2015

Compositions, logratios and geostatistics: An application to iron ore

Clint Ward
Edith Cowan University

2015

Compositions, logratios and geostatistics: An application to iron ore

Clint Ward
Edith Cowan University

Edith Cowan University

Copyright Warning

Compositions, logratios and geostatistics: An application to Iron Ore

by

Clint Ward

A Thesis Submitted to the
Faculty of Health, Engineering and Science
Edith Cowan University
Perth, Western Australia

In Partial Fulfilment of the Requirements for the
Masters of Science (Mathematics and Planning)

Supervisors:
Associate Professor U Mueller
Dr J Lo

11^{th} February 2015

USE OF THESIS

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

Copyright and Access Declaration

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

1. Incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;
2. Contain any material previously published or written by another person except where due reference is made in the text; or
3. Contain any defamatory material.
4. Contain any data that has not been collected in a manner consistent with ethics approval.

The Ethics Committee may refer any incidents involving requests for ethics approval after data collection to the relevant Faculty for action.

Dated: 11^{th} February 2015

Abstract

Common implementations of geostatistical methods, kriging and simulation, ignore the fact that geochemical data are usually reported in weight percent, sum to a constant, and are thus compositional in nature. The constant sum implies that rescaling has occurred and this can be shown to produce *spurious correlations*. Compositional geostatistics is an approach developed to ensure that the constant sum constraint is respected in estimation while removing dependencies on the spurious correlations. This study tests the applicability of this method against the commonly implemented ordinary cokriging method. The sample data are production blast cuttings analyses drawn from a producing iron ore mine in Western Australia. Previous studies using the high spatial density blast hole data and compositional geostatistical approach returned encouraging results, results other practitioners suggested were due to the high spatial density. This assertion is tested through sub-sampling of the initial data to create four subsets of successively lower spatial densities representing densities, spacings, and orientations typical of the different stages of mine development. The same compositional geostatistical approach was then applied to the subsets using jack-knifing to produce estimates at the removed data locations. Although other compositional geostatistical solutions are available, the additive logratio (alr) approach used in this study is the simplest to implement using commercially available software. The advantages of the logratio methodology are the removal of the constant sum constraint, allowing the resulting quantities to range freely within the real space and, importantly, the use of many proven statistical and geostatistical methods. The back transformation of linear combinations of these quantities and associated estimation variances to the constrained sample space is known to be biased; this study used numerical integration by Gauss-Hermite quadrature to overcome this drawback. The Aitchison and Euclidean distances were used to quantify both the univariate and compositional errors between the estimates and original sample values from each estimation method. The errors of each method are analysed using common descriptive and graphical criteria including the standardised residual sum of squares and an assessment of the accuracy and precision. The highest spatial density dataset is equally well

reproduced by either method. The compositional method is generally more accurate and precise than the conventional method. In general the compositional error analyses favour the compositional techniques, producing more geologically plausible results, and which sum to the required value. The results support the application of the logratio compositional methodology to low spatial density data over the commonly implemented ordinary cokriging.

Acknowledgments

Teaching mathematics to geologists is challenging. Geologists are by nature ethereal, verbose and ambiguous; traits completely at odds with mathematicians who pride themselves as succinct practitioners of the purest of all the sciences. As a result, this thesis is the end product of much work, not only by the author, but also by the unfortunate group of people I call my family, supervisors, mentors, colleagues and friends. I would like to thank each of you for the support during what became a difficult journey during trying personal circumstances; thank you for not wavering, for not flagging, for not criticising and for not complaining on the nights and weekends when we could have been doing other things. This has been a goal of mine for many years, a goal interspersed by babies, emigrating and illness, but a goal achieved nonetheless. Thank you Catherine, for supporting what I know you consider madness, thank you Camryn and Sean for not complaining when Dad had to work on nights and weekends, thank you Marlene for instilling in me the desire to learn and better myself, and last and certainly not least thank you Ute for your tireless contributions. You have motivated me and kept me on point throughout the journey, epitomising what supervisors should aspire to. Johnny, welcome to the mad club of geostatisticians, I fear you will fit in well; without any degrees of freedom. And finally, I must thank Cliffs Natural Resources, who provided the dataset and supported me during the whole of the journey including a memorable trip to Norway to present the initial findings.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Compositional data are multivariate data where each variable represents some part of a whole and which carries only relative information. Compositional data occur in many fields including natural resources, biology, forestry, and economics. In the context of mining or natural resources, compositional data can take a number of different forms. One form is geochemical data; such as whole rock chemical analyses, or the sum of the major oxide components. Quantitative mineralogical data is another, where the proportions of each mineral species present in a sample are totaled. Resource evaluation, mine planning and geometallurgy all use these quantities and perform mathematical operations on them, ultimately making business decisions based upon the values and the relationships between them. Geochemical data used in natural resource evaluation, in most measurement systems (weight percent, parts per million etc), are compositional data in that a rescaling or normalisation of the original units has occurred, resulting in a constant sum constraint. The so called *natural* iron ores, such as those mined in the Yilgarn and Pilbara regions in Western Australia, are sold to customers in much the same form they occur naturally; no beneficiation occurs between mining and delivery to customer with the exception of crushing and screening the ore to produce lump and fines products. Thus the relationships noted in the whole rock geochemical analyses of in-situ material must be present in the final products delivered to customers. Geostatistical methods are routinely used to estimate or simulate geochemical data for use in downstream economic and mine engineering applications. However, the compositional nature of the input data is rarely considered, and appropriate techniques to correctly represent the correlations and totals sums are seldom, if ever, utilised [1]. The correlations noted in the whole rock analyses are required both by the mine planning staff to optimise the delivery of ore and by customers operating blast furnaces producing iron and steel products. Errors in the correlations, both locally and globally are almost guaranteed if the compositional nature of the data are not taken into consideration. The decision as to whether any part of an iron ore deposit is economic to process is achieved through considering all of the components, therefore misrepresenting the correlations between components will result in incorrect classification of ore mate-

rial as waste and vice versa. The following sections will examine the available published works which use compositional data with natural resource applications.

1.1 Outline of the problem

During a study of animal proportions, Pearson [2] observed and demonstrated the *spurious correlation* that arises between uncorrelated measures that are assessed using a common denominator and attributed this to the rescaling. He proposed adjustments to account for the spurious correlation, however, these adjustments did not take into account the range of the data and essentially ignored the compositional nature of the problem; focusing instead on the scaling. It was more than half a century later that Chayes [3] described the spurious correlation problem in the context of compositional data. His contribution demonstrated that due to the constant sum constraint, rather than the rescaling, some of the correlations must be negative. The most significant non-spatial research was conducted by Aitchison starting in 1982 with his address to the Royal Statistical Society [4] and is summarised in his 1986 monograph [5]. In these early works he set up the basic compositional data framework adhered to today, including the terminology. The major thrust of his research is the recognition of the relative nature of compositions, being ratios, Aitchison was the first to propose working with the logarithm of the ratio rather than the unwieldy ratios directly. The main feature of the logarithmic transform is to move the problem from a constrained sample space imposed by the compositional ratio into an unconstrained one, enabling the use of a much larger variety of existing statistical methods [5].

Theoretical developments

Pawlowsky-Glahn [6] furthered the study of compositional data by recognising that directly applying geostatistical techniques to compositional data would carry the same spurious correlation penalty as non-spatial techniques. In later works with others she proposed combining Aitchison's logratio approach with traditional geostatistical techniques [7], [8]. The most complete treatment of the proposed solutions with derivations of key findings and a proposed work flow on an oil sands dataset focused particular attention to the treatment and modelling of the spatial covariance structure of compositional data, a basic requirement for most geostatistical techniques [7]. The transformation of the data using the selected logratios presents theoretical and practical difficulties in the modelling of the spatial covariances. It is shown that the different transformations are related to one another, creating pathways for modelling the spatial covariances via conventional techniques and software. The approach implemented by Pawlowsky-Glahn and Olea [7] uses Fast-Fourier Transforms and bespoke software code. When the technique proposed is

used with an example dataset, superior results were generated when compared against conventional techniques using compositional metrics which included the Aitchison and Mahanalobis distances. Critically, the constant sum total is honoured when using the proposed work flow while errors in excess of one percent of the required total were noted when using conventional techniques which ignore this feature of the input data. A one percent total discrepancy is not large, however, if it is largely confined to a single component upon which a categorisation or classification is dependent such as in a mining study, material misclassification could result.

This is not the only possible solution to the problem of estimating spatially distributed compositional data where the resulting estimates honour the compositional nature of the input data; nevertheless, many practitioners merely ignore the issue and treat compositional data as they would any other continuous additive variable. Indeed, in most cases the sum of components fall within ranges deemed acceptable, even if the constant sum constraint is rarely honoured, and therefore the correlations cannot be accurately reproduced. In essence the estimates are not optimal, normally an important aim of an estimation exercise. Optimal estimates of regionalised compositional variables should comply with the non-negativity and constant sum constraints; examples of methods that can achieve this are nearest neighbour, inverse distance, triangulation and local sample mean interpolation. These methods will always yield compositional estimates because the columns of the matrix of weights are identical, the weights sum to one, the weights range between zero and one and the conditioning data are compositions. However, the spatial covariance structure is not taken into account [9]. Variogram-based geostatistical methods including ordinary cokriging, are designed specifically to take the spatial character of the input data into consideration. However, ordinary cokriging of nonnegative input data can generate negative estimates, frequently noted when the *screening effect* occurs [10]. Additionally, estimates honouring the constant sum constraint are only produced in the special case of proportional semi-variogram models which is seldom possible in practice. One way to ensure that kriging obeys the compositional restrictions is to add one or more conditions to those the estimator must satisfy. Walvoort and de Gruijter [9] proposed a new interpolation method, compositional kriging, based on ordinary kriging that satisfies the constraints imposed by compositional data outlined above. The compositional kriging system is a modification of ordinary kriging and considers each component of the compositional whole sequentially, rather than ordinary cokriging which considers all components simultaneously. This is performed to avoid the following features of cokriging compositions namely:

1. estimating and modelling the cross semi-variograms,
2. forcing the system to honour correlations known to be spurious,
3. enforcing linear relationships between components that may be non-linear, and
4. the positive-definite restriction on the linear model of coregionalisation can result

in a poor fit of models to experimental data.

In order to satisfy the requirements of the desired compositional estimator the compositional kriging system is a direct extension of ordinary kriging with the addition of the nonnegativity and constant sum constraints. Compositional kriging has a number of drawbacks:

1. the cross-correlations are not taken into account omitting potentially valuable information,
2. due to difficulties in its interpretation, the prediction error variance cannot be utilised as a measure for constructing confidence intervals,
3. commercial software is not readily available for testing and implementation, and
4. the solutions to the compositional kriging systems are iterative and can be computationally expensive.

The compositional kriging method was compared with the logratio approach advocated by Olea and Pawlowsky-Glahn [7] using the direct back-transformation of the logratio estimates to the constrained sample space in two case studies. The results of the comparison were not decisive with regards to selecting the optimal method in that the logratio method produced results superior to the proposed compositional kriging method in one of the case studies. Koushavand and Deutsch [11] examine the problem of spatial estimation of compositional data and suggest that solving a kriging system with an additional constraint could solve the compositional data estimation problem, from the univariate to the multivariate case. This was performed by adding an additional constraint to the simple cokriging system and solving only one system for each data location to force the sum of the resulting estimates to a specified value. The drawbacks with the system are that negative estimates can still result even though the total sums to the required constraint. A second step to remove possible negative estimates by setting those estimates to zero, removing them and resolving the kriging system adds computational expense. Additionally this creates a relationship between the kriging weights and the informing data and therefore the potential for bias exists [11]. Moreover, this constrained cokriging approach does not address the spurious correlation issue as the correlations evident in the data would be transfered through the linear model of coregionalisation. The modifications were only tested on the simple kriging system which is globally unbiased under appropriate conditions of stationarity but local bias is expected in most real data applications.

