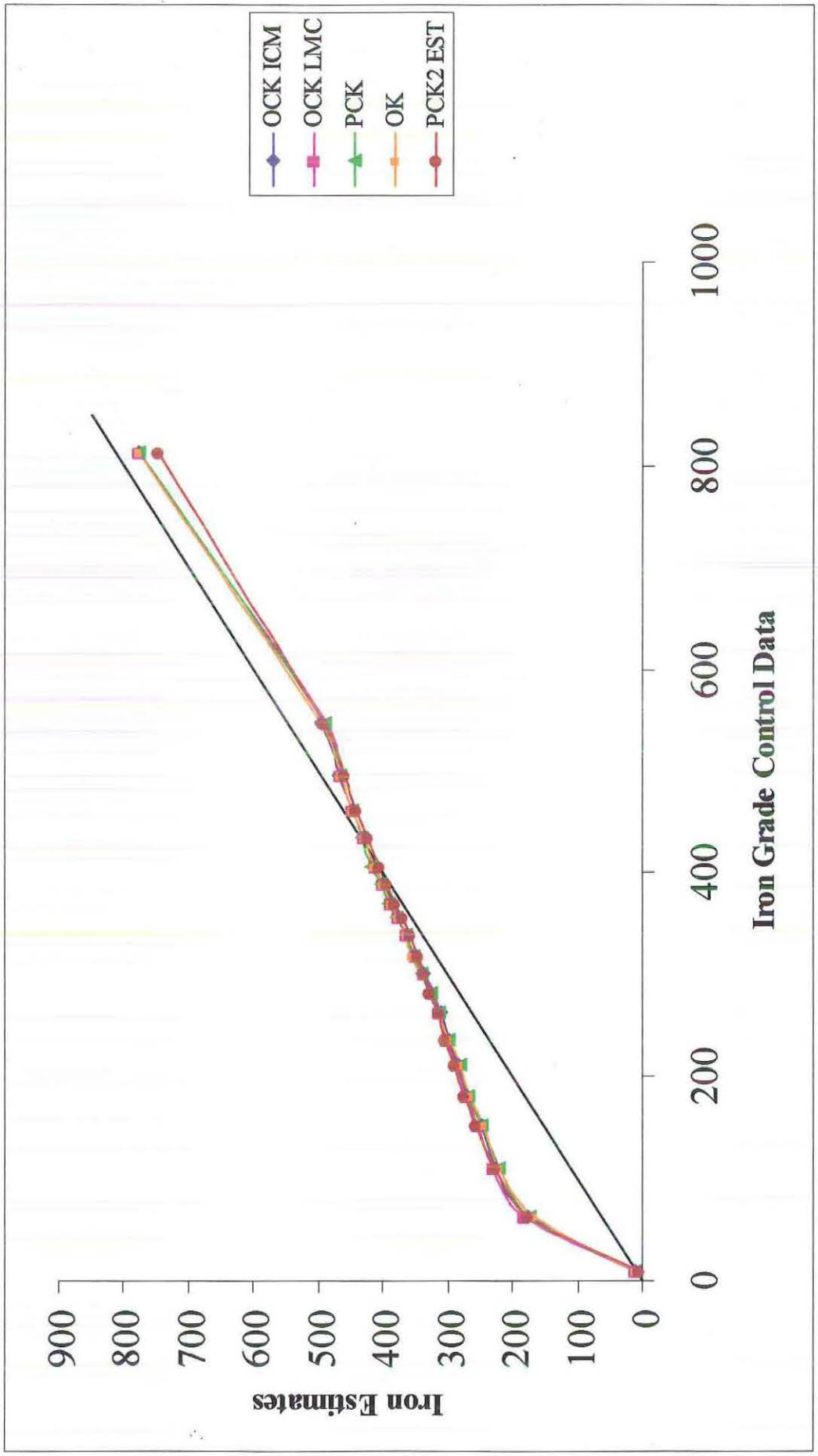


Figure 7.9: Q-Q plots for iron

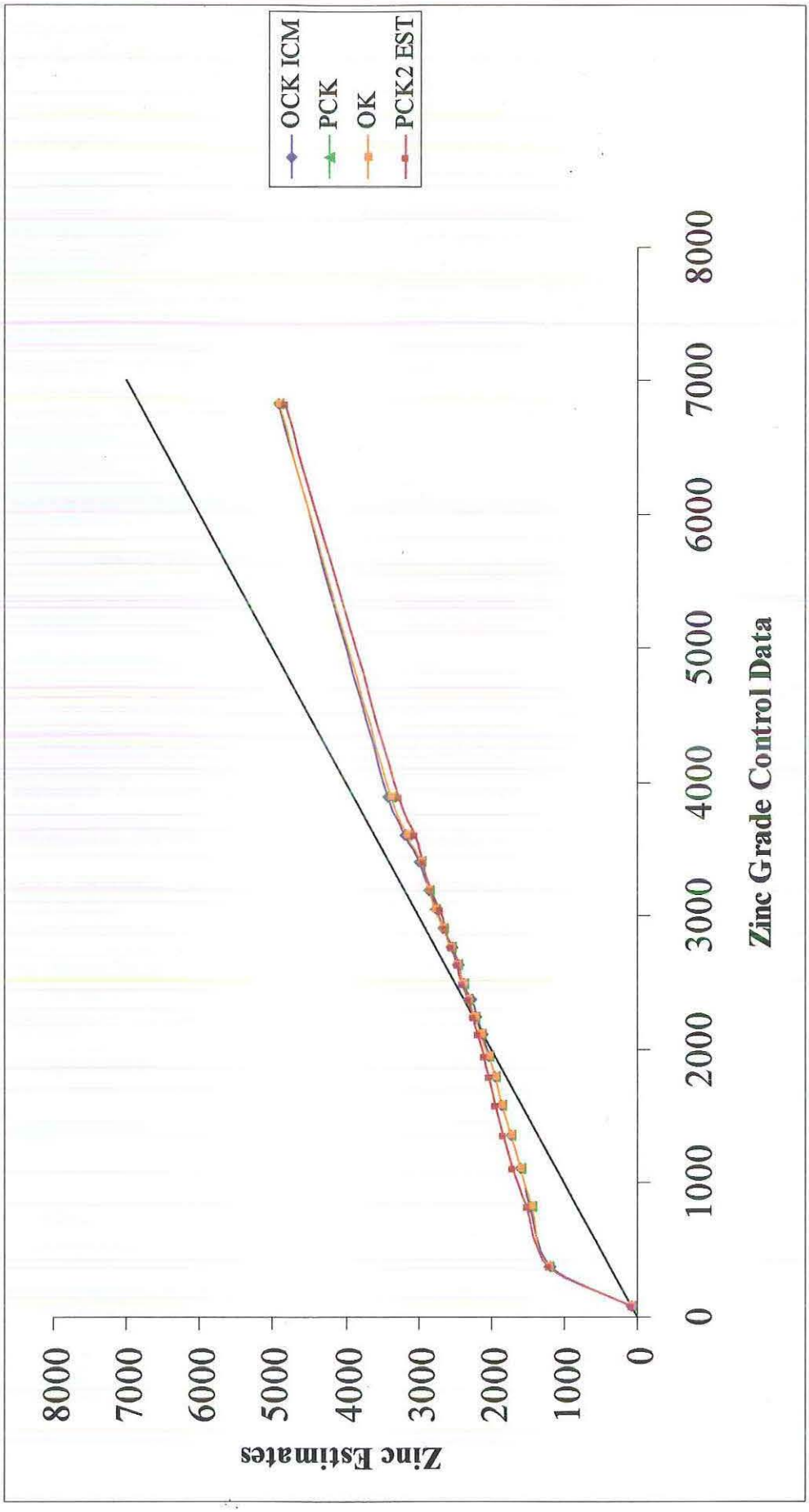


Figure 7.10: Q-Q plots for zinc

Finally we examine the estimation variance for each of the kriging methods as displayed in Figures 7.11 and 7.12. For comparative purposes we display the estimation variance for ordinary kriging using the standardised exploration data. The estimation variance is clearly highest in the far north-western corner, along the north-western boundary, in the far south-eastern corner and along the southern boundary of the study region. These areas all fall outside the region bounded by the exploration data, h[illegible]e we are extrapolating rather than interpolating in these areas. The estimation variance is lowest around the locations of the exploration data, and is, as expected, zero at these locations (kriging is an exact interpolator). The estimation variance was identical for all of the variables estimated using the intrinsic coregionalisation model. The estimation variance of the principal components gets successively smaller with each principal component.

While the estimation variance is expected to be lower for the cokriging estimates this was true in our case for nickel and iron only for the intrinsic coregionalisation model and nickel only for the linear model of coregionalisation. Overall the estimation variance for nickel was lower around the locations of the exploration data, however as we move further away from these points the estimation variance becomes the same as the ordinary kriging estimation variance. The estimation variance of the intrinsic coregionalisation model is lower still than the other methods. For cobalt the ordinary kriging estimation variance was the same as that for the intrinsic coregionalisation model, both of which were better than those of the linear model of coregionalisation. For magnesium and iron the ordinary kriging estimation variance was lower overall than the other methods. For zinc the intrinsic coregionalisation model yielded the lowest estimation variance.

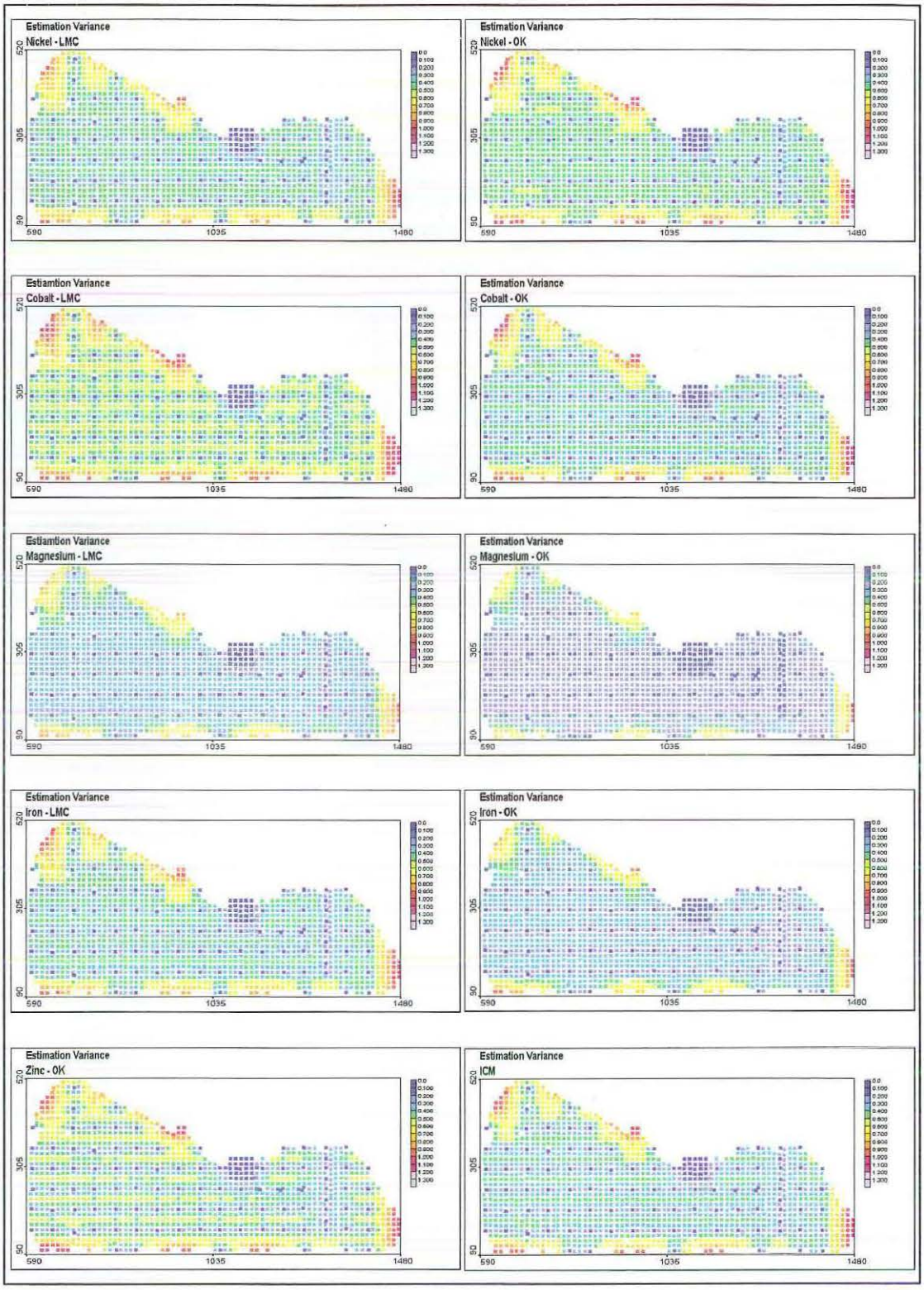


Figure 7.11: Post plots of estimation variance for ordinary cokriging and ordinary kriging