Case studies

Jaruta-Bragulat *et al* [12] compare four different methods of geostatistical interpolation techniques applied to a compositional bauxite dataset. Four techniques are employed;

ordinary cokriging of the variables in original units, two logratio transformations with no mention of the approach used to model the spatial covariance structure, and the direct application of the Pawlowsky-Glahn and Olea [7] workflow by using Fast-Fourier Transforms to model the spatial covariances. The graphical results shown in the paper suggest that the four techniques were equivalent with some subjective advantage allocated to the Fast-Fourier Transform method. No mention of the back transformation method or the effect on bias was made. Manchuk [1] provides a holistic mathematical and practical overview of compositions including considering geostatistical applications. Although not the primary focus of the work, the problem of zero values components with logratio transforms is addressed, and several solutions are proposed. Geochemical data are unlikely truly barren in any one component, due partially to the analytical accuracy of the equipment, rendering the problem one of interest rather than practicality. The theoretical aspects of implementing geostatistics using the logratio techniques is discussed but no case studies are presented.

The isometric logratio transformation is used in conjunction with ordinary kriging [14] to model and estimate granite proportions and compare against fuzzy kriging and ordinary kriging. The results compare positively for the compositional approach by increasing confidence in the estimates over the other techniques. The compositional geostatistical approach is applied to particle size distributions in [15] combining functional analytical techniques with compositional techniques and ordinary kriging. The procedure which utilises the centred logratio transformation, cross-validates well and generates excellent results at unsampled locations. Further recognition of the compositional nature of grain-size data is described in [13] where indicator kriging, ordinary kriging and additive logratio cokriging are compared. The techniques are used to produce two dimensional maps of the sand, silt and clay and the results compared through cross-validation. The additive logratio cokriging technique produced the best results using the mean absolute error and confusion matrix criteria. The merits of compositional kriging, ordinary cokriging, logratio cokriging and additive logratio cokriging of soil particle size fractions are compared in two case studies [16]. The study focuses on the performance of each technique measured with traditional and compositional measures including the Mahalanobis and Aitchison distances. In these applications the outcomes did not favour any particular compositional technique; the compositional techniques did however, produce better results when compared with techniques that do not take the constant sum into account.

Boezio *et al* [17] applied ordinary cokriging to additive logratio transformed exploration iron ore data and compared the results of this method with ordinary cokriging using univariate criteria such as global and local means. When compared with the input data the logratio estimates display bias for some variables, bias not noted in the ordinary cokriging results. Boezio *et al* [18] then applied the Minimum / Maximum Autocorrelation Factors (MAF) technique to iron ore data which had been initially transformed using the

alr. The MAF technique was applied to decorrelate the data, thereby removing the requirement to use the restrictive linear model of coregionalisation. The published results in the paper displayed a bias of the iron grades when compared against the input data. The precursor studies [19], [20] to this thesis on production blast hole data from an operating iron mine used the alr cokriging procedure outlined by Pawlowsky-Glahn and Olea [7]. These precursor studies, using only the three principal analytes of interest in iron ore mining, compared the alr cokriging method against the sample data using cross-validation and against the traditional ordinary cokriging method. The results compared favourably with the input data and with the ordinary cokriging. In all the geostatistical case studies mentioned above that utilise the logratio method to transform the data to an unconstrained space followed by interpolation, another function, such as the additive generalised logistic (agl) is applied to return the estimates to the constrained sample space.

Elimination of bias

Job [21] concludes that although the results from the geostatistical case studies available in the literature appear unbiased, in fact the direct application of inverse logratio functions such as the the agl, necessarily generate estimates that are biased. A small example is worked in his thesis to illustrate the error induced by applying a linear process to non-linearly transformed variables and then applying the inverse non-linear transformation resulting in bias. This problem is reinforced by the comments of Chilès and Delfiner [22] who dismiss logratio methods as biased. Aitchison [5] acknowledges this phenomenon and states that although the moments (such as the expected value and variance) of logratio transformed distributions exist, they do not reduce to a simple form. He then outlines the use of numerical integration to generate approximations of the required moments. Pawlowsky-Glahn and Olea [7] describe this procedure by using Gauss-Hermite quadrature as a method to compute the expected value and estimation variance at an estimation location in the original sample space. Tolosana-Delgado *et al* [23] illustrate that another mechanism can be used. Under the assumption that alr transformed data at a certain support come from a Gaussian random field, Monte Carlo simulation can provide an upscaling. This concept was tested on exploration iron ore data where the target blocks were partitioned into units of that support, and LU simulation on the discretised block applied, conditional to the available data. Simulated values were back-transformed with the agl, and averaged to obtain a simulated block mean composition. The results of the block simulation technique compared favourably against results from Gauss-Hermite back transformation of cross-validated point estimates and the input data.

The effect of data density

Ward and Mueller [19] compared estimates generated using the agl back transformation with the input data which were reasonably good but did exhibit minor bias. However, in addition to the agl back transformation, the numerical integration approach was also used to generate estimates in the original units. These estimates compared more favourably with the input data and the ordinary cokriging results than the agl back transformed results. Other practitioners asserted that the reason for the low bias noted in the agl back transformed estimates (and by proxy the estimates derived by numerically integration) is the high spatial density of the input data (*pers. comm.* Job, 2013). In order to validate this, a follow up study used the same base dataset and three analytes from the first study, but created successively lower data densities and used the jack-knifing technique to test this [20]. The study focused only on the now proven numerical integration approach, as early (unpublished) results using the lower density input data and the agl back-transformation were strongly biased. The numerical integration results were comparable with the ordinary cokriging results in terms of bias, but with the added advantage of summing to the required constant at all the spatial densities tested. The review of current literature highlights that logratio techniques are increasingly seen and tested as an alternative to the traditional methods for the spatial interpolation of compositional data. It is noted however, that four areas specific to mining would benefit from further examination.

1. Bias in the back-transformed results of interpolated logratio estimates is of primary importance. It is difficult to convince practical mining professionals to consider alternative methods if the accuracy is inferior, even if the constant sum is respected.

2. Correlations between the estimated components is important; indeed, validation reports of estimation studies which use traditional estimation or simulation techniques focus heavily on the reproduction of the correlations in the estimates, especially when a decorrelating technique has been used to remove the onerous requirement for fitting of the Linear Model of Coregionalisation. In the literature however, very little, if any time is spent verifying this critical feature.

3. Spatial covariance modelling; the application of most geostatistical techniques requires a model of the spatial covariance, an aspect that Pawlowsky-Glahn dedicates much effort to provide a theoretically sound framework for logratio transformed variables [7]. The majority of practical applications of variogram-based geostatistical methods ignore much of this theory and merely apply standard software and modelling practices to logratio transformed data.

4. The impact of data density on the resulting estimates. The effects are well understood with linear estimators, furthering the understanding of logratio techniques is necessary to drive acceptance of these methods with other practitioners.

1.2 Objectives

Three overarching objectives are identified based upon the literature review and the perceived gaps in the current set of knowledge and specifically case studies.

Objective 1

An examination of the logratio framework, both in the classical as well as the spatial sense. As part of this objective, the theory of regionalised compositions will be discussed, along with the methodology for ordinary cokriging and compositional cokriging. The approach will be illustrated through the application to a compositional iron ore geochemical data set, comprised of 7 analytes (Al_2O_3, Fe, Loss on Ignition (LOI), Mn, P, S, SiO_2) and a filler variable. Of particular interest will be an assessment of the stationarity of the input data. Two logratio transformations will be applied: Isometric Logratio (ilr) (for stationarity assessment) and alr (for cokriging of the logratios).

Objective 2

A comparison of compositional cokriging via logratios with ordinary cokriging in the context of the compositional iron ore geochemical data set. The comparison will be based on the cross validation results and the assessment will include:

- accuracy and precision of the estimates,
- univariate and multivariate errors calculated using compositional and standard Euclidean approaches,
- standardised Residual Sum if Squares (STRESS), and
- scatterplots to check the reproduction of bivariate relationships between analytes.

Objective 3

An analysis of the impact of spatial data density on the estimation results. The original data will be subsampled to generate 4 subsets with properties matching the typical drilling available at varying stages of mine development. These subsets comprise successively sparser samples, each containing 27%, 15%, 9%, and finally 4.4% of the total dataset respectively. The methodology outlined in objective 2 will be used to examine the impact of the subsampling.

1.3 Structure of the thesis

Chapter 1 addresses the nature of the spurious correlation problem, the proposed solutions and the testing approach of these solutions. Additionally the available literature relating to the problem and solutions is presented. The mathematical background and theory is presented in Chapter 2, starting with the basic description of random variables as applied to spatially correlated examples. Thereafter the link between the spurious correlation problem and random variables is examined, with an emphasis on the logratio transformations available. The geostatistical requirement for stationarity is followed by the description of the cokriging systems used. The chapter concludes with a description of the Gauss-Hermite back-transformation measures employed and the testing measures used to evaluate the applied techniques. Chapter 3 is focused on the data used in this study; beginning with the geological description and upgrade theories. The sampling methodology used to collect the data is described. Chapter 4 contains details of the implementation and steps followed in the process with a flow-sheet. Chapter 5 presents the exploratory data analysis and modelling of the spatial covariance in the simplex and the logratio space together with results of the estimation and performance comparisons presented in Chapter 6. The errors are comprehensively examined using traditional and compositional error measures for bias and information loss. In addition, the precision and accuracy of the estimation techniques are compared as is the totals of the analytes. Chapter 7 draws conclusions from the work and future questions are posed.

1.4 Notation and terminology

Mathematical notation

Symbol	Description
$\mathbf{x}$	row vector
S^D	*D-Simplex*
u	datum location
u_i	datum location for estimation quality tests
$\mathbf{A}$	Study area
$\mathbf{X}(u)$	vector of continuous random variable at location u
$X_i(u)$	i^{th} component of continuous random variable at location u
$\mathbf{Z}(u)$	vector of logratio transformed continuous random variable at location u
$Z_i(u)$	i^{th} component of logratio transformed continuous random variable at location u
g_i	geometric mean of a D-component continuous random variable
$\mathbf{g}$	closed geometric mean of g_i
$\mathbf{T}$	variation matrix

D	number of parts of a composition
$\mathbb{R}^D$	D-dimensional real space
$\mathbf{w}_i$	component in orthonormal basis generating system
b	balance between two subgroups of an orthonormal basis
a_+	numerator weight of the i^{th} balance
a_-	denominator weight of the i^{th} balance
x^*	coordinate of $\mathbf{x}$
$E[\mathbf{Z}(u)]$	expected value of $\mathbf{Z}(u)$
$Var[\mathbf{Z}(u)]$	variance of $\mathbf{Z}(u)$
h	separation or lag distance
$Cov(h)$	covariance of pairs separated by h
$\gamma(h)$	experimental auto semivariogram for lag h
$\gamma_{i.j}(h)$	experimental cross-semivariogram between component i and j for lag h
$\mathbf{\Psi}(h)$	semivariogram function for lag h
$\mathbf{\Sigma}(h)$	covariance function for lag h
$\mathbf{Z}^*(u)$	estimate $\mathbf{Z}(u)$ of at location u
Γ_k	matrix of weights
ν	Lagrange multiplier
$\overline{\mu}_{\mathbf{z}}(u))$	approximation of $E[\mathbf{Z}^*(u)]$
$\Sigma_Z(u)$	approximation of estimation variance of $\mathbf{Z}^*(u)$
$\Sigma_X(u)$	estimation variance of $\mathbf{X}^*(u)$
δ_u	univariate error at datum location u
$\varepsilon(u_i;p)$	indicator function at each datum location u_i for each analyte
A	Accuracy statistic
P	Precision measure