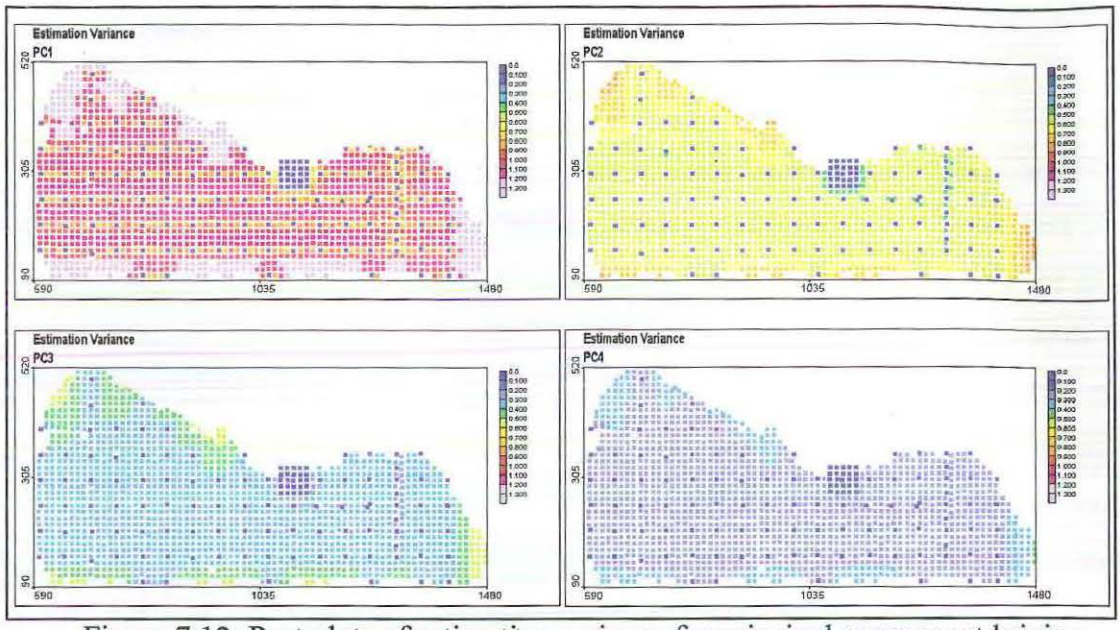


Figure 7.12: Post plots of estimation variance for principal component kriging

8 DISCUSSION AND CONCLUSIONS

Data collection in the earth sciences is rarely limited to just one variable at each location. Such is the case in mining, where it is common that each drill core taken will be sampled for not just the primary variable of interest but also for other attributes. These secondary variables are often spatially cross correlated with the primary variable as well as each other. Hence each of the variables sampled may contain useful information about the others. Multivariate geostatistics allows us to exploit the relationships among and between the variables with the aim of improving their estimates at unsampled locations. This is particularly the case when the primary variable of interest is undersampled in relation to the secondary variables.

This thesis presented the theory and the process of the modelling and estimation of spatially dependent multivariate data. These methods were illustrated by application to a multivariate data set from the Murrin Murrin nickel mine near Laverton in Western Australia. We have demonstrated the use of the classical multivariate statistical technique principal component analysis in a geostatistical environment. We have exhibited the development and application of three models of spatial continuity, namely the intrinsic coregionalisation model, the linear model of coregionalisation and the linear model of regionalisation. We have used each of these models to estimate values of the attributes at unsampled locations using ordinary cokriging, principal component kriging and ordinary kriging.

The data suite MM22D used in this study came from an actual mineralisation and consisted of two data sets, a grade control (exhaustive) data set and an exploration (sample) data set. As the data are isotopic we did not have the case where the primary variable is undersampled in relation to the secondary variables. This would be unrealistic in a mining situation when all data are coming from the same drill core samples. The data suite comprises three dimensional grade and thickness measurements on eight variables: nickel, cobalt, magnesium, iron, aluminium, chromium, zinc and manganese. For the purposes of this study the data were transformed to accumulations, hence were considered two dimensional. In order to demonstrate the different types of models of spatial variability yet remain within the

scope of an honours thesis we decided to limit the analysis to two four variable subsets created from the exploration data. One subset consisted of highly correlated variables (nickel, cobalt, iron and zinc), the other subset was chosen for the economic importance of the variables (nickel, cobalt, magnesium and iron).

A detailed exploratory data analysis was carried out on each of the variables in the MM22DGC and MM22DEXP data sets. Overall the exploration data reproduced the grade control summary statistics well. Cobalt, aluminium, magnesium and manganese were all strongly positively skewed for both the grade control and exploration data. It was however decided that a transformation of these variables was not warranted. A principal component analysis was performed on the MM22DEXP data set and it was determined (by examining the cross correlograms of the principal components) that the data were not intrinsically correlated. We then considered the principal component analysis of the two four variable subsets MM22DHC4 and MM22DTOP4. It was determined that the former subset was intrinsically correlated. In both cases the principal components extracted could be reasonably well interpreted.

The first estimation technique we investigated was ordinary cokriging. In order to implement the cokriging algorithm we required a permissible model of the spatial variability of the variables being estimated. The first model that we demonstrated was the intrinsic coregionalisation model using the MM22DHC4 data set. The intrinsic coregionalisation model is a particular example of the linear model of coregionalisation where all of the sills of the basic semivariogram models are proportional to each other. Under the assumption of second-order stationarity the matrix of coefficients was for our data set the correlation matrix. As we were dealing with standardised data we required that the sills of the basic models summed to one. The actual modelling process then for this data set was relatively simple in that it consisted of identifying the basic structures that captured the spatial variability of the variables and determining their corresponding sills. The conditions for this type of model though are restrictive and in practice the intrinsic model rarely fits experimental coregionalisations (Gooevaerts, 1997, p. 117).

The second model we investigated was the more general linear model of coregionalisation which was demonstrated using the MM22DTOP4 data set. The difficulty with this type of model lies not only in the fact that each of the

$N_v(N_v + 1)/2$ (10 in our case) direct and cross semivariogram models must be jointly inferred, but that each of the coregionalisation matrices must be positive semi-definite. In order to ensure that the linear model of coregionalisation we were building was indeed permissible it was necessary to first determine jointly the basic set of structures that best captured the characteristics of the direct semivariograms (nickel, cobalt, magnesium and iron). It was then necessary to restrict our choice of structures for the cross semivariograms to those used for the direct ones. Having obtained a model with a suitable fit on all direct and dross semivariograms it was necessary to make the coregionalisation matrices positive semi-definite. As we were dealing with fairly small coregionalisation matrices we did this manually using the property of diagonal dominance. This process can indeed be tedious, time consuming and frustrating, and in addition may not be optimal, so for larger coregionalisation matrices an iterative procedure, such as that outlined in Goovaerts (1997) is recommended.

The estimates obtained from cokriging using both models were compared to the grade control data, which we considered to be reality for this study. We compared the postplots of the estimates with those of the grade control data. In both cases the cokriging estimates captured the overall spatial distribution of the grade control data. In addition we compared the descriptive statistics of the estimates obtained for each variable with those of the grade control data. In all cases the mean and median of the estimates has over estimated that of the exploration and grade control data. The only exception is zinc where the mean and median of the estimates were close in value to those of the grade control data. The standard deviation of the estimates from both cokriging cases was considerably lower than that of the grade control and exploration data, which is an indication of the smoothing of the variability due to the cokriging process. The residual analysis for both cokriging methods showed the residuals to be approximately normally distributed with the only exception being cobalt. In both cases the histograms of the cobalt residuals were negatively skewed.

The second estimation technique we investigated was principal component kriging. The overriding assumption of this technique is that the data are intrinsically correlated. Hence we used the principal components extracted from the MM22DHC4 data set for this method. In order to perform the ordinary kriging of the principal

component scores we were required individually model the spatial variability of the principal component scores using a linear model of regionalisation. We then used ordinary kriging to estimate the principal component scores at unsampled locations and recalculated the estimates of the original variables using the corresponding coefficients from the matrix of eigenvectors. In addition to estimating the variables using all four principal components we kriged only the first two (which accounted for 89% of the total variability of the original data) and calculated our estimates using only these.

The estimates obtained from both forms of principal component kriging were compared with the grade control data. Overall, for all variables, the post plots of the estimates in both cases have reproduced those of the grade control values reasonably well. In addition the descriptive statistics of the estimates were compared to those of the grade control data. Once again the mean and median of the nickel, cobalt and iron estimates have overestimated those of the grade control data while those for zinc are close in value. The standard deviation is considerably lower for all four variables estimated than for the grade control data. The residual analysis confirmed that the residuals were approximately normally distributed with cobalt once again being the only variable to not conform this. What is most significant here is that the estimates obtained by using only the first two principal components are remarkably close in value to those obtained using all four.

The ordinary kriging estimates were also obtained for the nickel, cobalt, magnesium, iron and zinc variables. As for the individual principal components this involved obtaining a linear model of regionalisation for each of the variables individually. As the data we used were isotopic it was anticipated that the estimates obtained in the ordinary kriging method would be similar to those obtained from the other methods. This was indeed the case with all estimates for each variable being almost identical. In particular in the case where the data are intrinsically correlated the ordinary cokriging and ordinary kriging estimates are expected to be equivalent. This was especially reflected in the Q-Q plots where the dark blue ordinary cokriging (intrinsic coregionalisation model) line is completely covered by the orange ordinary kriging line.

The only significant difference between methods was for magnesium where the ordinary cokriging estimates slightly improved on the overestimation experienced

across all values for this variable. The only other major difference between methods was that estimates obtained from the principal component kriging using only the first two principal components consistently underestimated the higher values to a greater extent than the other methods.

In conclusion then the estimates obtained from the various kriging methods have confirmed that in the case of isotopic data, in particular when the data are intrinsically correlated, there is little practical benefit to be obtained from cokriging. The true benefits of cokriging are only fully appreciated when the primary variable is undersampled in relation to the secondary variables and those secondary variables are well correlated with the primary variable. A combination of technological advances has lessened the practical and computational disadvantages of multivariate geostatistical methods. There have been marked improvements in the development of geostatistical software that allows interactive modelling simultaneously of both direct and cross semivariograms. In addition there are now available automated iterative procedures, such as those offered in the software package AGROMET, for determining the positive semi-definite coregionalisation matrices. With increased processing power the computational expense solving of large kriging matrices is becoming less of an issue.

This study also raised the question as to the benefits of using principal component kriging. As this technique relies on the data being intrinsically correlated we are able to individually model and krige the principal components. However if the data are intrinsically correlated then there is no benefit in using cokriging as these estimates are equivalent to the kriging estimates. Either way we are required to individually model and krige either N_v variables or N_v principal components. From the estimates obtained in this study using principal component kriging with both four and two principal components there was little difference in the results. This would suggest then that the real benefit in using principal component kriging is when we are able to reduce the number of principal components retained, effectively reducing the number of variables to be modelled and estimated. To be of real benefit however one would want the number of principal components to be retained to be significantly less than the number of original variables, yet still account for the majority of the variability of the original data.

REFERENCES

Afifi, A. A., & Clarke, V. (1996). *Computer-Aided Multivariate Analysis*. Boca Raton: Chapman & Hall.

Armstrong, M. (1998). *Basic Linear Geostatistics*. Berlin: Springer-Verlag.

Bleines, C., Deraisme, J., Geffroy, F., Jeannee, N., Perseval, S., Rambert, F., Renard, D., & Touffait, Y. (2001). *ISATIS: Software Manual* (3rd ed). Paris: Geovariances.

Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modelling Spatial Uncertainty*. New York: Wiley-Interscience.

Datta, B. N. (1995). *Numerical Linear Algebra and Applications*. Pacific Grove: Brooks/Cole Publishing Company.

Deutsch, C. V.,& Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide* (Second ed.). New York: Oxford University Press.

Dunteman, G. H. (1984). *Introduction to Multivariate Analysis*. Beverley Hills: Sage Publications, Inc.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Goulard, M., & Voltz, M. (1992). Linear Coregionalization Model: Tools for Estimation and Choice of CrossVariogram Matrix. *Mathematical Geology, 24* (3), 269-286.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* (24), 417-441.

Isaaks, R. M. & Srivastava, E. H. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

Johnson, R. A. & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis* (5th ed.). New Jersey: Prentice Hall.

Journel, A. G. & Huijbregts, C. J. (1978). *Mining Geostatistics*. London: Academic Press Inc.

Kanevski, M., Chernov, S. & Demyanov, V. (1998). *3PLOT* [Computer Software]. Moscow: IBRAE.

Lay, D. C. (1997). *Linear Algebra and its Applications*. Reading: Addison-Wesley.

Matheron, G. (1970). *The Theory of Regionalised Variables and its Applications*. Fasiscule 5, Les Cahiers du Centre Morphologie Mathematique, Ecole des Mines de Paris, Fontainebleau.

Murphy, M., Bloom, L. M., & Mueller, U. A. (2002). Geostatistical Optimisation of Mineral Resource Sampling Costs for a Western Australian Nickel Deposit. *IAMG 2002: Proceedings of the 8th Annual Conference of the International Association for Mathematical Geology*. (Volume 1, pp. 105-110).

Pannatier, Y. (1996). *VARIOWIN: Software for Spatial Data Analysis in 2D* [Computer software]. New York: Springer-Verlag.

Pearson, K. (1901). On lines and planes of closedst fit to systems of points in space. *Philosophical Maagazine, 6*(2), 559-572.

Wackernagel, H. (1998a). *Multivariate Geostatistics*. Berlin: Springer-Verlag.