Abbreviations

Abbreviation	**Expansion**
alr	Additive logratio
clr	Centred logratio
mlr	Multiplicative logratio
ilr	Isometric logratio
agl	Generalized additive logistic transformation
LMC	Linear Model of Coregionalisation
LOI	Loss on Ignition
BH	Blast hole
RC	Reverse Circulation
OCK	Ordinary Cokriging
GH	Gauss-Hermite
STRESS	Standardised Residual Sum of Squares

ROM	Run of Mine
QKNA	Quantitative Kriging Neighborhood Analysis
EDA	Exploratory Data Analysis

Chapter 2

Mathematical background and Theory

2.1 Regionalised variables, Random Variables and Compositions

The study and application of geostatistical methods is dependent upon the concept of *random functions*; drawn partially from the field's mining origin which was dedicated to describe geochemical data. Geochemical data are the final output of many natural processes, each of which is imperfectly known, including the sampling of the material. This uncertainty in the processes means that deterministic methods are not suitable to model them. The complexity in the phenomenon under study appears to be the result of random processes; this is not the case, the complexity is a function of the limited information available to describe the phenomenon. However, the complexity does fit well with the concept of random variables; defined as variables that have been generated according to a probabilistic or stochastic process. This approach has advantages; specifically the mathematical framework with which we can work with the random variables and functions of random variables [10]. This set of concepts was extended to compositional data distributed spatially [7]. A D-part composition is defined as a row vector $\mathbf{x}$ where

$$\mathbf{x} = [x_1, x_2, \ldots, x_D] \quad (2.1)$$

and all the D-parts or components are strictly positive real numbers and only carry relative information [8]. The units of the compositional implicitly state that the information carried is relative in that they are always parts of a whole and add up to a constant c. The sample space of compositional data is the *D-simplex*, defined as

$$S^D = \mathbf{x} = [x_1, x_2, \ldots, x_D] \quad (2.2)$$

where $x_i > 0$ and $\sum_{i=1}^{D} x_i = c$. *Regionalised variables* are sets of measurements distributed through space that display dependencies as functions of their proximity to one another [24]. Analogously, a *regionalised composition* is a vector random function $\mathbf{X} : \mathbf{A} \longrightarrow \{\mathbf{X}(\mathbf{u}) : \mathbf{u} \in \mathbf{A}\}$ such that for each $u \in \mathbf{A}$ the vector $\mathbf{X}(u)$ is a composition and the sample space for $\mathbf{X}(u)$ is the D-simplex $\mathbf{Z} \in S^D$.

2.2 Sample spaces

The common implementations used to estimate regionalised compositions ignore the differences between the sample space of the inputs and the possible outputs. This is important both from a theoretical as well as a practical perspective, even if as mentioned in Chapter 1 in some cases the penalties are not so severe as to make the technique practically useless. The *sample space* is, in the usual statistical context, defined as the set of possible results. The constant sum constraint implicit in compositional data implies a constrained sample space. This is critical to the application of statistical techniques to compositional data as the vast majority of multivariate statistical techniques were developed to address unconstrained sample spaces such as the Euclidean space. In fact most statistical methods assume compositional data are drawn from Euclidean space, an assumption implicitly declared when the squared *Euclidean* distance is used to measure errors between predictions and targets [24]. The sample space is, in the usual case of compositional data $\mathbf{X}(u)$ the D-simplex S^D. When D components are to be accounted for, then the corresponding composition is a vector $\mathbf{x} \in S^D$ of positive components which add up to a constant value.

2.3 Compositional operations

Tolosana-Delgado *et al* [25] provide an overview of compositional operations; the critical examples are reproduced hereafter considering a D-part composition of positive components $X_{initial} = [x_1, x_2, \ldots, x_D]$ totaling less than c. A summary of these is required for the development of later properties of compositions (Section 2.5.3), which require the simplex to have a vector space structure. The required operations are perturbation, analogous to addition, and powering analogous to multiplication; with these a vector space structure can be described and therefore a geometry [25]. These components are *closed* or normalised to c through the closure operation $Cl[.]$

$$Cl[x_1, x_2, \ldots, x_D] = \frac{c}{\sum_{i=1}^{D} x_i} \times [x_1, x_2, \ldots, x_D]. \tag{2.3}$$

Compositions are not always static and may undergo changes, geological examples of these changes or *perturbations* are described in Section 3.1.4. In mathematical terms the perturbation operation denoted by $\oplus$ of the D-part composition $X_{initial} = [x_1, x_2, \ldots, x_D]$ being perturbed by proportions $P = [p_1, p_2, \ldots, p_D]$ respectively is the scaling of each component by the corresponding proportion followed by the closure operation

$$\begin{aligned} X_{perturbed} &= X_{initial} \oplus P \\ X_{perturbed} &= [x_1, x_2, \ldots, x_D] \oplus [p_1, p_2, \ldots, p_D] \\ X_{perturbed} &= Cl[x_1 \times p_1, x_2 \times p_2, \ldots, x_D \times p_D]. \end{aligned} \tag{2.4}$$

Powering is multiple (α) iterations of the perturbation operation and is denoted by $\otimes$.

$$\begin{aligned} [X_{initial}]\alpha P &= (\alpha \otimes P) \oplus X_{initial} \\ [X_{initial}]\alpha P &= Cl[x_1 \times {p_1}^{\alpha}, x_2 \times {p_2}^{\alpha}, \ldots, x_D \times {p_D}^{\alpha}] \end{aligned} \tag{2.5}$$

The *compositional differences* operation is analogous to determining the proportion removed (δ) by a perturbation from the example system and is written

$$\begin{aligned} \delta &= X_{perturbed} - X_{initial} \\ \delta &= [x_1, x_2, \ldots, x_D] \ominus [p_1, p_2, \ldots, p_D] = X_{perturbed} \oplus (\alpha \otimes X_{initial}) \end{aligned} \tag{2.6}$$

where α = -1. In order to define a linear vector space structure the inner product is required. The inner product of $X_{initial} = [x_1, x_2, \ldots, x_D]$ and $Y_{initial} = [y_1, y_2, \ldots, y_D]$ is:

$$\langle X_{initial}, Y_{initial} \rangle = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \ln \frac{x_i}{x_j} \ln \frac{y_i}{y_j}. \tag{2.7}$$

This inner product associates a norm of $X_{initial}$

$$\| X_{initial} \| = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(\ln \frac{x_i}{x_j} \right)^2} \tag{2.8}$$

and distance between $X_{initial}$ and $Y_{initial}$

$$d(X_{initial}, Y_{initial}) = \sqrt{\frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2}. \tag{2.9}$$

These concepts are utilised in the discussion of the isometric logratio transformation (Section 2.5.3) and the distance (Section 2.12.2).

2.4 Compositional statistical measures

2.4.1 Numerical Compositional techniques

As traditional statistical measures of central tendency and dispersion do not adequately describe compositional data, alternatives are described below. Pawlowsky-Glahn *et al* [8] noted that the *centre* of a compositional distribution is a better descriptor than the arithmetic mean. For a D-part compositional data set of size n the centre is the closed geometric mean $\mathbf{g}$ defined as

$$\mathbf{g} = Cl[g_1, g_1, \ldots, g_D], \tag{2.10}$$

with $g_i = \left(\prod_{j=1}^{n} x_{ij}\right)^{\frac{1}{n}}$, $i = 1, 2, \ldots, D$. Dispersion in a compositional dataset is described by the *variation matrix* $\mathbf{T}$ defined as

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \ldots & t_{1D} \\ t_{21} & t_{22} & \ldots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & \ldots & t_{DD} \end{bmatrix}, \; t_{ij} = var\left(\ln \frac{x_i}{x_j}\right). \tag{2.11}$$

The global dispersion of the dataset can be measured through the *total variance* which is given by

$$\text{totvar}[\mathbf{X}] = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} var\left(\ln \frac{x_i}{x_j}\right). \tag{2.12}$$

The diagonal of the variation matrix $\mathbf{T}$ will contain only zeros and values indicate how the variability is split between the components (or their logratios). It can be demonstrated that the total variance and the variation matrix values do not change with changing c as the ratios are relative [8].

2.4.2 Visual Compositional techniques

Further to the numerical description Aitchison [5] adapted existing plots to assist with the analysis of compositional data. The first plot utilised is the *ternary diagram*, familiar to geologists, made more useful through *centring* of the points to the barycentre of the simplex. The centring results in easily visualised diagrams ready for interpretation. The centring is accomplished by determining the inverse of the geometric mean of the component $\mathbf{g}^{-1}$ and perturbing the sample by this quantity. An example of the changes induced by the centring operation is shown in Figure 2.1. The second plot is the *biplot*, a type of exploratory graph used in statistics, which is a generalization of the simple two-variable

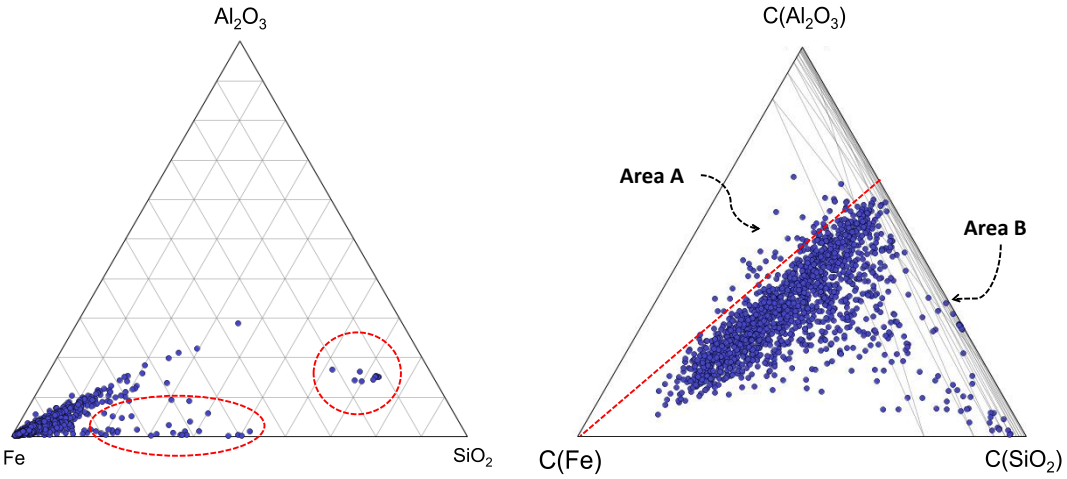


Figure 2.1: Ternary diagram (left) and centred ternary diagram(right)

scatterplot. These were also adapted for use with compositional data and can be used for exploratory data analysis to explain sources of variability. A detailed discussion of the construction of the compositional biplot is beyond the scope of this document; suffice it to say the biplot graphically displays the best rank-2 approximation $\mathbf{Y}(u)$ to the coefficients $\mathbf{Z}(u)$ in clr coordinates in the least squares sense given by the singular value decomposition. Interested readers are directed to the discussion in Pawlowsky-Glahn *et al* [8]. A schematic example of a biplot is given in Figure 2.2, the parts of which are:

1. the origin O which is the centre of the dataset,
2. the *rays* joining the origin O to each vertex of the D parts,
3. the *links* joining the end point of each ray.

The biplot is interpreted with the following guidelines:

1. if links of two components are a right angles then zero correlation of the (sub)compositions is expected and possible independence, or
2. if links are coincident then $\text{var}(\ln(\frac{x_i}{x_j}))$ is zero and the parts x_i and x_j can be considered redundant, and finally
3. co-linear links indicate one dimensional variability; i.e. plotting along a compositional line.

The schematic biplot in 2.2 shows that the link joining clr(a) and clr(b) is perpendicular to the link joining clr(b) and clr(c), this indicates low or zero correlation with possible independence of these components. By contrast the link joining clr(a) and clr(b) is at an acute angle to the link joining clr(a) and clr(c), this indicates some redundance between the information carried by those components with respect to one another.