Wackernagel, H. (1998b). *Principal Component Analysis for autocorrelated data: a geostatistical perspective*. (N-22/98/G). Fontainebleau: Centre de Geostatistique. Retrieved 5 November, 2002 from http://citeseer.nj.nec.com/cache/papers/cs/5260/ftp:zSzzSzcg.ensmp.frzSzpub zSzhanszSzpcaeof.pdf/wackernagel98principal.pdf

APPENDICES

APPENDIX A: PRINCIPAL COMPONENT ANALYSIS OF TRANSFORMED DATA

Here we consider the appropriateness of the transformation of the strongly skewed variables cobalt, magnesium, aluminium and manganese. We firstly review the eigenanalysis of the original data as covered in section 3.4.1

Table A1: Eigenanalysis of the correlation matrix of MM22DEXP

Eigenvalue	4.293	1.184	1.047	0.607	0.326	0.279	0.154	0.109
Proportion	0.537	0.148	0.131	0.076	0.041	0.035	0.019	0.014
Cumulative	0.537	0.685	0.815	0.892	0.932	0.967	0.986	1.000
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
NI	-0.409	0.082	0.133	-0.277	-0.401	-0.739	0.096	0.121
CO	-0.407	0.153	0.368	0.201	-0.258	0.233	-0.213	-0.685
MG	-0.180	0.650	-0.365	-0.571	0.186	0.195	-0.063	-0.095
FE	-0.416	-0.288	-0.168	-0.004	0.343	0.036	0.730	-0.253
AL	-0.243	0.301	-0.565	0.698	-0.062	-0.154	-0.073	0.106
CR	-0.340	-0.422	-0.331	-0.201	-0.510	0.459	-0.130	0.261
ZN	-0.404	-0.320	0.007	-0.055	0.582	-0.172	-0.599	0.065
MN	-0.351	0.305	0.507	0.162	0.130	0.308	0.169	0.600

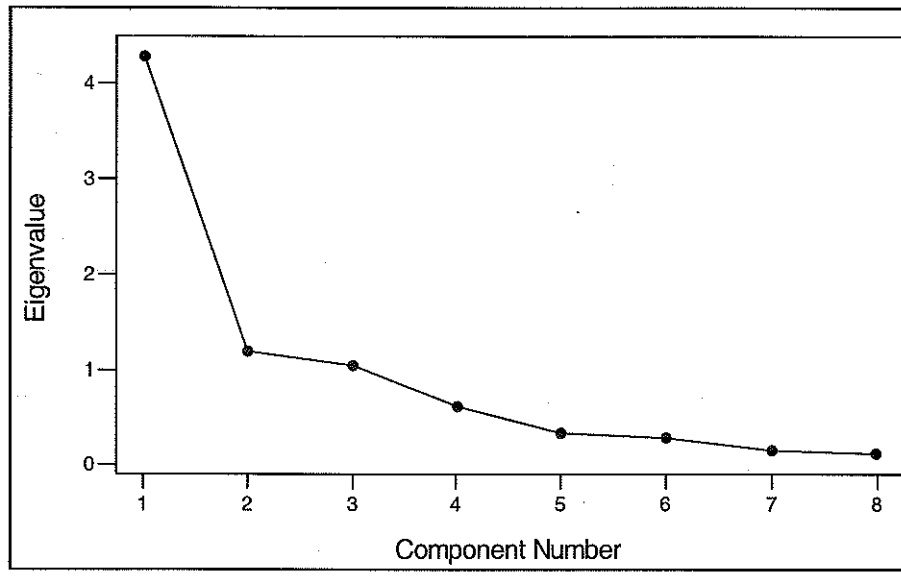


Figure A1: Scree plot of eigenvalues for the principal component analysis of MM22DEXP

We then performed a principal component analysis on the correlation matrix of the raw nickel, iron, chromium and zinc data together with the log transformed cobalt, magnesium, aluminium and manganese data. The results of this principal component analysis are very similar to those obtained using the raw data and are displayed in Table A2. The principal components extracted 58.2%, 15.7%, 8.7%, 6.6%, 4.0%, 3.8%, 1.9% and 1.0% of the total variance. Only the first two components have eigenvalues greater than one and together account for only 74% of the variability of the original data. By incorporating the third principal component this value is increased to 82.6% and by the fourth principal component we have accounted for 89.3% of the variability of the original data. As with the untransformed case the scree plot in Figure A2 reveals that the eigenvalues appear to level off after the fifth component. Again the eigenvalue for this component is small, only 0.322, hence it would be of no benefit to retain it. As for the untransformed case if data reduction was the objective only the first four principal components (which account for 89.3% of the variability of the original data) would be retained.

Table A2: Eigenanalysis of the correlation matrix of MM22DEXP with log transformed cobalt, magnesium, aluminium and manganese

Eigenvalue	4.659	1.259	0.694	0.530	0.322	0.306	0.150	0.080
Proportion	0.582	0.157	0.087	0.066	0.040	0.038	0.019	0.010
Cumulative	0.582	0.740	0.826	0.893	0.933	0.971	0.990	1.000
Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
NI	-0.388	0.071	0.230	-0.203	-0.645	0.564	0.031	0.123
TRCO	-0.412	0.198	0.141	0.377	0.212	0.178	-0.108	-0.735
TRMG	-0.234	0.609	-0.131	-0.667	0.094	-0.293	0.053	-0.124
FE	-0.386	-0.366	-0.022	-0.121	-0.119	-0.359	-0.746	0.041
TRAL	-0.295	0.086	-0.869	0.307	-0.175	0.019	0.111	0.117
CR	-0.327	-0.450	-0.137	-0.364	0.587	0.397	0.173	0.089
ZN	-0.382	-0.325	0.215	0.061	-0.191	-0.523	0.619	-0.074
TRMN	-0.368	0.369	0.300	0.356	0.325	-0.058	-0.046	0.633

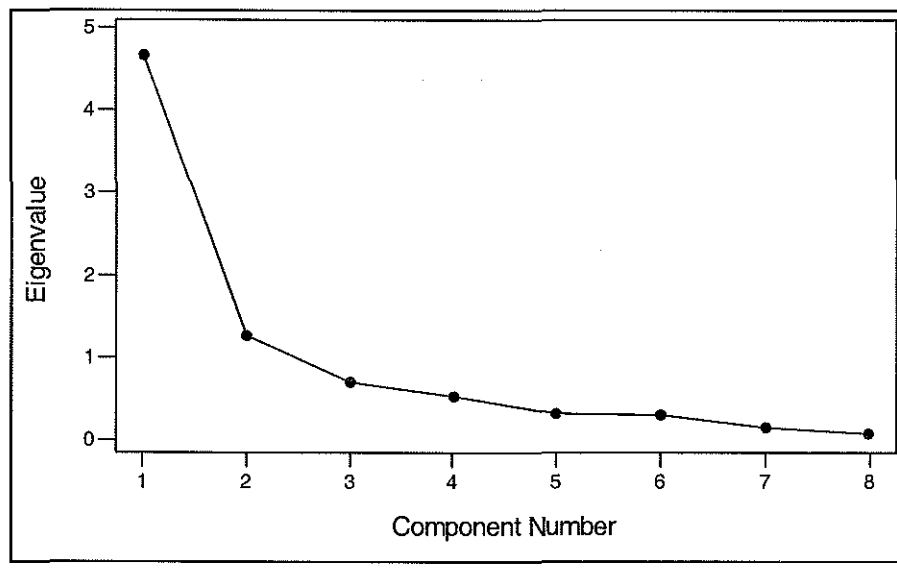


Figure A2: Scree plot of eigenvalues for the principal component analysis of MM22DEXP with log transformed cobalt, magnesium, aluminium and manganese

We next consider the viability of the transformation of the highly skewed variables cobalt, magnesium, aluminium and manganese by assessing the effect on the principal component analyses. The eigenvalues in each are case very close in value with the cumulative proportion of variability almost identical by the fourth component. The eigenvalues of the first and second principal components are somewhat higher for the transformed data while the third eigenvalue for this data set is somewhat lower. There have been numerous small changes in some of the elements of the coefficient vectors, namely in the first analysis nickel, cobalt, iron and zinc dominate the first principal component, all with fairly equal weighting. In the second analysis, the transformed cobalt receives the greatest weight while nickel, iron, zinc and transformed manganese have lower, fairly even values. Similar changes have occurred in the subsequent principal components, though they all appear to be small.

Figures A3 and A4 display the scatter plots of the scores of the first four principal components for each analysis. There does appear to be some slight curvature in some of the plots in Figure A3, in particular those between PCS1, PCS2 and PCS3, suggesting that the data may benefit from a transformation. By comparison, the scatter plots in Figure A4 do appear to be less correlated and circular in shape, suggesting that the transformation has somewhat improved the analysis. However given that the results of the two principal component analyses are very similar, the curvature in the plots in Figure A3 is not sufficiently marked and the dangers of exaggeration of estimation errors we feel that the data do not warrant the transformation.