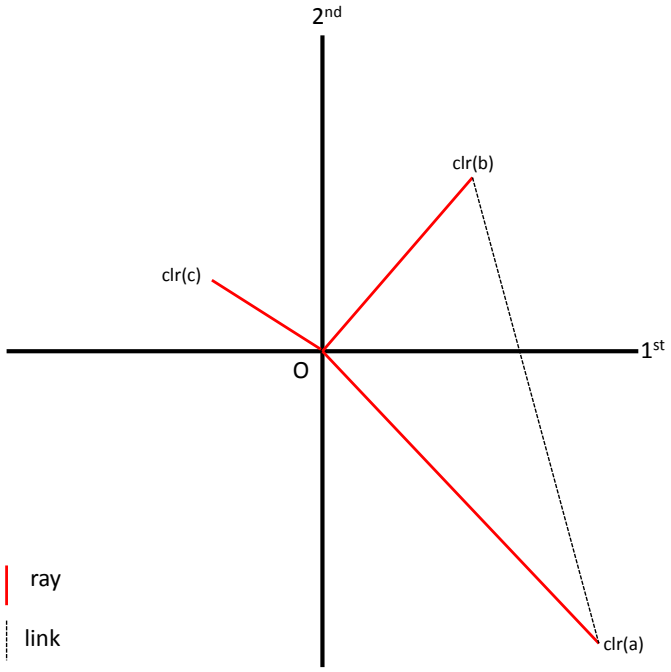


Figure 2.2: 3 part schematic Biplot

2.5 Logratios

The relative nature of compositional data lends itself to analysis with ratios. In his seminal publication, Aitchison [5] introduced transformations based on logratios to use these relative properties. Early work focused on the *additive logratio* transformation (alr) and the *centred logratio* transformation (clr). The other transformation of importance is the *isometric logratio* transformation (ilr); developed to overcome the shortcomings of both the alr and the clr which are expanded upon in their respective sections below. The ilr is an association of coordinates with compositions in an orthonormal system and results in an isometry between S^D and $\mathbb{R}^{D-1}$ [26].

2.5.1 Additive logratio transformation

The *additive logratio* transformation alr : $S^D \longrightarrow \mathbb{R}^{D-1}$ (relative to the i^{th} component), given by

$$\text{alr}(\mathbf{Z}(u)) = \mathbf{X}(u) = \left(\ln\left(\frac{Z_1(u)}{Z_i(u)}\right), \ln\left(\frac{Z_2(u)}{Z_i(u)}\right), \ldots, \ln\left(\frac{Z_D(u)}{Z_i(u)}\right) \right) \tag{2.13}$$

embeds the D-simplex into $(D-1)$-dimensional space and removes the constant sum constraint. In practice any one of the parts can be used as the divisor [7]; and should be selected so as to minimise any computational difficulties in later processes. The *multiplicative logratio* (mlr) is a further development of the alr in which a filler variable is calculated to ensure the constant sum constraint is honoured and is then used as the

denominator. Due to this the mlr and the alr have the same properties when the filler variable is considered equivalent to the original parts of the composition. The alr transformation results in $D-1$ components; advantageous when considering the fitting of the linear model of coregionalisation (LMC) to the resulting components. This property is also a drawback as by changing the divisor the alr transformation changes too (it is defined as asymmetric). The major problem when applying the alr transform is that it is not an isometric transformation from the simplex, with the Aitchison metric, onto the real space with the Euclidean metric [26]. The alr transform is invertible and its inverse is the *additive generalised logistic* (agl) transform:

$$\text{agl}(\mathbf{X}(u)) = \mathbf{Z}(u) = \frac{(\exp(x_1(u)), \ldots, \exp(x_{D-1}(u)), 1)}{1 + \sum_{j=i}^{D-1} \exp(x_i(u))} \tag{2.14}$$

2.5.2 Centred logratio transformation

The *centred logratio* transformation $\text{clr} : S^D \longrightarrow \mathbb{R}^D$ given by

$$\text{clr}(\mathbf{Z}(u)) = \mathbf{X}(u) = \left(\ln \left(\frac{z_1(u)}{g(Z(u))} \right), \ldots, \ln \left(\frac{z_{D-1}(u)}{g(Z(u))} \right) \right) \tag{2.15}$$

where

$$g(Z(u)) = \left(\prod_{i=1}^{D} z_i(u) \right)^{\frac{1}{D}} \tag{2.16}$$

is the geometric mean. The advantage of the clr is symmetry in the components but this is offset by a new constraint on the transformed variable; the sum of the components is equal to zero [26]. Thus the covariance matrix of $\mathbf{X}(u)$ is singular; this requires the use of generalised inverses for cokriging, an option not available on commercial software [7]. The clr transform is invertible and its inverse is the following:

$$\text{clr}^{-1}(\mathbf{X}(u)) = \mathbf{Z}(u) = Cl[\exp(x_1(u)), \exp(x_2(u)), \ldots, \exp(x_D(u))] \tag{2.17}$$

2.5.3 Coordinates and the Isometric logratio transformation

Mathematical statistics relies on real analysis which is commonly performed using coordinates with respect to an orthonormal basis [27]. In statistics, the real space $\mathbb{R}$ is assumed to be the sample space for a set of observations and the Euclidean geometry as the algebraic-geometric structure. If the sample space of a given set of observations is different then an alternative strategy must be used or one runs the risk of incompatibilities or incoherences arising; this is epitomised by the spurious correlation referred to in Chapter 1. One such strategy uses coordinates and although very similar to the logratio approach some differences are noted. Using coordinates is called *Stay in the simplex* or

stay and the logratio approach is called *move*. An alternative geometry is required for the stay approach available through the properties of the Aitchison inner product, norm and distance which honour the important principles of compositional analysis; scale invariance, sub-compositional coherence and permutation invariance. These properties are not defined here, however further information is available in [5] and summarised in [8]. Through the so called Aitchison simplicial geometry; the properties of Euclidean spaces can be used to analyse compositional data, which importantly, allows the use of standard statistical techniques and software [27]. A critical element in this geometry is the construction of orthonormal bases and their corresponding coordinates. The construction of orthonormal coordinates has been called the *isometric logratio* transformation (ilr). The first step in defining an orthonormal basis is finding a generating system to build the basis. Pawlowsky-Glahn *et al* [8] demonstrate that one method is

$$\mathbf{w}_i = C(\exp(\mathbf{e}_i)) = C[1,1,\ldots,e,\ldots,1], \quad i = 1,2,\ldots,D, \tag{2.18}$$

where in each $\mathbf{w}_i$ the number e is in the i-th column and the exponentiation operates component-wise. Using the fact that the closure operation is unaffected by the choice of c we can represent any vector $\mathbf{x} \in S^D$ as

$$\mathbf{x} = \bigoplus_{i=1}^{D} \ln x_i \otimes \mathbf{w}_i \quad = \ln x_1 \otimes [e,1,\ldots,1] \oplus \ln x_2 \otimes [1,e,\ldots,1] \oplus \cdots \oplus \ln x_D \otimes [1,1,\ldots,e]. \tag{2.19}$$

It can be further shown that the clr coefficients are generated by dividing each component of Equation 2.19 by the geometric mean (Equation 2.16). The complete derivation is beyond the scope of this study, however interested readers are directed to Pawlowsky-Glahn *et al* [8]. If one of the vectors from the generating system in Equation 2.19 is omitted then a basis is obtained, however this is not an orthonormal basis. Egozcue *et al* [26] demonstrate that the Gram-Schmidt procedure can be used to generate one of infinitely many orthonormal bases in any Euclidean space. The relationships between groups of components can be useful in statistical analysis of compositions. If a geological reason is available to subdivide the composition into groups then a relevant and useful orthonormal basis can be generated using *sequential binary partitioning* as discussed in detail in Egozcue and Pawlowsky-Glahn [28]. A sequential binary partition is a hierarchy of the parts of a composition which, importantly, can be performed in a manner meaningful to the geology of the data. The total composition is split into to two groups of parts. Each group is then sequentially split into two groups until no further subdivision is possible and each group has one part. The *sign matrix* is shown in Table 2.1, where the first D columns contain the entries for each component and the last two contain the number of parts of the first subgroup, r and the number of parts in the second subgroup, s. The *balance (b)* is the

Table 2.1: Example Binary Sequential Partition

Order	x_1	x_2	...	x_D	r	s
1	x_{1_1}	x_{1_2}	...	x_{1_D}	r_1	s_1
2	x_{2_1}	x_{2_2}	...	x_{2_D}	r_2	s_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
$D-1$	x_{D-1_1}	x_{D-1_2}	...	x_{D-1_D}	r_{D-1}	s_{D-1}

normalised logratio of the geometric mean of the two groups

$$b = \sqrt{\frac{rs}{r+s}} \ln \frac{(x_{i_1} x_{i_2} \ldots x_{i_r})^{1/r}}{(x_{j_1} x_{j_2} \ldots x_{j_s})^{1/s}} \tag{2.20}$$

where for the i^{th} balance each part receives a weight of

$$a_+ = +\frac{1}{r}\sqrt{\frac{rs}{r+s}} \tag{2.21}$$

for those on the numerator,

$$a_- = -\frac{1}{s}\sqrt{\frac{rs}{r+s}} \tag{2.22}$$

for those on the denominator, and $a_0 = 0$ for those not involved in the splitting. It is possible to populate a matrix Ψ of orthonormal basis with entries a_{ij} where a_{ij} is a_+ if the code from the sign matrix, at the i^{th} order partition, is +1 for the j^{th} part, is a_- if the code is -1 and a_0 if the code is a null. The matrix Ψ satisfies the requirement that $\Psi\Psi' = I_{D-1}$ and is used in the isometric logratio transformation (ilr) to produce coordinates x^* where:

$$x^* = ilr(x) = clr(x)\dot{\Psi}' \tag{2.23}$$

The coordinates are real variables and can be analysed using conventional statistical techniques. The inverse function restoring the coordinates to the simplex is:

$$x = C(\exp(x^*\Psi)). \tag{2.24}$$

A convenient summary diagram is the *balance-dendrogram* which carries information for the sequential binary partition in the form of a tree structure, the sample mean and variance of each ilr coordinate, as well as a box-plot summarising the order statistics of each ilr coordinate. In this case, the order statistics are the sample minimum, maximum, median and interquartile ranges. The horizontal portion of the plot contains each coordinate, the limits of which corresponds to the range (identical for every coordinate). As the range of each coordinate is symmetric, the relative position of the box-plot on the horizontal axis (i.e. closer to one part or group than another) indicates abundance of that

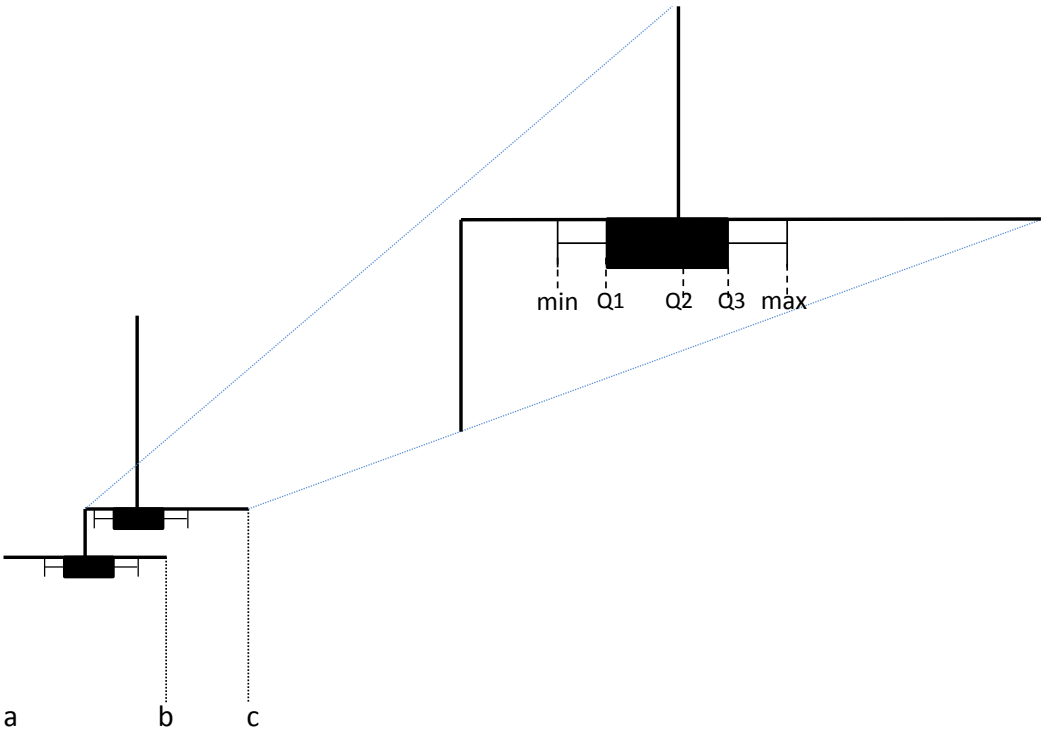


Figure 2.3: Schematic Balance-dendrogram (left) with exploded view of intersection and box-plot (right)

part (or group). The length of each vertical bar represents the variance of that coordinate and the contact point reflects the coordinate mean, about which the order statistics in the form of the box-plot are displayed. For example, in Figure 2.3, the information contained is that the majority of the variance is associated with the coordinates of a and c and that the relative abundance of c is greater than that of the group containing a and b.