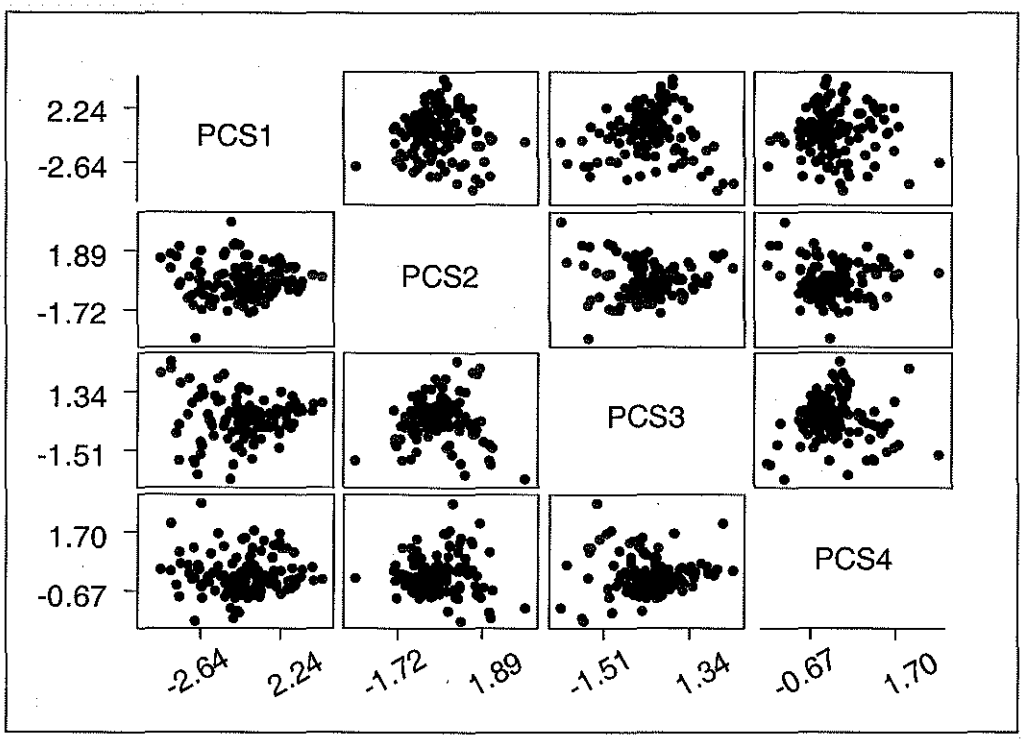


Figure A3: Scatter plots of the principal component scores of MM22DEXP

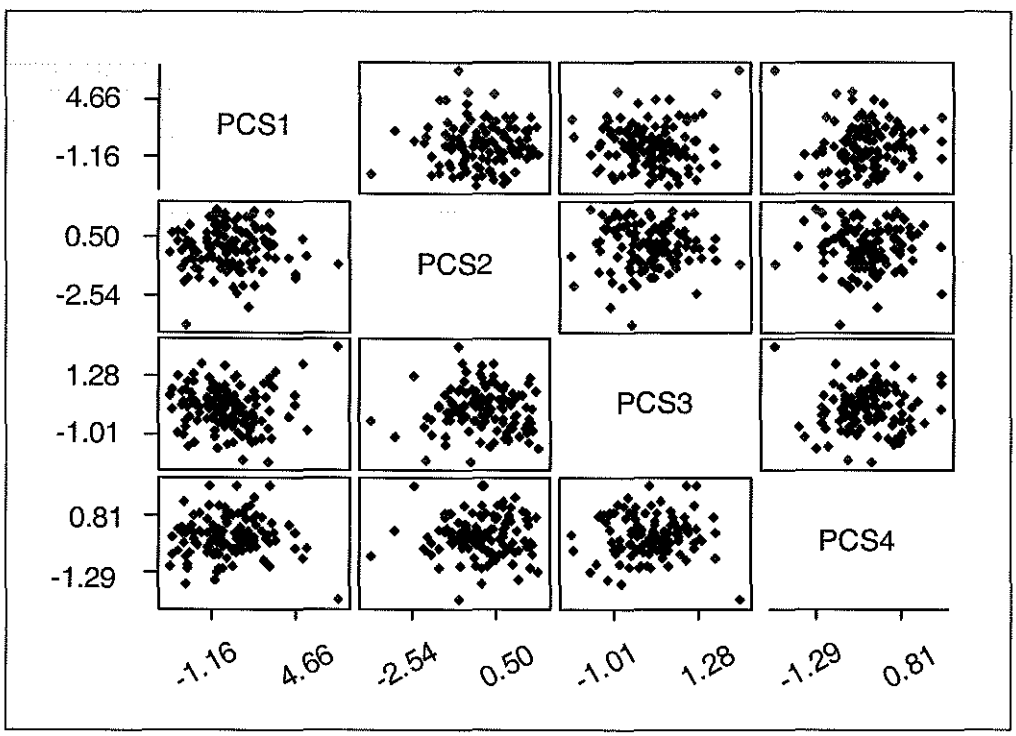
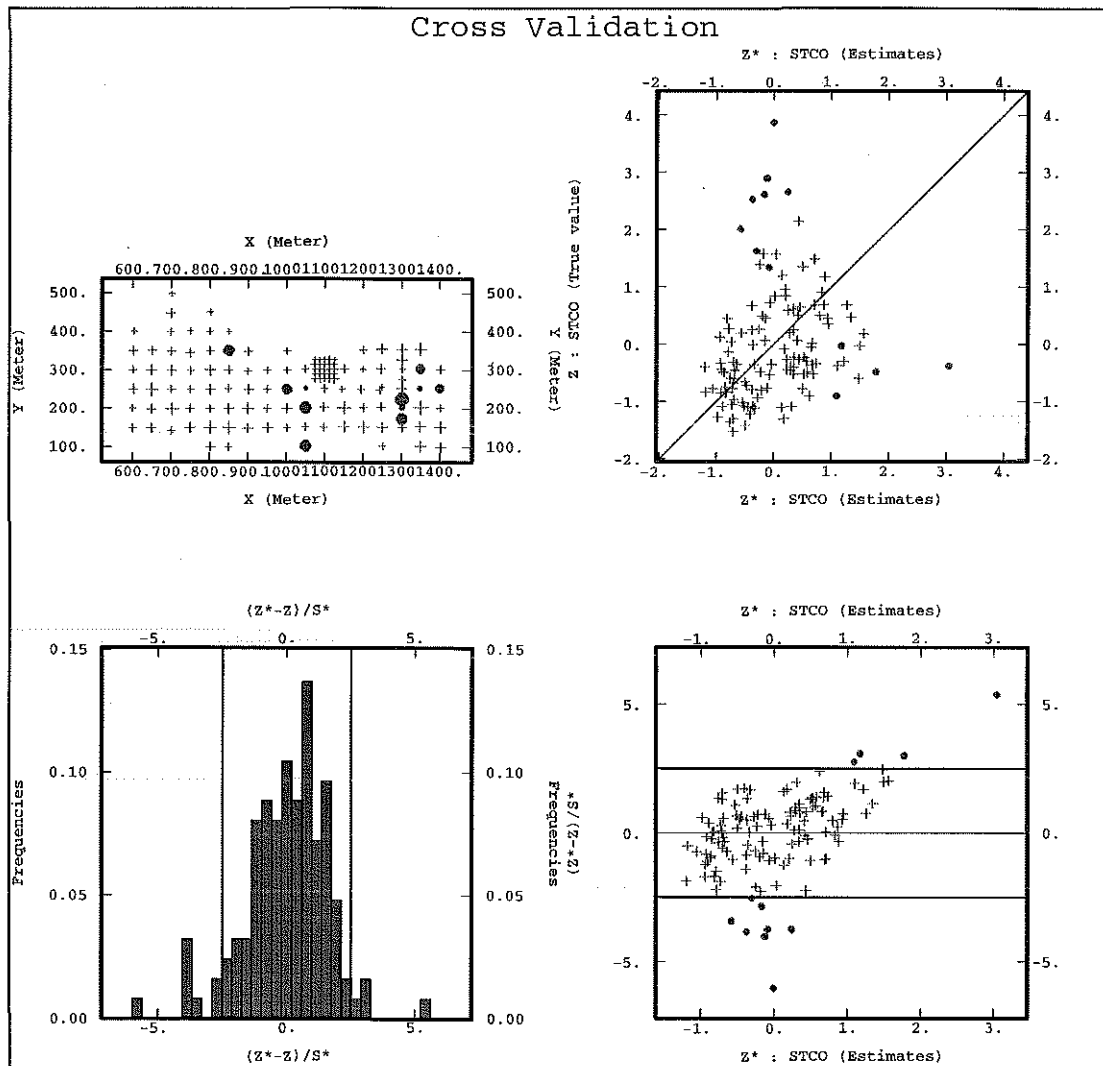


Figure A4: Scatter plots of the principal component scores of MM22DEXP with log transformed cobalt, magnesium, aluminium and manganese

APPENDIX B: CROSS VALIDATION

Cokriging Cross Validation of Cobalt, Iron and Zinc from MM22DHC4



Isatis

data/stexp

- Variable #1 : STNI
- Variable #2 : STCO
- Variable #3 : STFE
- Variable #4 : STZN

Standard Parameter File for Model: hc4
Standard Parameter File for Neighborhood: neighborhood
Cross validation statistics based on 124 test data

	Mean	Variance
Error	0.01820	1.24065
Std. Error	0.01552	2.56868

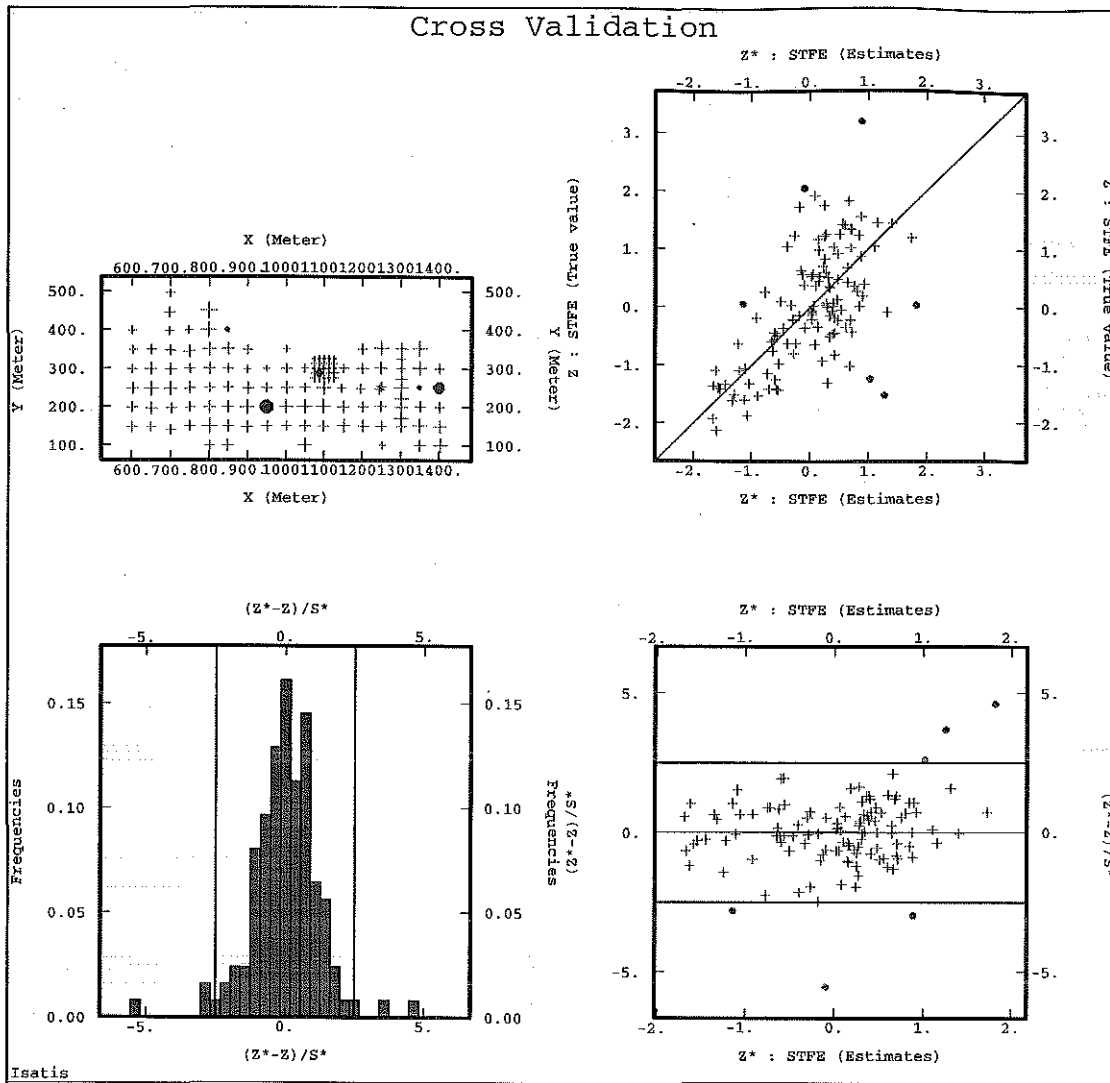
Cross validation statistics based on 112 robust data

	Mean	Variance
Error	0.12837	0.63143
Std. Error	0.15833	1.25607

A data is robust when its Standardized Error
lies between -2.500000 and 2.500000

Ellen M. Bandarian
May 13 2003 13:29:45
ellen

Figure B1: Cross validation of cobalt



Isatis

data/stexp

- Variable #1 : STNI
- Variable #2 : STCO
- Variable #3 : STFE
- Variable #4 : STZN

Standard Parameter File for Model: hc4
Standard Parameter File for Neighborhood: neighborhood
Cross validation statistics based on 124 test data

	Mean	Variance
Error	0.01315	0.65802
Std. Error	0.01033	1.54736

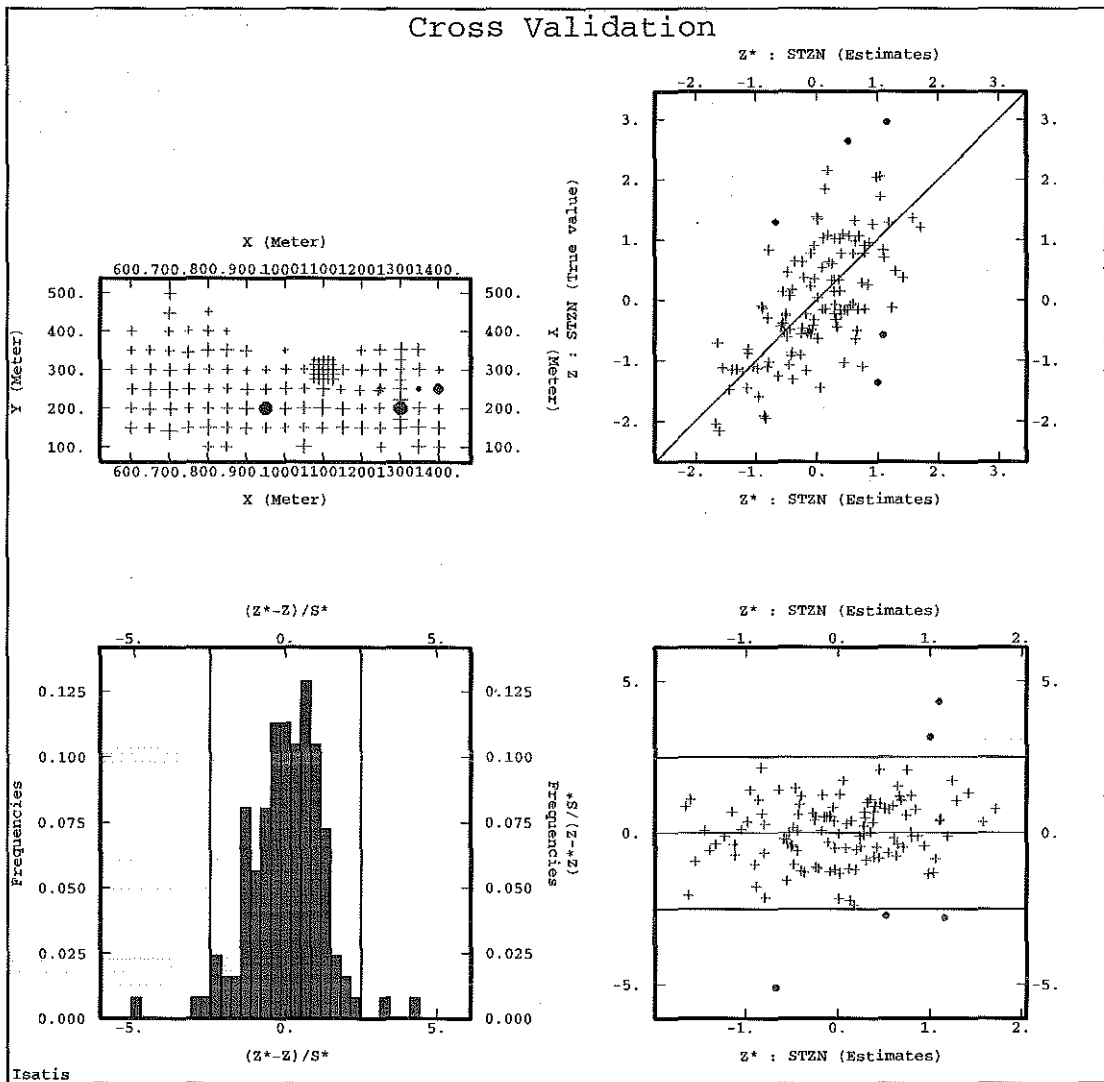
Cross validation statistics based on 118 robust data