No matter which approach is used, either *stay* or *move* the important property is that the transformations result in a one-to-one mapping of $\mathbb{S}^D \longrightarrow \mathbb{R}^{D-1}$ (Figure 2.4)

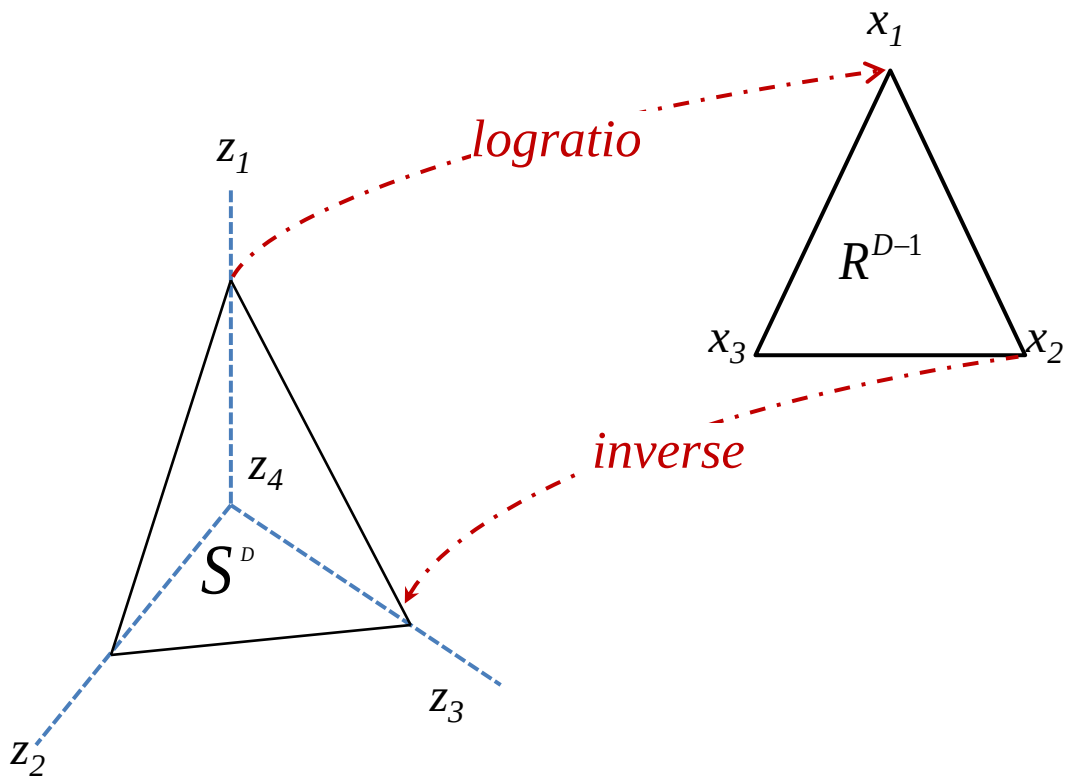


Figure 2.4: Representation of mapping using logratios between the *simplex* and the real space

2.6 Stationarity

Stationarity is an important consideration when deciding upon an interpolation method. In geostatistics a stationary process is a stochastic process whose joint probability distribution is invariant under translation over a given study area. A stationary random function is homogeneous and self-repeating in space and consequently parameters such as the mean and variance, in fact *all* the moments if they are present, do not change or follow any trends as a function of location. In practical terms stationarity is a decision made by the practitioner after consideration of the type and scale of the problem. *Second-order* stationarity is a weaker hypothesis in which only the mean and variance are invariant. This is captured by the following expectation:

$$\begin{aligned} E[\mathbf{Z}(u)] &= m \\ E[(\mathbf{Z}(u+h)-m)(\mathbf{Z}(u)-m)] &= Cov(h) \end{aligned} \tag{2.25}$$

for all distance vectors h and where $Cov(h)$ is the covariance of all pairs separated by h. Some authors suggest that stationarity is not a hypothesis that can be tested and proven (or otherwise) by the data [29], while others [30] suggest that testing the hypothesis is

possible using the following relationship:

$$\lim_{h \to +\infty} \frac{\gamma(h)}{h}^2 = 0 \tag{2.26}$$

When considering the problem we are applying geostatistical techniques to, it is important to understand the impact of scale on the decision of stationarity. Armstrong [30] provides a simple rule of thumb which states if the *scale* we are considering is dominated by trend, then assuming stationarity is not advised; however if the random fluctuations dominate, then our decision of stationarity is likely acceptable. Although in a theoretical context the domain in which we are working may not be perfectly stationary; the practical drawbacks caused by assuming a form of stationarity may be inconsequential. This has driven the acceptance of lower orders of stationarity in mining applications, specifically the *Intrinsic Hypothesis* which is a milder hypothesis characterised by the following relationships:

$$\begin{aligned} E[\mathbf{Z}(u+h) - \mathbf{Z}(u)] &= 0 \\ Var[\mathbf{Z}(u+h) - \mathbf{Z}(u)] &= 2\gamma(h) \end{aligned} \tag{2.27}$$

where $\gamma(h)$ is the *semivariogram* function and which can account for some drift in the mean [22].

2.7 Spatial covariance structure

The spatial distribution of random variables and functions is described using spatial covariances or semivariograms and in the case of intrinsic random functions only with semivariograms. An important consideration in the selection of the appropriate spatial covariance structure is the presence or absence of symmetry as it is a prerequisite for cokriging with semivariograms and cross-semivariograms [31]. Each of the logratio specifications discussed in Section 2.5 has a (generally distinct) spatial covariance structure resulting from the transformation. Pawlowsky-Glahn and Olea [7] discuss each of these in detail; interested readers are suggested to refer to this excellent overview. They highlighted that for a complete characterisation of the spatial covariance structure for stationary regionalised compositions D^4 covariance functions are needed and in their calculation all possible logratios are accounted for [7]. Alternatively the clr cross-covariances appear on first inspection to offer another solution due to symmetry in the components (unlike the alr cross-covariances) however the associated matrix function is singular; conventional software is not able to correctly deal with this. The alr cross-covariances are useful in that asymmetry is retained, if present, but the asymmetric consideration of the components is shown to create differing spatial covariance structures. However, the study completed by Job [21] demonstrated that the overall outcome is independent of the selection of the divisor. In addition the direct logratio semi-variograms (*lr semi-variograms*) provide the

optimal solution as they can be modelled using existing software. The intrinsic spatial covariance structure of a regionalised D-part composition is completely defined by the the D^4 covariograms $\gamma_{ij.kl}(h)$

$$\gamma_{ij.kl}(h) = \frac{1}{2}\text{Cov}\left[\ln\left(\frac{Z_i(u)}{Z_k(u)}\right) - \ln\left(\frac{Z_i(u+h)}{Z_k(u+h)}\right), \ln\left(\frac{Z_j(u)}{Z_l(u)}\right) - \ln\left(\frac{Z_j(u+h)}{Z_l(u+h)}\right)\right]. \quad (2.28)$$

where $i, j, k, l = 1, 2, \ldots, D$. In the case of an intrinsic alr random function which is symmetric, then the number of covariance terms required reduces to $(D-1)D/2$, the corresponding cross-semivariograms are required and they are defined by

$$\gamma_{i.j}(h) = \gamma_{ii.jj}(h), \quad (2.29)$$

and the alr cross-semivariogram function is given by

$$\mathbf{\Psi}(h) = \psi_{ij}(h) = \gamma_{ij.DD}(h) \quad (2.30)$$

where the alr semivariogram is the special case where $i = j$. Interestingly; each alr cross-semivariogram is dependent only on the auto-semivariograms of the pairwise logratios of the involved components and the component used as a divisor. This is due to the following correspondence:

$$\ln\left(\frac{Z_i(u)}{Z_j(u)}\right) = \ln\left(\frac{Z_i(u)}{Z_D(u)}\right) - \ln\left(\frac{Z_j(u)}{Z_D(u)}\right) \quad (2.31)$$

and therefore the estimation and modelling of the semivariogram of the following quantity

$$\ln\left(\frac{Z_i(u)}{Z_D(u)}\right) - \ln\left(\frac{Z_j(u)}{Z_D(u)}\right), \quad (2.32)$$

is the same as the estimation and modelling of $\gamma_{i.j}(h)$. The correspondence for every lag h is the following:

$$\psi_{ij}(h) = \frac{1}{2}(\gamma_{i.D}(h) + \gamma_{j.D}(h) - \gamma_{i.j}(h)) \quad (2.33)$$

As in classical geostatistics, the covariance function and the semivariogram function $\mathbf{\Psi}$ are related via the equation

$$\mathbf{\Psi}(h) = \mathbf{\Sigma}(0) - \mathbf{\Sigma}(h), \quad (2.34)$$

so that fitting of a linear model of coregionalisation (LMC) using standard variography tools is possible.

2.8 Cokriging

A common problem in earth sciences is the best linear unbiased (BLU) estimation of single or multiple values at unsampled locations. Kriging is a method of interpolation

for which the interpolated values are modelled by a process, which may or may not be Gaussian, governed by prior covariances. Under suitable assumptions on the priors (see Section 2.6 describing stationarity), kriging gives the best linear unbiased prediction of the intermediate values. Chilès and Delfiner [22] note that kriging appears especially well suited to Gaussian random functions and while kriging still provides the best linear estimator for non-Gaussian cases, the linear estimators may not be efficient due to heteroscedasicity. Cokriging utilises the cross-correlations sometimes present between variables to reduce the variance of the estimation error further than kriging each variable independently thus cokriging is most useful when one of the variables is under-sampled with respect to another [10]. The following description primarily follows Myers' [31] discussion of cokriging focusing initially on the stationary and then the intrinsic case.

2.8.1 The stationary case

Let $Z_1, \ldots, Z_D$ represent the continuous random variables in the multivariate estimation problem. The informed or sampled locations are denoted by $\mathbf{I} = \{u_1, u_2, \ldots, u_n\}$. Given these data the objective is to estimate $\mathbf{Z}(u)$ using the $\mathbf{I}$ informing data where:

$$\mathbf{Z}(u) = [Z_1(u), \ldots, Z_D(u)]. \tag{2.35}$$

In order to fulfill the aim of a BLU estimator a linear combination of of $\mathbf{Z}(u_i)$, $1 \leq i \leq n$ such that the estimation variance is minimised must be constructed. The estimator is of the form:

$$\mathbf{Z}^*(u) = \sum_{k=1}^{n} \mathbf{Z}(u_k)\Gamma_k \tag{2.36}$$

where each $\Gamma_k = [\lambda_{ij}^k], 1 \leq i, j \leq D$ is an $D \times D$ matrix of weights. Thus the λ_{ij}^k element represents the contribution of the i_{th} variable at location u_k, to the estimate of the j_{th} variable. Since this is a linear expectation of random functions:

$$E[\mathbf{Z}^*(u)] = \sum_k E[\mathbf{Z}(u_k)]\Gamma_k, \tag{2.37}$$

in the stationary case the necessary and sufficient condition for $\mathbf{Z}^*$ to be unbiased is for

$$\sum_{k=1}^{n} \Gamma_k = I. \tag{2.38}$$

The estimation variance can be defined as

$$\sum_{i=1}^{D} Var[Z_i(u) - Z_i^*(u)]. \tag{2.39}$$

When equation 2.38 is true equation 2.39 can be written as

$$E[(\mathbf{Z}(u)-\mathbf{Z}^*(u))(\mathbf{Z}(u)-\mathbf{Z}^*(u))^T] \tag{2.40}$$

where T denotes the transpose [31]. Equation 2.40 can be reformulated in terms of covariances through the following relation:

$$E[Z_i(u)Z_j(v)] = Cov_{ij}(u-v) \tag{2.41}$$

where v is another, possibly different location. In the case of *ordinary cokriging (OCK)* the unbiasedness requirement demands the use of Lagrange multipliers. Let the Lagrange Multipliers ($\overline{\nu}$) be an $D \times D$ matrix of elements ν_{ij}. These are necessary in the minimisation exercise which is achieved by solving the system of equations:

$$\begin{aligned} &\frac{\partial \phi}{\partial \lambda_{ij}^k} = 0, \quad k = 1,\ldots,n; \quad i,j = 1,\ldots,D \\ &\frac{\partial \phi}{\partial \nu_{ij}} = 0 \end{aligned} \tag{2.42}$$

where ϕ is the restatement of equation 2.40 in terms of Γ_k and $\overline{\nu}$. These partial derivatives simplify to:

$$\begin{bmatrix} \overline{Cov}(u_1-u_1) & \cdots & \overline{Cov}(u_1-u_n) & I \\ \vdots & \vdots & \ddots & \vdots \\ \overline{Cov}(u_n-u_1) & \cdots & \overline{Cov}(u_n-u_n) & I \\ I & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \\ \overline{\mu} \end{bmatrix} = \begin{bmatrix} \overline{Cov}(u_1-u) \\ \vdots \\ \overline{Cov}(u_n-u) \\ I \end{bmatrix} \tag{2.43}$$

In the case of kriging a single variable, equation 2.43 has the identical form except that entries are scalars, not matrices [31].

2.8.2 The Intrinsic case

The special case of *Intrinsic Random Functions* discussed in Section 2.6 allows for the use of variogram functions. This is also true of the cokriging system which can utilise cross variograms in the case of symmetrical cross-covariances [31]. Because cokriging considers multiple variables, and therefore equations of matrices, equation 2.27 must be altered to the following:

$$E[Z_i(u+h)-Z_i(u)] = 0 \quad i = 1,\ldots,\ldots,m \tag{2.44}$$

which implies

$$E[\mathbf{Z}(u+h)-\mathbf{Z}((u)] = [0,0,\ldots,0] = \mathbf{0} \tag{2.45}$$

and that

$$Cov[Z_i(u+h)-Z_i(u),Z_j(u+h)-Z_j(u)] = 2\gamma_{ij}(h) \tag{2.46}$$

exists and is dependent only upon h for all $i, j = 1,\ldots,m$. When equation 2.45 is true then equation 2.46 becomes

$$\frac{1}{2}E[\mathbf{Z}((u+h)-\mathbf{Z}((u)]^T[\mathbf{Z}((u+h)-\mathbf{Z}((u)] = \overline{\gamma}(h) = [\gamma_{ij}(h)]. \tag{2.47}$$

If $C_{ij}(h)-C_{ij}(-u)-C_{ij}(u+h) = C_{ij}(-h)-C_{ij}(u)-C_{ij}(-u-h)$ which is true when $C_{ij}(v) = C_{ij}(-v)$ for all i, j, v i.e. symmetry then we can also derive the following

$$E[\mathbf{Z}((u+h)-\mathbf{Z}((u^*)]^T[\mathbf{Z}((u-h)-\mathbf{Z}((u^*)] = \overline{\gamma}(u+h)+\overline{\gamma}(u)-\overline{\gamma}(h). \tag{2.48}$$

In the intrinsic case the estimator Z^* has the same form as in the stationary case equation 2.36 with the same unbiasedness condition shown in equation 2.38. In order to maintain the same unbiasedness constraint the equation 2.48 is used to restate the estimation variance in terms of the variogram. Similarly to the stationary case partial derivatives are taken to obtain the following system of equations:

$$\begin{bmatrix} \overline{\gamma}(u_1-u_1) & \cdots & \overline{\gamma}(u_1-u_n) & I \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\gamma}(u_n-u_1) & \cdots & \overline{\gamma}(u_n-u_n) & I \\ I & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \vdots \\ \Gamma_n \\ \overline{\nu} \end{bmatrix} = \begin{bmatrix} \overline{\gamma}(u_1-u) \\ \vdots \\ \overline{\gamma}(u_n-u) \\ I. \end{bmatrix} \tag{2.49}$$

The minimised estimation variance (*kriging variance*) can then be expressed in terms of the variogram function as the following:

$$\sigma_K^2 = Tr[\sum_{i=1}^{n}\gamma(u-u_i)\Gamma_i]+Tr\overline{\nu} \tag{2.50}$$

where Tr denotes the trace of a matrix.

2.8.3 Compositional Cokriging

Myers' work deals with the general case of unconstrained sample spaces; the work of Pawlowsky-Glahn and Olea [7] is an extension of this to constrained samples spaces. The core of the approach used in this work is known as *alr cokriging*. The central premise used in this approach is the mapping of the constrained sample space (*simplex*) to the unconstrained real space $\mathbb{R}$ using the alr transformation described in section 2.5. In essence the aim is to estimate $\mathbf{Z}$

$$\mathbf{Z}^*(u) = agl\left(\sum_{i=1}^{n}\overline{X}(u_i)\Gamma_i\right). \tag{2.51}$$

based on the covariance structure of $X(u_i)$ where

$$X(u_i) = \text{alr}(Z(u_i)) \tag{2.52}$$

Under the conditions of second-order stationarity, complete knowledge of the covariance functions for every lag h, and $\mathbf{X}(u)$ following a multivariate normal, lognormal or additive normal distribution then equation 2.51 satisfies the following conditions:

1. The estimates are of the center of the distribution defined as $agl(E[\mathbf{X}(u)])$ and are unbiased in that the construction of the ordinary cokriging estimator $\mathbf{X}^*(u)$ is unbiased with respect to $\mathbf{X}(u)$ at any given location u.
2. The squared Euclidean distance between $\mathbf{X}^*(u)$ and $\mathbf{X}(u)$ - the estimation variance, is minimised.

Importantly Pawlowsky-Glahn and Olea [7] noted that as there is *no* explicit equation which relates $E[\mathbf{Z}(u)]$ and $E[alr(\mathbf{Z}(u))]$ only ordinary cokriging and not simple cokriging can be applied to alr transformed vectors.

2.9 Gauss-Hermite approximations

In equation 2.51 the agl back transformation function is used to estimate the expected value. In fact, due to the non-linear nature of the alr transform and the linear combinations used in cokriging, bias of the back transformed variables is *guaranteed* in the simplex if the agl is applied directly. Job [21] illustrated the non-additivity of logratios and highlighted that the bias will be greatest for the dominant component. Unbiased estimates of the expected value of components can be obtained by calculating the integral

$$\overline{\mu}_{\mathbf{z}}(u) = E(\mathbf{Z}(u)) = \int_{S_D} \mathbf{Z}(u) f(\mathbf{Z}(u)) d\mathbf{Z}(u), \tag{2.53}$$

and estimates of the estimation variance through

$$\Sigma_z(u) = Cov[\mathbf{Z},\mathbf{Z}] = \int_{S_D} (\mathbf{Z}(\mathbf{u}) - \overline{\mu}_{\mathbf{z}}(\mathbf{u})) (\mathbf{Z}(\mathbf{u}) - \overline{\mu}_{\mathbf{z}}(\mathbf{u}))^T f(\mathbf{Z}(u)) d\mathbf{Z}(u). \tag{2.54}$$

When the data follow a multivariate normal distribution with mean $\overline{\mu}_{\mathbf{z}}(u))$ and variance-covariance matrix Σ_x (obtained from ordinary compositional cokriging) then R can be calculated as the square root of Σ_x obtained via the Cholesky decomposition, $\Sigma_{\mathbf{X}} = R'R$. Thus the integrals in 2.53 and 2.54 can be calculated as

$$\overline{\mu}_{\mathbf{z}}(u) = \int_{\mathbb{R}^{D-1}} \pi^{-\frac{D-1}{2}} \text{agl}\left(\sqrt{2} R^T \mathbf{Y} + \mathbf{X}^*(u)\right) \exp\left(-\mathbf{Y}^{\mathbf{T}}\mathbf{Y}\right) d\mathbf{Y} \tag{2.55}$$

and

$$\Sigma_{\mathbf{z}}(u) = \int_{\mathbb{R}^{D-1}} \pi^{-\frac{D-1}{2}} \mathbf{M}\mathbf{M}^{\mathbf{T}} exp(-\mathbf{Y}^{\mathbf{T}}\mathbf{Y}) d\mathbf{Y} \tag{2.56}$$

respectively, where $\mathbf{M} = \text{agl}(\sqrt{2}R'\mathbf{Y} + \mathbf{X}^*(\mathbf{u})$ and $\mathbf{Y} = \frac{1}{\sqrt{2}} \left(R^{-1}\right)^T (\text{alr}(\mathbf{Z}) - \mathbf{X}^*(\mathbf{u}))$. The evaluation of the equation must be done via numerical integration and due to the multivariate normality Gauss-Hermite Quadrature can be applied such that

$$\int_{\mathbb{R}^{D-1}} g(\mathbf{Y}) \exp(-\mathbf{Y}\mathbf{T}\mathbf{Y}) d\mathbf{Y} = \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_D=1}^{k} \prod_{\ell=1}^{D} w_{i_\ell} g(Y_{i_1}, Y_{i_2}, \ldots Y_{i_D}) \tag{2.57}$$

where $w_1, w_2, \ldots, w_k$ are the known weights, $Y_1, Y_2, \ldots, Y_k$ are the known abscissa. Using Equation 2.57 to compute $\overline{\mu}_{\mathbf{z}}(u)$ take

$$g(\mathbf{Y}) = \pi^{-\frac{D-1}{2}} \text{agl}\left(\sqrt{2}R^T\mathbf{Y} + \overline{\mu}_{\mathbf{x}(\mathbf{u})}\right) \tag{2.58}$$

whereas for the computation of $\Sigma_{\mathbf{Z}(\mathbf{u})}$ take

$$g(\mathbf{Y}) = \pi^{-\frac{D-1}{2}} \left(\text{agl}\left(\sqrt{2}R^T\mathbf{Y} + \overline{\mu}_{\mathbf{x}(\mathbf{u})}\right) - \overline{\mu}_{\mathbf{z}(\mathbf{u})}\right) \left(\text{agl}\left(\sqrt{2}R^T\mathbf{Y} + \overline{\mu}_{\mathbf{x}(\mathbf{u})}\right) - \overline{\mu}_{\mathbf{z}(\mathbf{u})}\right)^T. \tag{2.59}$$

Functionally, numerical integration assigns weights to the integrand at each abscissa to quantify the approximation, hence in general the higher the order k of the quadrature implemented, the more accurate the approximation.

2.10 Estimation quality test procedures

Two techniques are used to evaluate the applicability and robustness of the estimation approaches selected for this study.