	Mean	Variance
Error	0.00384	0.45980
Std. Error	0.01486	0.87059

A data is robust when its Standardized Error
lies between -2.500000 and 2.500000

Ellen M. Bandarian
May 13 2003 13:28:53
ellen

Figure B2: Cross validation of iron



Isatis

data/stexp
- Variable #1 : STNI
- Variable #2 : STCO
- Variable #3 : STFE
- Variable #4 : STZN

Ellen M. Bandarian
May 13 2003 13:29:18
ellen

Standard Parameter File for Model: hc4
Standard Parameter File for Neighborhood: neighborhood
Cross validation statistics based on 124 test data

	Mean	Variance
Error	0.01876	0.64322
Std. Error	0.01526	1.50370

Cross validation statistics based on 119 robust data

	Mean	Variance
Error	0.03477	0.50105
Std. Error	0.04257	0.97769

A data is robust when its Standardized Error
lies between -2.500000 and 2.500000

Figure B3: Cross validation of zinc

Cokriging Cross Validation of Cobalt, Magnesium and Iron MM22DTOP4

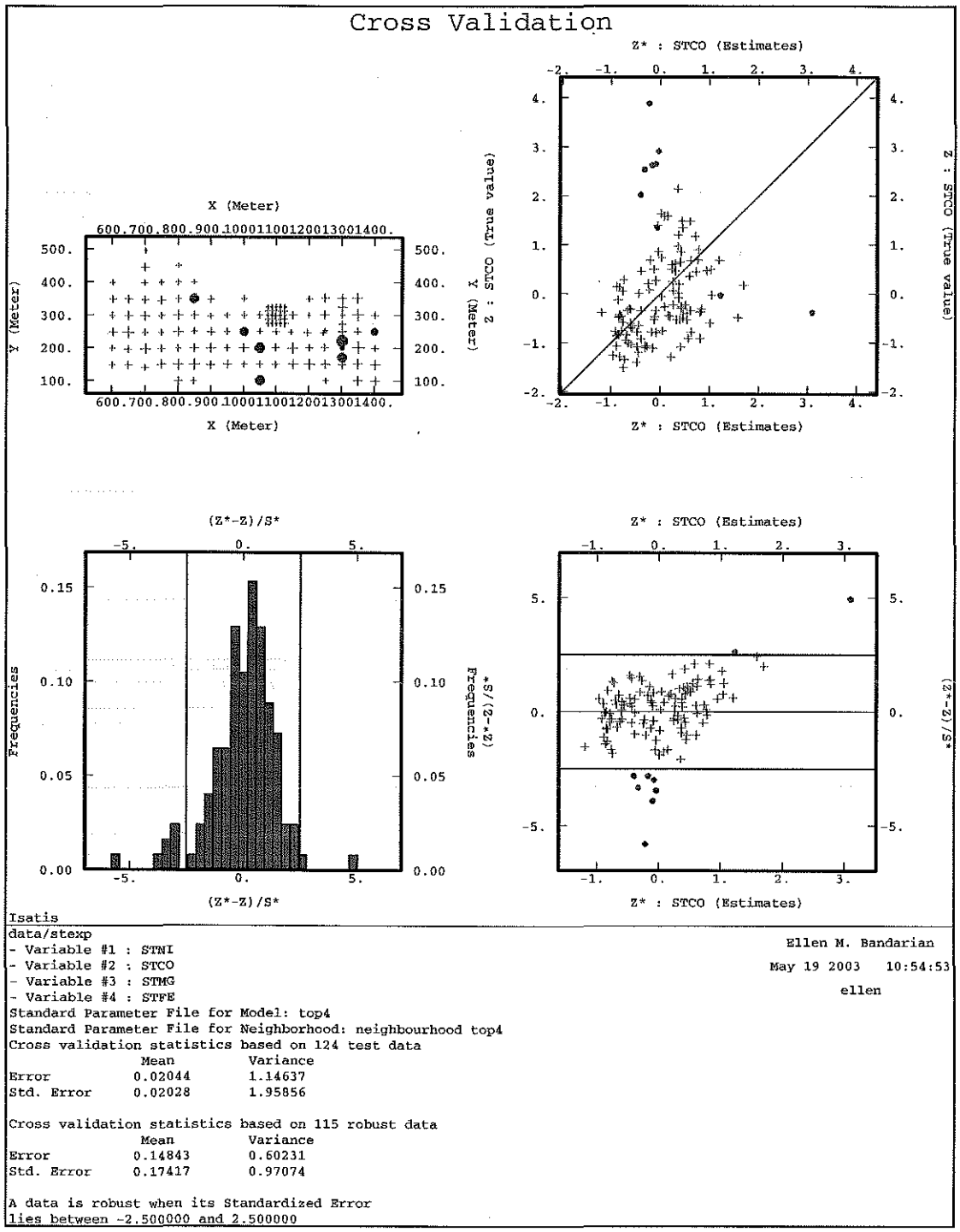


Figure B4: Cross validation of cobalt

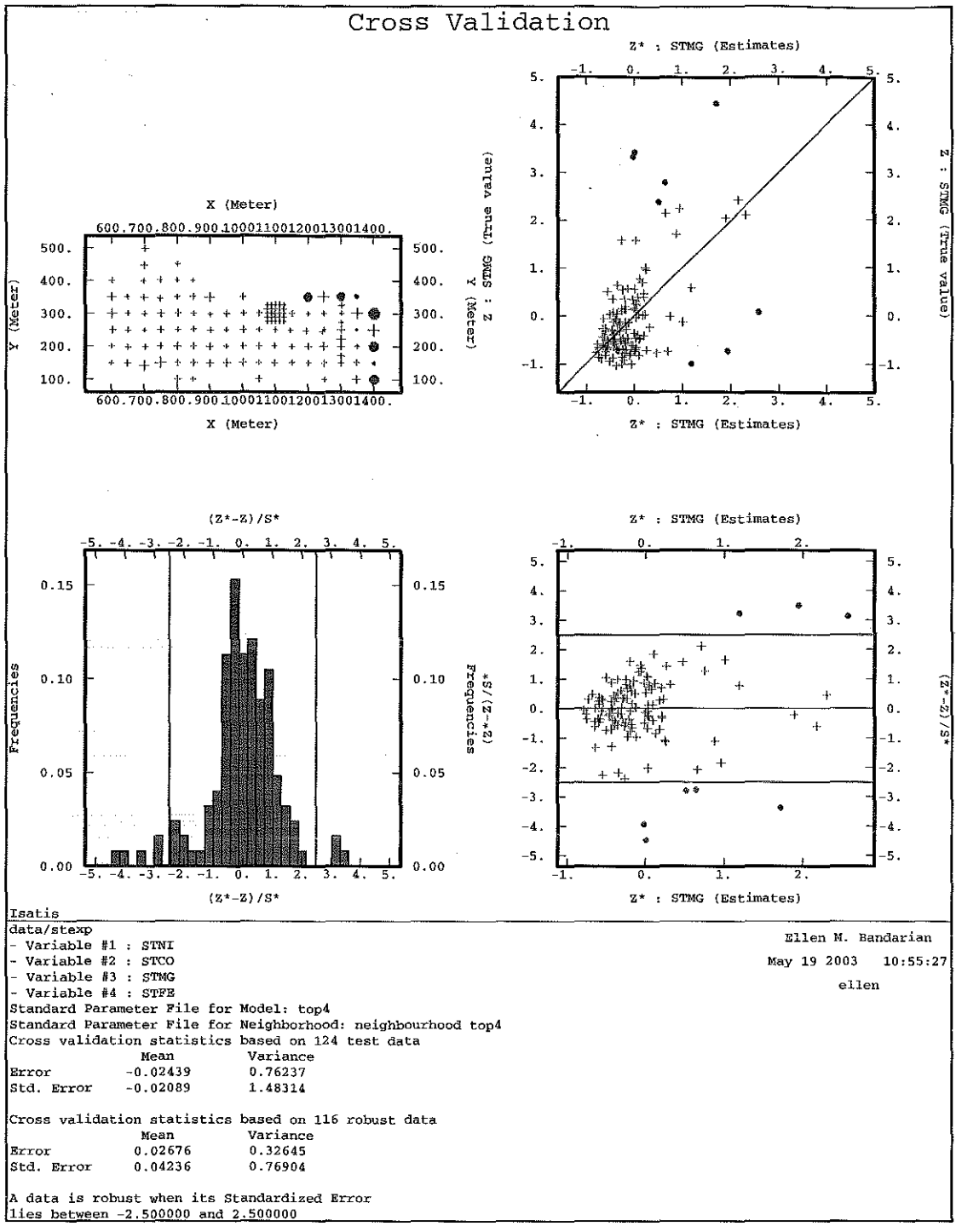


Figure B5: Cross validation of magnesium

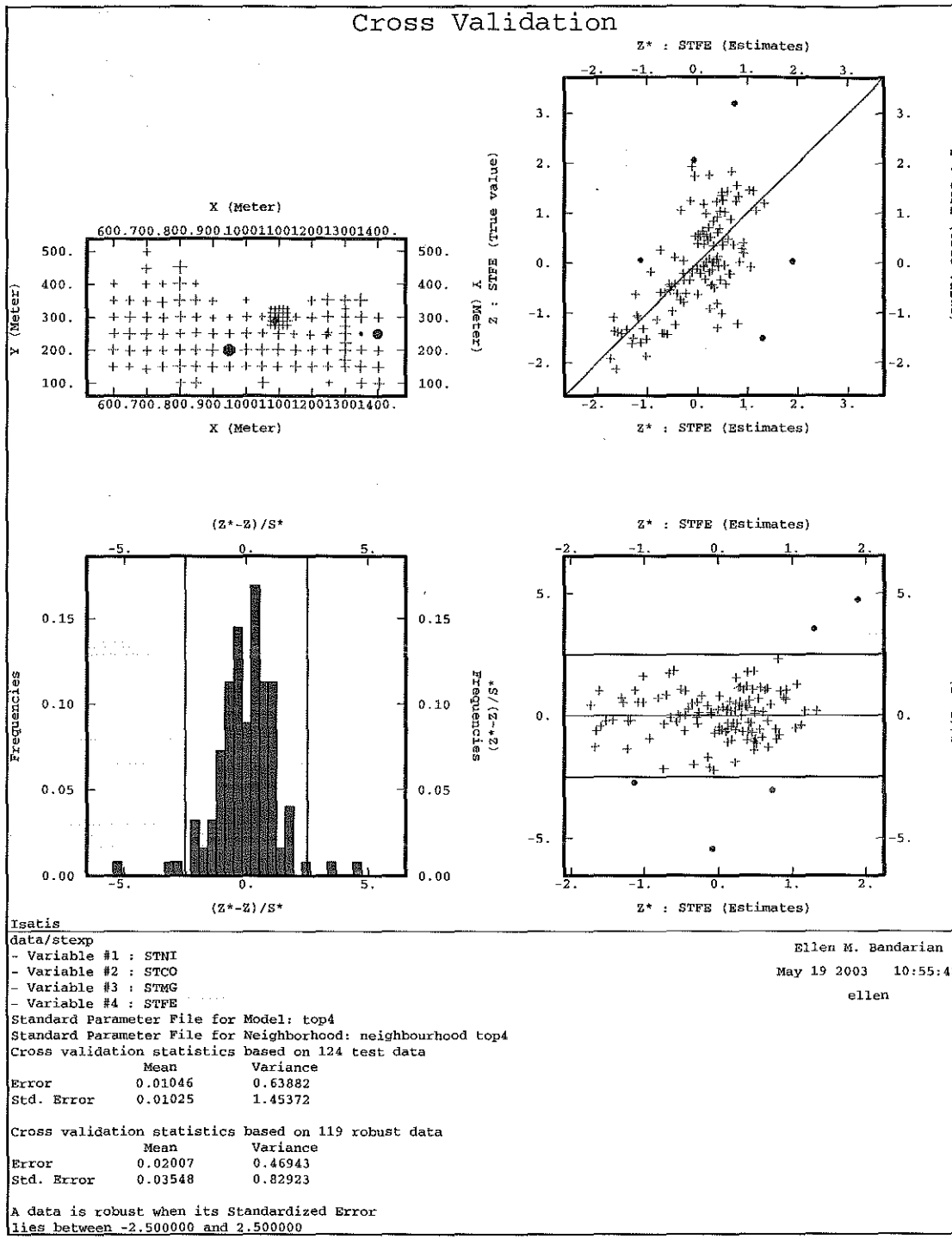


Figure B6: Cross validation of iron

Ordinary Kriging Cross Validation of Principal Components

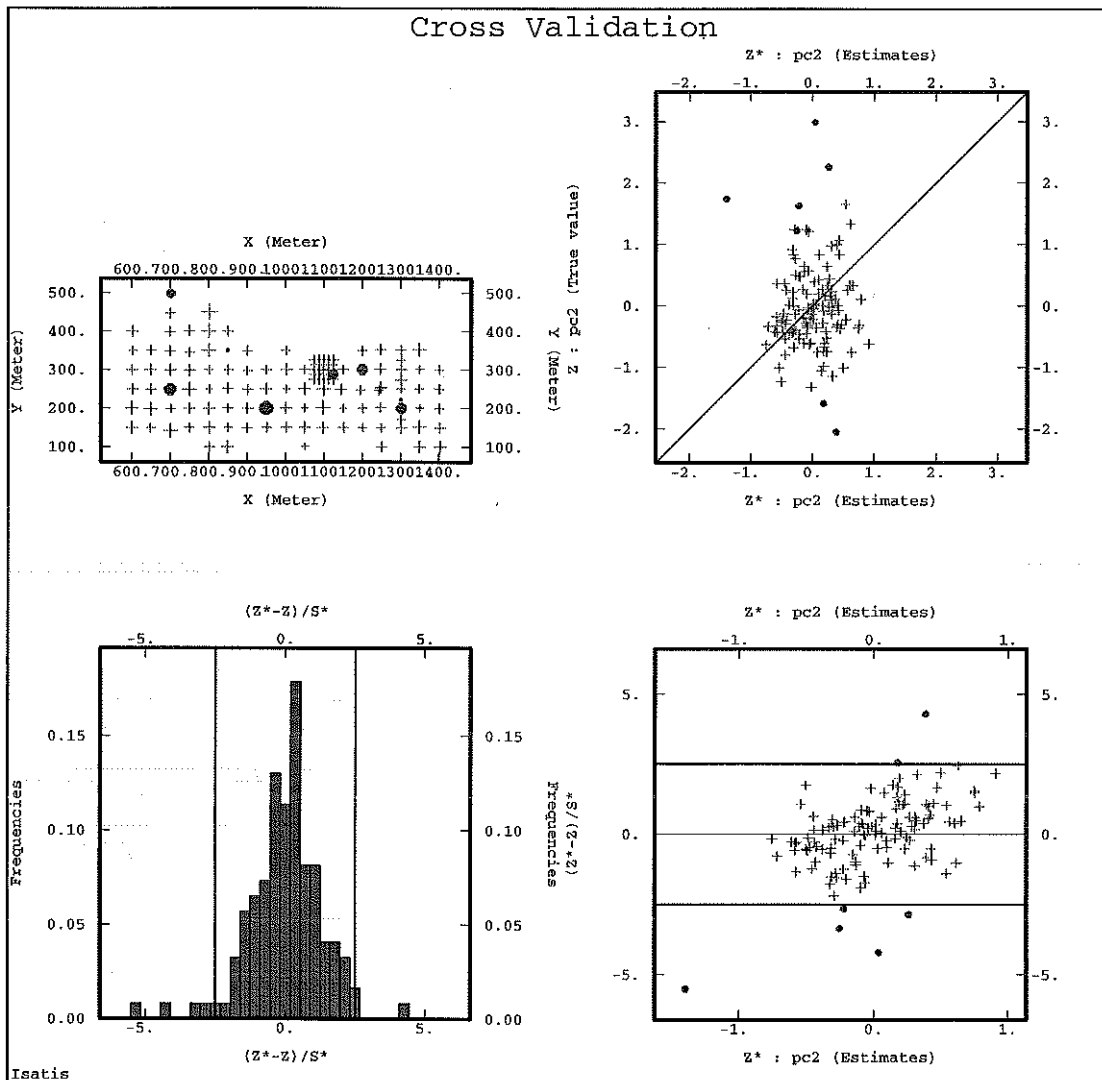


Figure B7: Cross validation of PC2

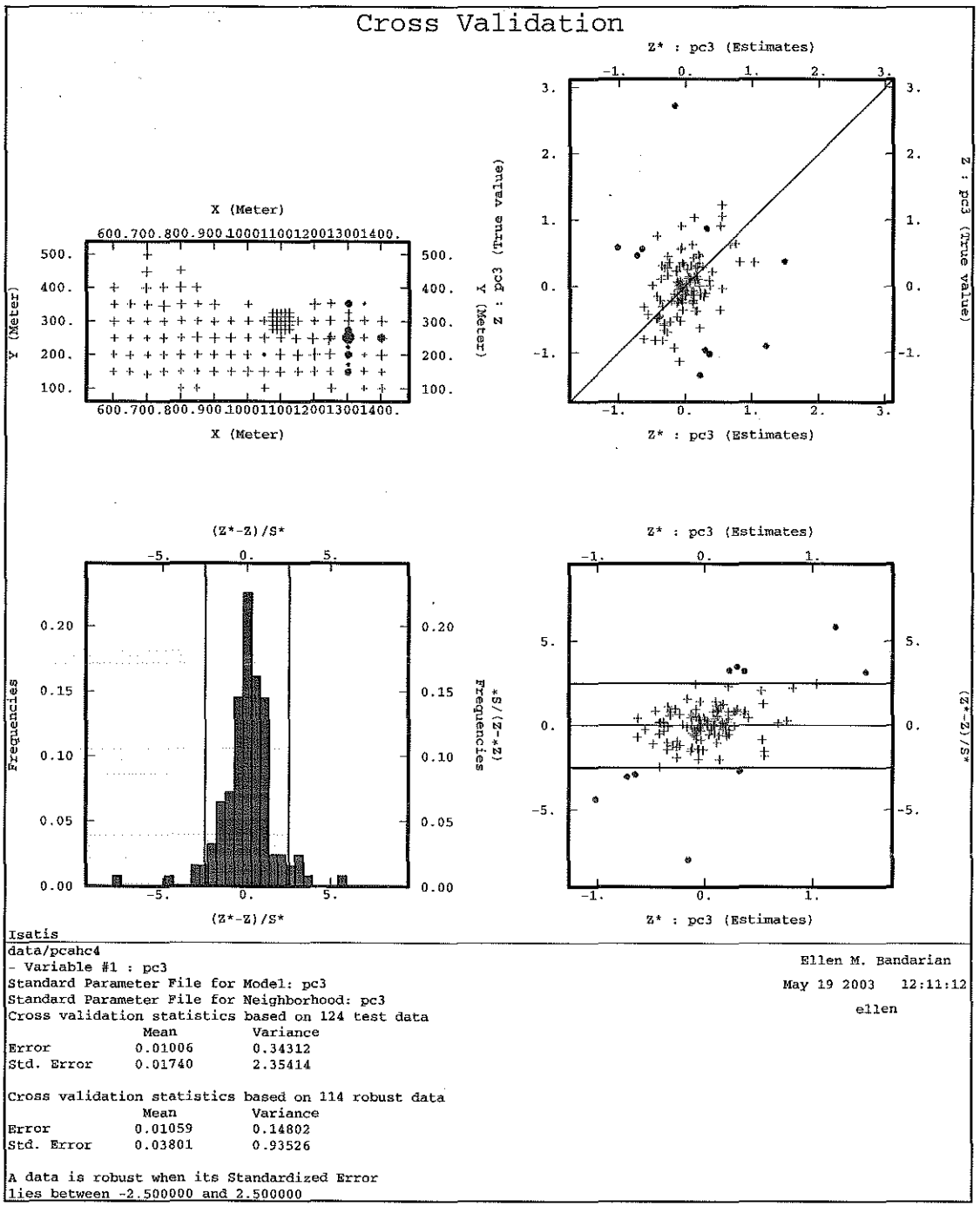


Figure B8: Cross validation of PC3

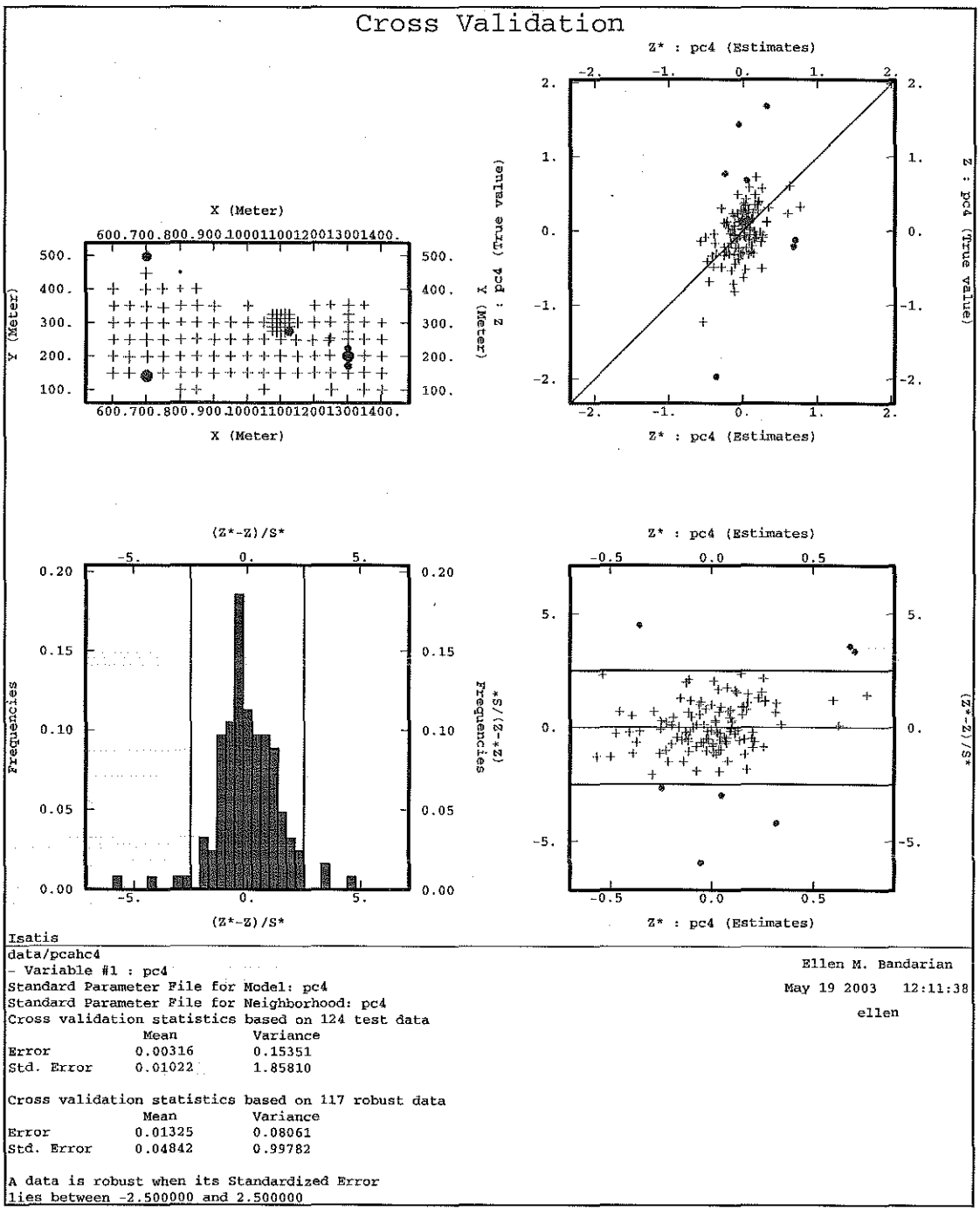


Figure B9: Cross validation of PC4

Ordinary Kriging Cross Validation of Cobalt, Magnesium, Iron and Zinc

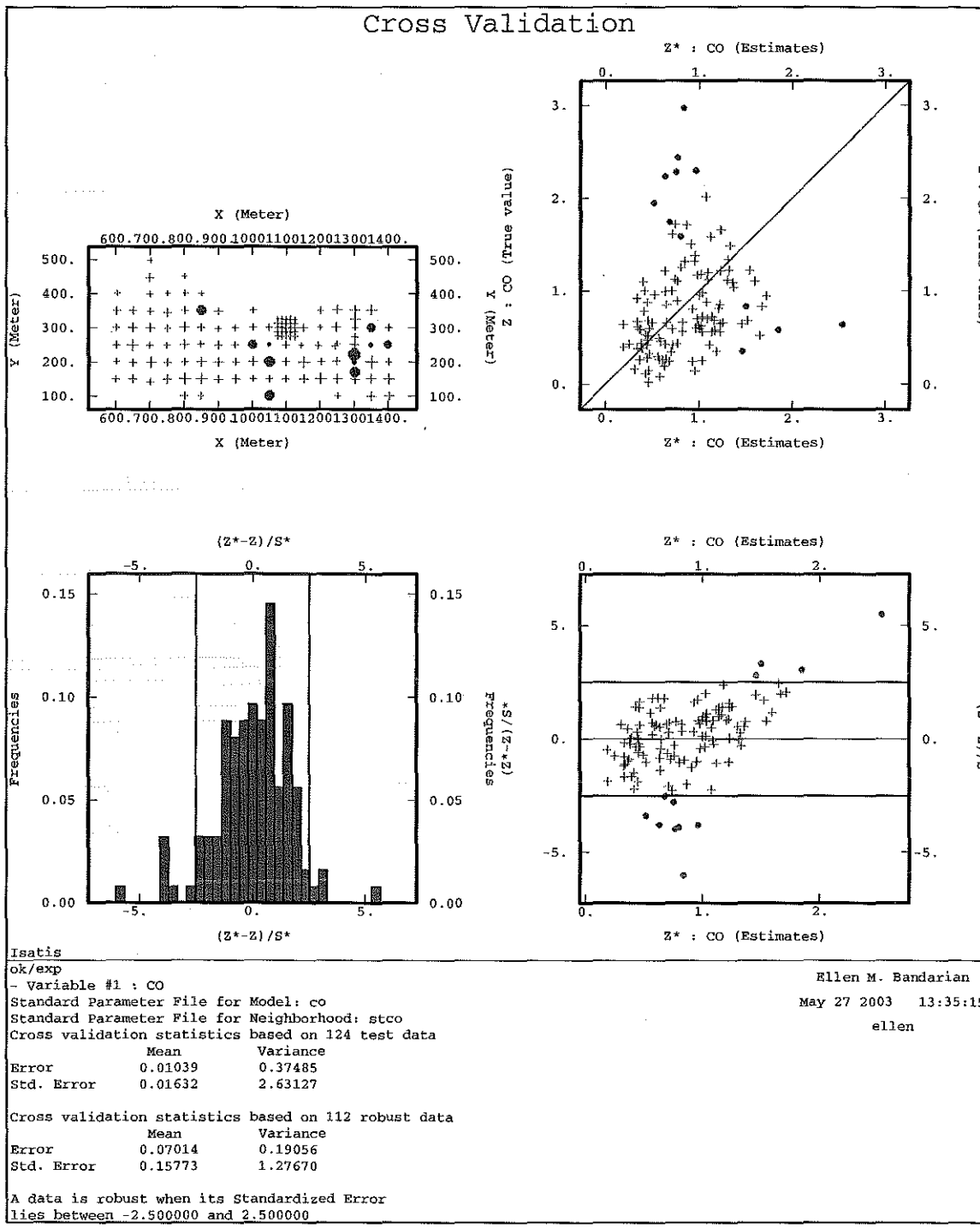
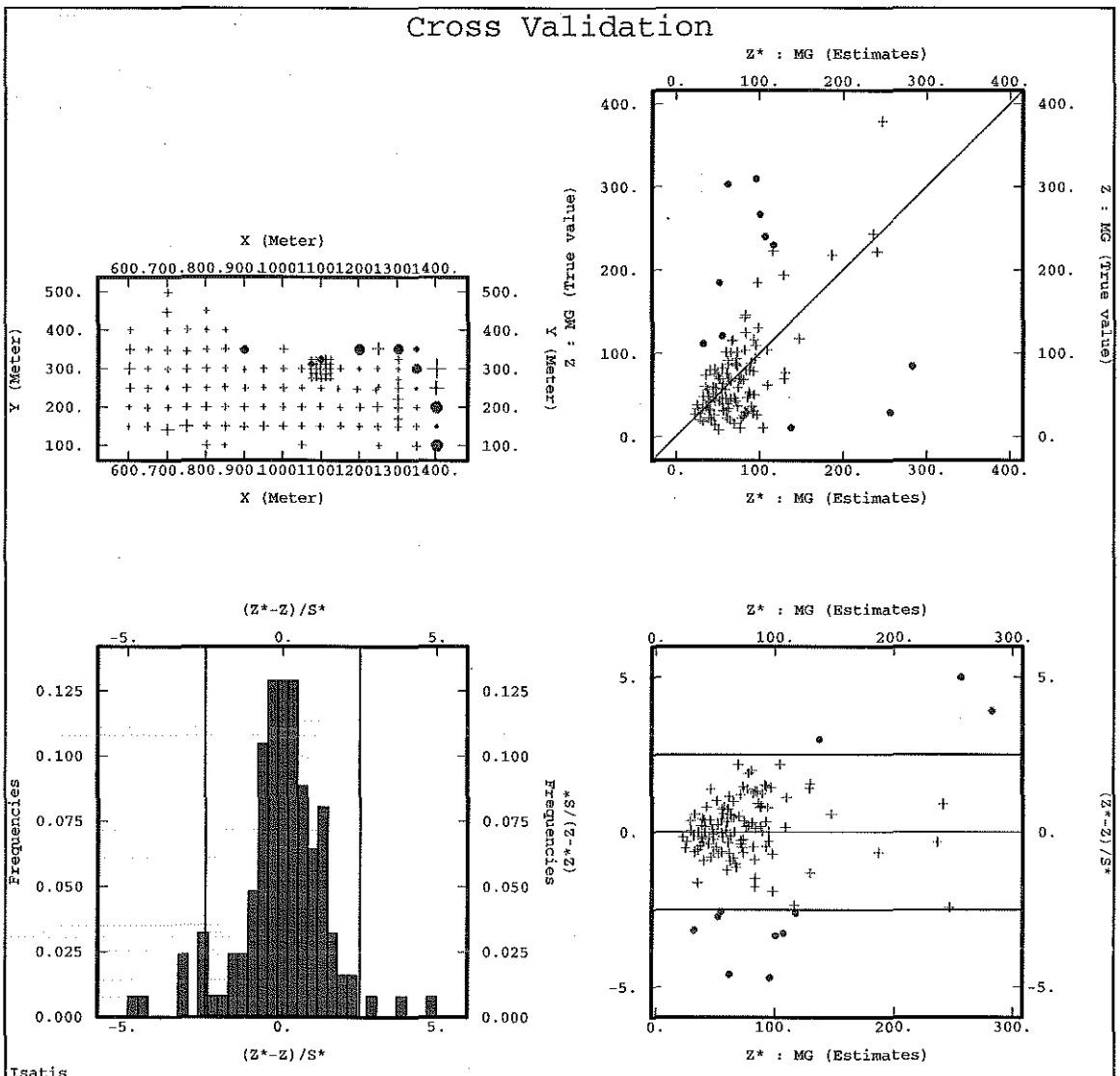


Figure B10: Cross validation of cobalt