2.10.1 Cross-validation

Chilès and Delfiner [22] define cross-validation of a dataset with n samples as the estimation of $Z^*(u_i)$ at each sample location u_i using the neighbouring data $Z(u_j) : i = j \neq 1, 2, \ldots, n$. Although cross-validation was developed and implemented originally as a variogram validation procedure, it is possible to evaluate other estimation techniques using it which is the approach used in this study. The critical knowledge gained is the *true error* ($\delta_i = Z^*(u_i) - Z(u_i)$. Further to δ_i, other useful statistics can be generated including the mean absolute error $\overline{\delta_u} = \frac{1}{n}\sum_{i=1}^{n} \delta_i$ and root mean square error $\overline{\Delta_u}^2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\delta_i)^2}$.

2.10.2 Jack-knifing

Jack-knifing is a precursor to the cross-validation technique described in subsection 2.10.1. Whereas cross-validation is quite specific to geostatistics, jack-knifing is a generalised statistical technique used originally to reduce bias and later developed to set appropriate confidence limits for estimates [32]. *Sensu stricto* jack-knifing refers to resampling without replacement. Geostatistical jack-knifing is similar to the cross-validation technique except a subset of the data set is removed rather than a single datum; this results in an alternative set of data being estimated from a *non-overlapping* set. Functionally, a data set of size n is split into two subsets, the locations to be estimated $Z(u_i) : i = 1,2,\ldots,n$, and $Z(u_j) : i = j \neq 1,2,\ldots,n$, the informing data. This can be repeated with alternate $Z(u_i)$ subsets of differing sizes and locations; the repetition filters out statistical fluctuation caused by selection of only one dataset [33]. Further, as the informing or non-overlapping data are not used to compare histograms or variograms with the resulting estimates, jack-knifing is seen as a more rigorous technique [33] [22]. The main advantage is the same as that from cross-validation, computation of the *true error* (δ_u) and the other statistics outlined in 2.10.1.

2.11 Accuracy and Precision

Confidence limits in the univariate case, and confidence regions in multivariate instances, have been used historically as alternatives to point estimates for unknown parameters. Confidence intervals and regions are helpful in assessing the accuracy and variability of estimators, and in making decisions about the unknown parameters [7]. Deutsch [34] provides a framework for the assessment of *local* accuracy and precision of estimates resulting from different methodologies. The estimation of an error variance associated with the expected value is one of the distinguishing features of probabilistic methods and kriging in particular (See Section 2.8). When used in conjunction with a method which leaves data with which to compare the estimates against true values such as cross-validation or jack-knifing, a powerful comparative tool can be constructed. In this application, Deutsch [34] defines *accuracy* as the proportion a true value falls within fixed symmetrical probability intervals defined by the probabilistic estimator. The *precision* is defined as the closeness of the accurate distribution to the ideal distribution. Precision is meaningless without accuracy when used in this framework, as a constant value (zero variability) would have the ultimate precision, but be a very poor estimator of the true value locally. The process described below follows very closely that suggested by Deutsch [34]. Local conditional cumulative distribution functions (ccdf) can be calculated from the expected value and estimation standard deviations at each location under the assumption of normally distributed errors. From these local ccdf's probabilities associated with the true values are calculated.

The ccdf is defined as:

$$F(u_i;z|(u_j)) = Prob(Z(u_i) \leq z|(u_j)), \tag{2.60}$$

where as in Section 2.10.1 and 2.10.2 $Z(u_j) : i = j \neq 1,2,\ldots,n$. The probabilities associated to the true values from Equation 2.60 are:

$$F(u_i;z(u_i)|u_(j)),\ i = 1,\ldots,n \tag{2.61}$$

Symmetric p-probability intervals are constructed by setting

$$p_{low} = \frac{(1-p)}{2} \text{ and } p_{upp} = \frac{(1+p)}{2}, \tag{2.62}$$

thus for $p = 0.9$ the 90% probability interval (PI), the lower limit is $p_{low} = 0.05$ and the upper limit $p_{upp} = 0.95$. Further an indicator function $\varepsilon(u_i;p)$ is defined at each sample location u_i

$$\varepsilon(u_i;p) = \begin{cases} 1, & \text{if } F(u_i;z(u_i)|(u_j)) \in (p_{low}, p_{upp}) \\ 0 & \text{otherwise.} \end{cases}$$

Averaging $\varepsilon(u_i;p)$ over the n data locations generates the proportion $\overline{\varepsilon(p)}$ of locations where the true value falls within the symmetrical PI. The algorithm is defined as accurate when $\overline{\varepsilon(p)} \geq p$, for all p. This quantity can be graphically displayed by plotting $\overline{\varepsilon(p)}$ against p, known as an *accuracy plot*. By construction, points plotting above the 45° line meet the definition of accuracy defined above, while points below below the 45° line do not. A quantitative statistic (A),

$$A = \int_0^1 a(p)dp, \tag{2.63}$$

which can be used to compare one distribution's accuracy against that of another, is generated by an indicator $a(p)$ defined for each probability interval $p \in [0,1]$

$$a(p) = \begin{cases} 1, & \text{if } \overline{\varepsilon(p)} \geq p \\ 0 & \text{otherwise.} \end{cases}$$

A is approximated numerically by a discrete summation over K probability values $p_k, k = 1,\ldots,K$. The best case is A = 1.0 indicating maximum accuracy and A = 0.0 for the worst case where no true values fall within the probability intervals [34]. The *precision* is defined as the closeness of $\overline{\varepsilon(p)}$ to p if the distribution is accurate over that interval i.e. $\overline{\varepsilon(p)} \geq p$. This measure ($P$) can be calculated by

$$P = 1 - 2\int_0^1 a(p)[\overline{\varepsilon(p)} - p]dp, \tag{2.64}$$

where similarly to the Equation 2.63:

- P =1.0 is maximum precision and P = 0.0 is the worst case
- P is approximated numerically by a discrete summation over K probability values $p_k, k = 1, \ldots, K$.

The summary statistics A and P do not account for the case where the true values falls outside of the PI, i.e. where $\overline{\varepsilon(p)} < p$. The goodness can be quantified through G the departure of the points from the 45° line with the inaccurate points penalised more severely. The integral

$$G = 1 - \int_0^1 [3a(p) - 2][\overline{\varepsilon(p)} - p]dp \tag{2.65}$$

is approximated numerically as for precision and accuracy with the maximum value for Goodness being 1 and the minimum 0.

2.12 Error measures

The objective of estimation is to maximise the accuracy and precision of the estimate with respect to the true value. This is explicit in the design of the kriging algorithms to be unbiased and with minimum error variance. This is captured using a distance measure of some type, of which there are many to choose. Martin-Fernandez *et al* [35] highlighted that the measure(s) should be selected for ease of computation and interpretation. They expand upon these criteria by suggesting that the measure should conform to the sample space of the data set; and ideally to be one dimensional. The one dimensional aspect is important as we need to test the *combined goodness* of the estimate of all the attributes at a particular location as well as each attribute separately. Traditional statistical description can then take place on the one dimensional measurement or error variable.

2.12.1 Univariate measures

In Section 2.10 the univariate error measure, the arithmetic difference (δ_u) was described. If this measure is applied directly in the simplex only the one dimensional criterion outlined above is satisfied as negative values can result. Taking the absolute value of the difference would address this but would require a different interpretation and would not be immediately useful as a measure of the bias. This method is commonly used in the earth sciences to generate an estimate of mean bias. We define this error measure as follows:

$$\delta_i = Z_i^*(u) - Z_i(u). \, , i = 1, \ldots, D \tag{2.66}$$

In fact this is the same error we seek to minimise in the kriging equations; except of course there is no knowledge of the true value in that case and thus the true error is unknown.

2.12.2 Compositional measures

In addition to the recommendation in [35] to use distance or error measures of one dimension, they also advocate using different errors for different sample spaces. The principal calculations are carried out in the D-simplex S^D where the *Euclidean* error, defined as follows, is used:

$$\delta_e^2(\mathbf{X}(u),\mathbf{X}^*(u)) = \sum_{i=1}^{D}(X_i(u) - X_i^*(u))^2. \tag{2.67}$$

where $\mathbf{X}(u)$ is a coregionalisation with $\mathbf{X}^*(u)$ as its estimate with matching number D of observations and estimates. In a strict statistical sense using Euclidean error is equivalent to assuming equal variances and absence of correlation which may or may not be met in practice [35]. The alternative choice is the *Aitchison* error δ_a^2 defined as:

$$\delta_a^2(\mathbf{Z}(u),\mathbf{Z}^*(u)) = \sum_{i=1}^{D}\left(ln\frac{Z_i(u)}{g(Z(u))} - \frac{Z_i^*(u)}{g(Z^*(u))}\right)^2. \tag{2.68}$$

where $\mathbf{Z}(u)$ and $\mathbf{Z}^*(u)$ are as defined in equation 2.67 and $g(\mathbf{Z}(u))$ and $g(\mathbf{Z}^*(u))$ are the geometric mean of the components of $\mathbf{Z}(u)$ and $\mathbf{Z}^*(u)$ respectively.

2.12.3 STRESS

Two different estimation methods will be compared; a global measure of performance well suited to this task is the *S*tandardised *Re*sidual *S*um of *S*quares (*STRESS*). The comparison is is therefore of δ_{ab}, the difference between observations $\mathbf{Z}(u_a)$ and $\mathbf{Z}(u_b)$ and δ_{ab}^* the difference between observations $\mathbf{Z}^*(u_a)$ and $\mathbf{Z}^*(u_b)$ where δ_{ab} and δ_{ab}^* use the same type error measure, say Euclidean. Then

$$STRESS = \left[\frac{\sum_{a<b}(\delta_{ab} - \delta_{ab}^*)^2}{\sum_{a<b}\delta_{ab}^2}\right]^{\frac{1}{2}}. \tag{2.69}$$

The STRESS can be interpreted as a normed loss of information induced by replacing the true values with the estimates; thus the smaller the value of STRESS, the better the quality of the estimate. As with the other one dimensional error measures, analysis of the base statistics such as central tendency and spread, are used to further analyse the results of the STRESS calculation [35].

Chapter 3

Data, Geology and Sampling methodologies

3.1 Geology

Cliffs Natural Resources operates the Koolyanobbing Iron Ore Project containing the Windarling mining hub. At Windarling iron ore is mined from open pits located in the Marda-Diemals Greenstone Belt of the Southern Cross Domain. Prior to mining operations commencing in 2004, the Windarling deposits formed a prominent east-west oriented range approximately 25 km north of Mt Jackson and approximately 90 km north of the Koolyanobbing mining hub complex (figure 3.1).

3.1.1 Windarling Geological Setting

The Marda-Diemals (Windarling) greenstone belt is located within the Yilgarn craton. The belt, placed within the Youanmi Terrane, an amalgamation of the Murchison and Southern Cross domains, is thought to represent a protocraton onto which the younger Eastern Goldfields and Narryer Terrane were accreted. This greenstone belt differs from surrounding belts due its west-northwest trend and is divided into a lower and upper sequence [36].