Isatis

ok/exp

- Variable #1 : MG

Standard Parameter File for Model: mg

Standard Parameter File for Neighborhood: stmg

Cross validation statistics based on 124 test data

	Mean	Variance
Error	-1.63389	3629.50520
Std. Error	-0.01960	1.94218

Cross validation statistics based on 113 robust data

	Mean	Variance
Error	3.43792	1353.50795
Std. Error	0.11234	0.83542

A data is robust when its Standardized Error
lies between -2.500000 and 2.500000

Ellen M. Bandarian

May 27 2003 13:34:44

ellen

Figure B11: Cross validation of magnesium

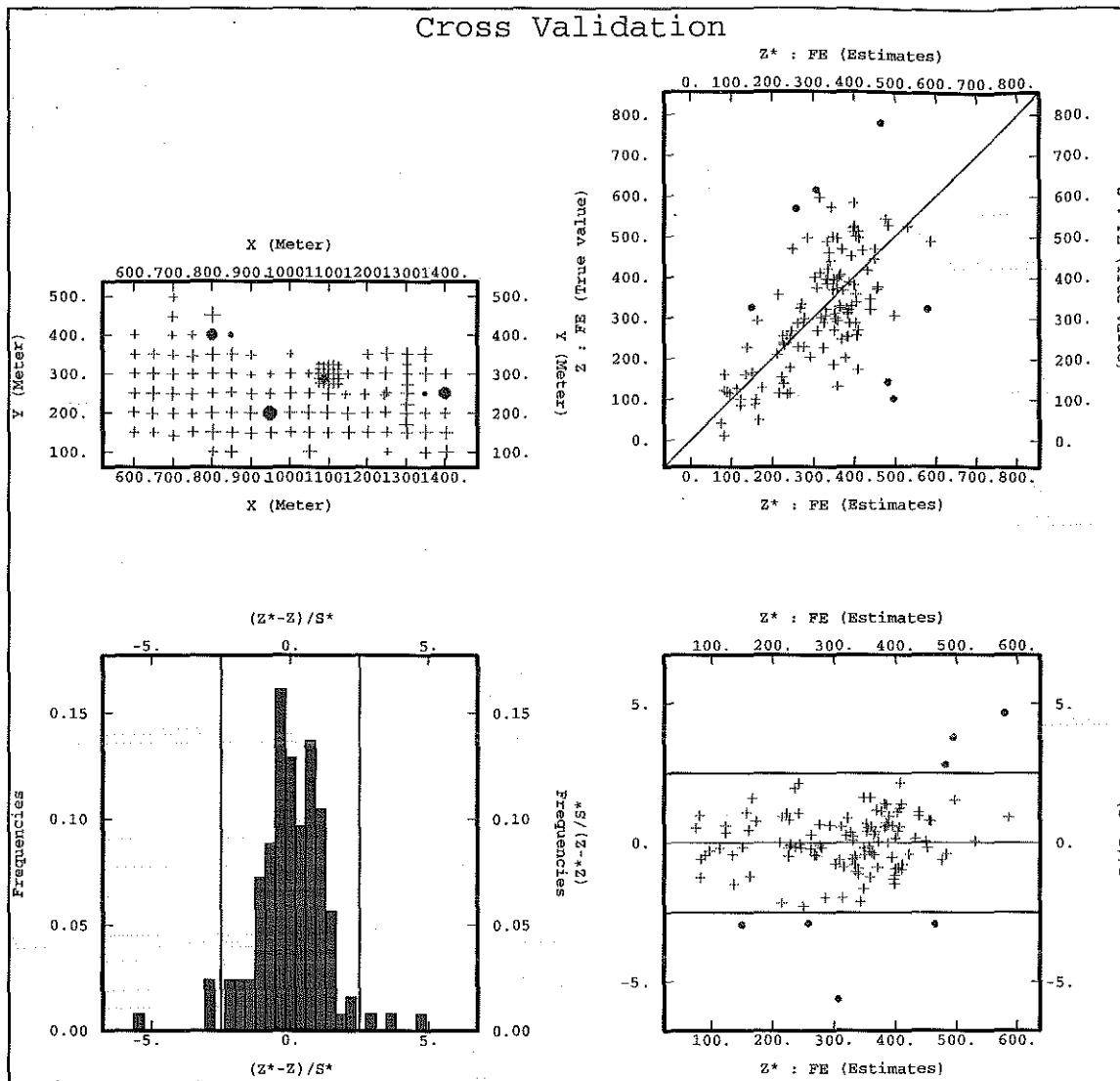


Figure B12: Cross validation of iron

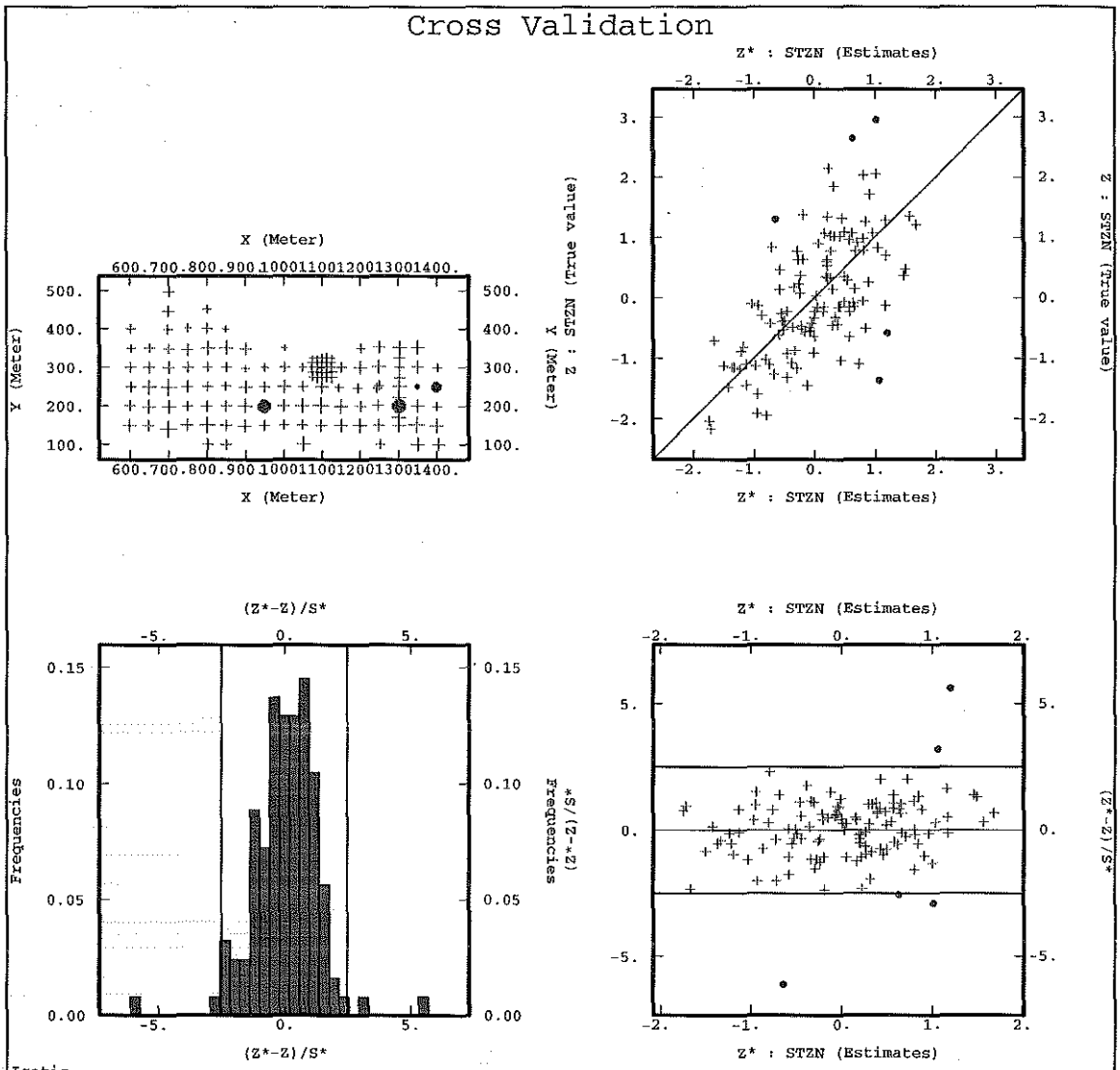


Figure B13: Cross validation of zinc

APPENDIX C: PARAMETER FILES

Ordinary Cokriging of MM2DHC4

| Kriging procedure |

Data File Information:

Directory = data

File = stexp

Variable(s) = STNI

Variable(s) = STCO

Variable(s) = STFE

Variable(s) = STAN

Target File Information:

Directory = data

File = hc4cokrig

Variable(s) = stniest

Variable(s) = nisd

Variable(s) = stcoest

Variable(s) = cosd

Variable(s) = stfeest

Variable(s) = fesd

Variable(s) = stznest

Variable(s) = znsd

Type = POINT (1717 points)

Model Name = hc4

Neighborhood Name = neighbourhood hc4 - MOVING

Successfully processed = 1717

Written to the disk = 1717

Neighbourhood parameters

==========

Type = MOVING

X-Radius = 220.00m

Y-Radius = 120.00m

Z-Radius = 228.12m

Rotation angle around Z = 0 degrees

Rotation angle around Y = 0 degrees

Rotation angle around X = 0 degrees

Minimum # of information = 4

Optimum # of information = 4

Number of angular sectors = 4

Automatic Sorting on

X-Mesh of sorting grid = 228.12m

Y-Mesh of sorting grid = 228.12m

Z-Mesh of sorting grid = 228.12m

No Heterotopic Search

No Minimum Distance used between Data Points

No Maximum Distance without Samples

No Maximum Number of Consecutive Empty Sectors

Model : Covariance part

==========

Number of variables = 4

- Variable 1 : STNI

- Variable 2 . STCO

- Variable 3 : STPE

- Variable 4 : STZN

Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.0500	0.0360	0.0320	0.0320
Variable 2	0.0360	0.0500	0.0290	0.0300
Variable 3	0.0320	0.0290	0.0500	0.0415
Variable 4	0.0320	0.0300	0.0415	0.0500

S2 : Spherical - Range = 100.00m

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.5000	0.3600	0.3200	0.3200
Variable 2	0.3600	0.5000	0.2900	0.3000
Variable 3	0.3200	0.2900	0.5000	0.4150

Variable 4	0.3200	0.3000	0.4150	0.5000

S3 : Spherical - Range = 100.00m

Directional Scales = (200.00m, 100.00m, 100.00m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.4500	0.3240	0.2880	0.2880
Variable 2	0.3240	0.4500	0.2610	0.2700
Variable 3	0.2880	0.2610	0.4500	0.3735
Variable 4	0.2880	0.2700	0.3735	0.4500

Ordinary Cokriging of MM22DTOP4

| Kriging procedure |

==========

Data File Information:

Directory = data

File = stexp

Variable(s) = STNI

Variable(s) = STCO

Variable(s) = STMG

Variable(s) = STFE

Target File Information:

Directory = data

File = top4cokrig

Variable(s) = stniest

Variable(s) = nisd

Variable(s) = stcoest

Variable(s) = cosd

Variable(s) = stmgest

Variable(s) = mgsd

Variable(s) = stfeest

Variable(s) = fesd

Type = POINT (1717 points)

Model Name = top4

Neighborhood Name = neighbourhood top4 - MOVING

Successfully processed = 1717

Written to the disk = 1717

Neighbourhood parameters

==========

Type = MOVING

X-Radius = 220.00m

Y-Radius = 100.00m

Z-Radius = 228.12m

Rotation angle around Z = 0 degrees

Rotation angle around Y = 0 degrees

Rotation angle around X = 0 degrees

Minimum # of information = 3

Optimum # of information = 4

Number of angular sectors = 4

Automatic Sorting on

X-Mesh of sorting grid = 228.12m

Y-Mesh of sorting grid = 228.12m

Z-Mesh of sorting grid = 228.12m

No Heterotopic Search

No Minimum Distance used between Data Points

No Maximum Distance without Samples

No Maximum Number of Consecutive Empty Sectors

Model : Covariance part

==========

Number of variables = 4

- Variable 1 : STNI

- Variable 2 : STCO

- Variable 3 : STMG

- Variable 4 : STFE

Number of basic structures = 3

S1 : Nugget effect

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.1000	0.0500	0.0300	0.0000
Variable 2	0.0500	0.0900	0.0500	0.0000
Variable 3	0.0300	0.0500	0.0800	0.0200
Variable 4	0.0000	0.0000	0.0200	0.0500

S2 : Spherical - Range = 75.00m

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.2700	0.1700	0.0300	0.1900
Variable 2	0.1700	0.5100	0.0000	0.1700
Variable 3	0.0300	0.0000	0.1000	0.0400
Variable 4	0.1900	0.1700	0.0400	0.4000

S3 : Spherical - Range = 140.00m

Directional Scales = (200.00m, 140.00m, 140.00m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Variance-Covariance matrix :

	Variable 1	Variable 2	Variable 3	Variable 4
Variable 1	0.6300	0.5000	0.3000	0.4500
Variable 2	0.5000	0.4200	0.1800	0.4000
Variable 3	0.3000	0.1800	0.8200	0.1200
Variable 4	0.4500	0.4000	0.1200	0.5500

Ordinary Kriging of the Principal Component Scores

Legend: pc1 (pc2) (pc3) (pc4)

| Kriging procedure |

==========

Data File Information:

Directory = data (data) (data) (data)

File = pcahc4 (pcahc4) (pcahc4) (pcahc4)

Variable(s) = pc1 (pc2) (pc3) (pc4)

Target File Information:

Directory = pc1est (pc2est) (pc3est) (pc4est)

File = pc1est (pc2est) (pc3est) (pc4est)

Variable(s) = pc1est (pc2est) (pc3est) (pc4est)

Variable(s) = pc1sd (pc2sd) (pc3sd) (pc4sd)

Type = POINT (1717 points) (1717 points) (1717 points) (1717 points)

Model Name = pc1 (pc2) (pc3) (pc4)

Neighborhood Name = pc1 (pc2) (pc3) (pc4) - MOVING

Successfully processed = 1717 (1717) (1717) (1717)

Written to the disk = 1717 (1717) (1717) (1717)

Neighbourhood parameters

==========

Type = MOVING

X-Radius = 220.00m (120.00m)(190.00m) (250.00m)

Y-Radius = 120.00m (120.00m) (100.00m) (250.00m)

Z-Radius = 228.12m (228.12m) (228.12m) (228.12m)

Rotation angle around Z = 0 degrees (25 degrees) (0 degrees) (0 degrees)

Rotation angle around Y = 0 degrees (0 degrees)(0 degrees) (0 degrees)

Rotation angle around X = 0 degrees (0 degrees)(0 degrees) (0 degrees)

Minimum # of information = 3 (3) (3) (3)

Optimum # of information = 4 (4) (4) (4)

Number of angular sectors = 4 (4) (4) (4)

Automatic Sorting on

X-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

Y-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

Z-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

No Heterotopic Search

No Minimum Distance used between Data Points

No Maximum Distance without Samples

No Maximum Number of Consecutive Empty Sectors

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : pc1

Number of basic structures = 3

S1 : Nugget effect

Sill = 0.1000

S2 : Spherical - Range = 90.00m

Sill = 0.4700

S3 : Spherical - Range = 90.00m

Sill = 2.4700

Directional Scales = (200.00m, 90.00m, 90.00m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : pc2

Number of basic structures = 2

S1 : Nugget effect

Sill = 0.0800

S2 : Spherical - Range = 85.00m

Sill = 0.4700

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : pc3

Number of basic structures = 3

S1 : Nugget effect

Sill = 0.0100

S2 : Spherical - Range = 50.00m

Sill = 0.0500

S3 : Spherical - Range = 98.00m

Sill = 0.2150

Directional Scales = (140.00m, 98.00m, 98.00m)

Local Rot (mathematician)= (25.00, 0.00, 0.00)

Local Rot (geologist) = (65.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : pc4

Number of basic structures = 3

S1 : Nugget effect

Sill = 0.0180

S2 : Spherical - Range = 100.00m

Sill = 0.0650

S3 : Spherical - Range = 250.00m

Sill = 0.0920

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Ordinary Kriging of Nickel, Cobalt, Magnesium, Iron and Zinc

Legend: nickel (cobalt) (magnesium) (iron) (zinc)

| Kriging procedure |

==========

Data File Information:

Directory = ok (ok) (ok) (ok) (ok)

File = exp (exp) (exp) (exp) (exp)

Variable(s) = ni (co) (mg) (fe) (zn)

Target File Information:

Directory = ok (ok) (ok) (ok) (ok)

File = niest (coest) (mgest) (feest) (znest)

Variable(s) = niest (coest) (mgest) (feest) (znest)

Variable(s) = nisd (cosd) (mgsd) (fesd) (znsd)

Type = POINT (1717 points) (1717 points) (1717 points) (1717 points) (1717 points)

Model Name = ni (co) (mg) (fe) (zn)

Neighborhood Name = ni (co) (mg) (fe) (zn) - MOVING

Successfully processed = 1717 (1717) (1717) (1717) (1717)

Written to the disk = 1717 (1717) (1717) (1717) (1717)

Neighbourhood parameters

==========

Type = MOVING

X-Radius = 180.00m (220.00m)(300.00m) (210.00m)(220.00m)

Y-Radius = 110.00m (120.00m) (160.00m) (110.00m)(110.00m)

Z-Radius = 228.12m (228.12m) (228.12m) (228.12m)(228.12m)

Rotation angle around Z = 0 degrees (0 degrees) (120 degrees) (0 degrees) (0 degrees)

Rotation angle around Y = 0 degrees (0 degrees)(0 degrees) (0 degrees) (0 degrees)

Rotation angle around X = 0 degrees (0 degrees)(0 degrees) (0 degrees) (0 degrees)

Minimum # of information = 3 (3) (3) (3) (3)

Optimum # of information = 4 (4) (4) (4) (4)

Number of angular sectors = 4 (4) (4) (4) (4)

Automatic Sorting on

X-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

Y-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

Z-Mesh of sorting grid = 228.12m (228.12m) (228.12m) (228.12m)

No Heterotopic Search

No Minimum Distance used between Data Points

No Maximum Distance without Samples

No Maximum Number of Consecutive Empty Sectors

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : NI

Number of basic structures = 3

S1 : Nugget effect

Sill = 5.0000

S2 : Spherical - Range = 90.00m

Sill = 12.1500

S3 : Spherical - Range = 89.60m

Sill = 28.3500

Directional Scales = (160.00m, 89.60m, 89.60m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : CO

Number of basic structures = 3

S1 : Nugget effect

Sill = 0.0120

S2 : Spherical - Range = 100.00m

Sill = 0.1600

S3 : Spherical - Range = 100.00m

Sill = 0.1300

Directional Scales = (200.00m, 100.00m, 100.00m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : MG

Number of basic structures = 3

S1 : Nugget effect

Sill = 135.0000

S2 : Spherical - Range = 140.00m

Sill = 2300.0000

S3 : Spherical - Range = 150.00m

Sill = 2162.0000

Directional Scales = (300.00m, 150.00m, 150.00m)

Local Rot (mathematician)= (120.00, 0.00, 0.00)

Local Rot (geologist) = (-30.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : FE

Number of basic structures = 3

S1 : Nugget effect

Sill = 1030.0000

S2 : Spherical - Range = 75.00m

Sill = 3100.0000

S3 : Spherical - Range = 99.90m

Sill = 16790.0000

Directional Scales = (185.00m, 99.90m, 99.90m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Model : Covariance part

==========

Number of variables = 1

- Variable 1 : ZN

Number of basic structures = 3

S1 : Nugget effect

Sill = 17700.0000

S2 : Spherical - Range = 85.00m

Sill = 274239.0000

S3 : Spherical - Range = 80.00m

Sill = 601500.0000

Directional Scales = (200.00m, 80.00m, 80.00m)

Local Rot (mathematician)= (0.00, 0.00, 0.00)

Local Rot (geologist) = (90.00, 0.00, 0.00)

Model : Drift part

==========

Number of drift functions = 1

- Universality condition

Date: 12th November 2002

David Selfe
Anaconda Operations
Locked Bag 4
Welshpool Delivery Centre
Pilbara Street Welshpool, WA

Anaconda

Ute Mueller
School of Engineering and Mathematics
Edith Cowan University, WA

Dear Ute,

Re: Anaconda Multivariate Data Set

I refer to the e-mail sent by you on the 5th November 2002. Anaconda gives its permission for publication of research results on the data supplied to you by Mark Murphy on the following basis:

- Raw data is not published,
- Original coordinate data will not be published and any published data will be made anonymous by changing locations.
- Anaconda will be informed and given a copy of any publications,
- Anaconda will receive due acknowledgement of it's support of the work, and it's permission for the publication of the results,
- This permission is only applicable to the studies of Ms Ellen Bandarian, any further use of the data will require separate permission to be granted from Anaconda.

Yours Sincerely

[redacted]

David Selfe

Chief Production Geologist