3.1.2 Windarling Regional Geology

The lower sequence of the Marda-Diemals greenstone belt is thought to be a mafic volcanic-dominated succession of about c. 3.0 Ga. This sequence is unconformably overlain by an upper (c. 2.73 Ga) sequence comprising felsic volcanic and associated volcanoclastic-

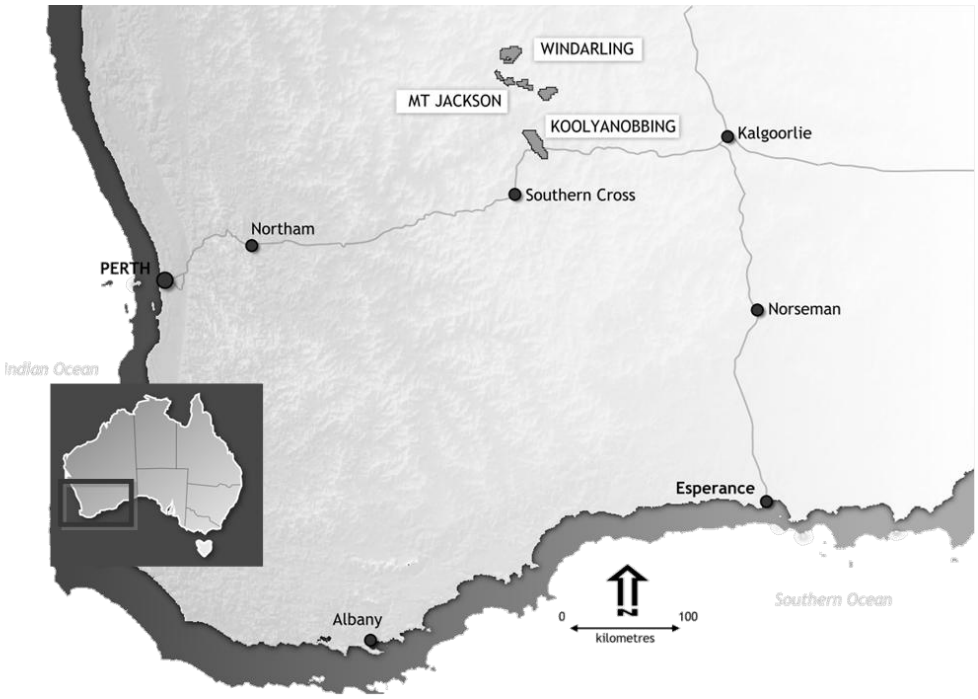


Figure 3.1: Location of the Cliffs NR Windarling open pits

sedimentary rocks of the Marda Complex as well as clastic metasediments of the Diemals Formation. Younger granitoid (principally monzogranitic) intrusions have been dated at c. 2.73-2.65 Ga; three major periods of deformation are further recorded and detailed Riganti and Chen [37]. Greenschist facies metamorphism predominates in the lower greenstone succession. High grades (up to amphibolite facies) have been recorded from granite-greenstone contacts and in sediments occurring in the northernmost part of the Marda Complex.

3.1.3 The local geology of W3

The W3 deposit forms the most prominent ridge of the Windarling Range. The stratigraphy dips approximately 50 degrees south, with mineralisation extending over a strike length of approximately 950 metres. Up to three parallel zones of iron mineralisation, over a width of approximately 40 metres, are separated by lenses of sheared magnesium and iron rich volcanic (mafic) material. The ore is hard, fine to medium grained, bedded, and comprised of goethite magnetite hematite. The W3 deposit is characterised by medium to high Fe grade (>63 weight percent Fe), relatively high phosphorus and uniformly low sulphur. The W3 deposit lies immediately to the south of the W5 deposit (Figure 3.2. W5 consists of up to three relatively narrow, discontinuous bands of hematite-goethite mineralisation forming a low arcuate ridge immediately south of the W3 deposit. The W5 deposit is characterised by relatively high Fe grades (>63 weight percent), moderate

phosphorus and uniformly low sulphur. The mineralisation is interspersed along strike with areas of hematite rich jaspilite. In plan view the BIF bands appear to have been thickened by longitudinal compression to create a series of imbricated structures with a strike direction change approximately at the midpoint (721165 m Easting) of the study area.

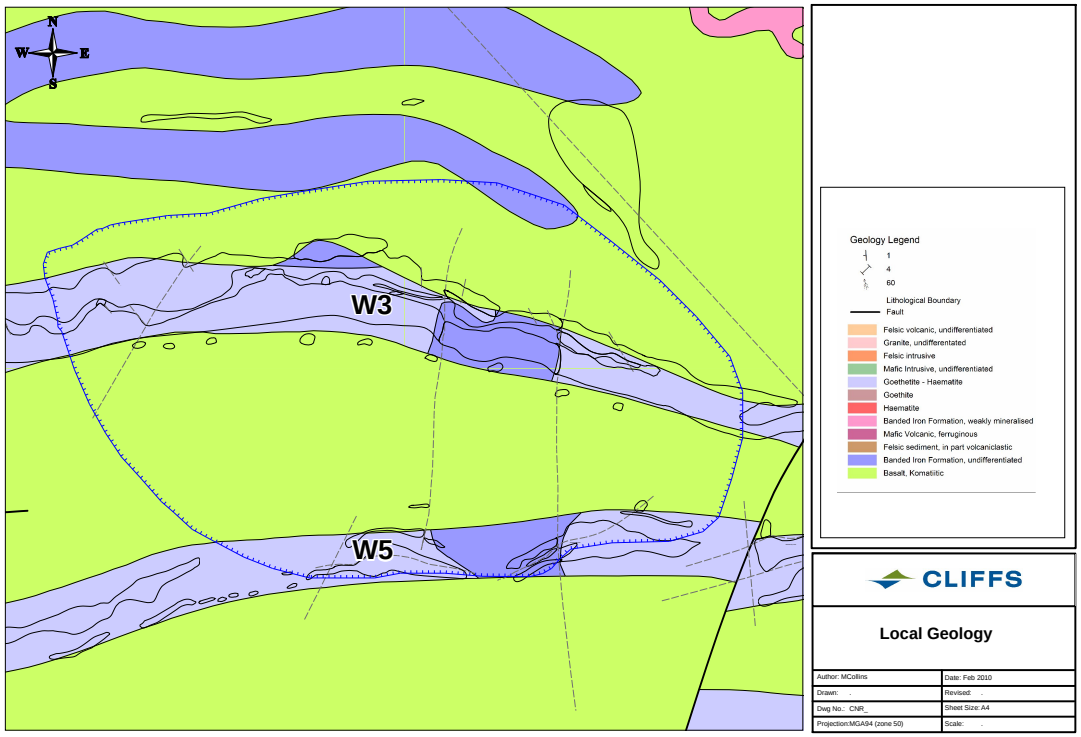


Figure 3.2: Local geological mapping of W3 and W5. Courtesy Cliffs NR

3.1.4 Iron concentrating mechanisms

Angerer and Hagemann [38] noted that the structural control of high grade (>58 weight percent Fe) BIF-hosted ore deposits is considered to be a critical factor. They further list the principal iron ore-formation models postulated to create economic iron oxide concentrations. These are:

1. syngenetic models that propose synsedimentary structures such as boudinage or extensional faults which lead to chert-free BIF formation,
2. supergene and supergene-metamorphic models which operate primarily by the circulation of meteoric fluids through existing structures eliciting upgrade of the BIF by leaching of gangue minerals and,
3. hypogene models in which hydrothermal fluid flow associated with deformation is deemed critical for iron oxide mineralisation in the low stress portions of structures.

The most likely processes that cause upgrade in the Windarling (Marda) greenstone belt are analogous to those which occurred in the Koolyanobbing greenstone belt to the south. Angerer and Hagemann [38] postulate the same multistage, structurally controlled iron ore genesis for the deposits in the Windarling range with added structural high-strain zones that have resulted in a remarkable variety of pre-, syn- and post-deformational ore textures. Angerer *et al* [39] describe the Windarling genesis in the following way (figure 3.3):

- Trace element analyses of the BIF indicate precipitation from Archean seawater that was fertilised by hydrothermal vent fluids with a basaltic signature. The ore genesis then proceeded through four main stages
- Stage 1 was a syn- to post-metamorphic metasomatism that produced local Fe-dolomite-magnetite alteration in BIF. Hydrothermal alteration was induced by hot fluid flow controlled by brittle-ductile reactivation of BIF-basalt margins and cross-cutting D1 faults. Analysis of dolomite and carbonate-altered BIF suggests that the basalts in the Windarling Range were the primary source of introduced metals.
- Stage 2 resulted in high-grade magnetite-hematite ore forming from syn-deformational fluid flow along BIF-basalt margins.
- Stage 3 is the relatively minor stage of remobilising iron oxides, carbonate and quartz to form veins and breccia.
- Stage 4 involved recent (Mesozoic to present) supergene oxidation and hydration in a weathering environment reaching down to depths of maximum 200 m below surface

Irrespective of the ore genesis theories, the outcome is a series of discrete enrichments of Fe within the broader Marda (Windarling) BIF range. In total, nine economic enrichments have been identified along the range and form a spectrum of grades. W3 is on the high end of the spectrum with a mean grade of ˜63 weight % Fe and locally blast hole analyses regularly in excess of 66 weight % Fe.

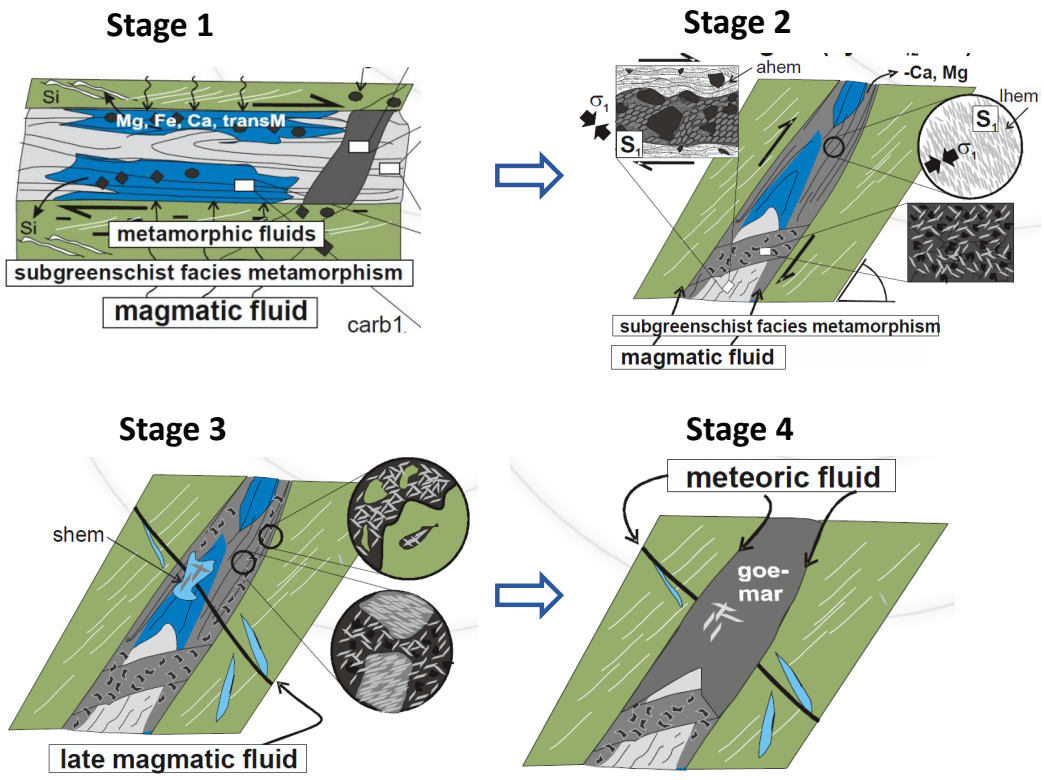


Figure 3.3: Windarling ore genesis model — Ore formation (after Angerer and Hagemann [38])

3.2 Sampling methods

The evaluation of mineral deposits is generally based upon geochemical analyses of rock, sand or other natural material obtained from a repeatable method. The kriging paradigm enforces the notion that (for a given variogram and search neighbourhood) more regularly gridded samples will result in superior estimates of the in-situ material than fewer samples [40]. As sampling of the material is expensive and the required confidence intervals around estimates are different for each stage of a mining project, different sampling methodologies and equipment are used for each stage. For example, at the scoping stage, fewer data are required as the output of the evaluation exercise will likely be a mean estimate with confidence limits. This contrasts with the requirement for a high degree of accuracy and local grade interpolation needed to delineate operational Run of Mine (ROM) material from waste during the mining stage. At the Windarling mining hub and specifically at W3, ROM sampling is achieved by taking ˜3 kg samples from blast hole cuttings piles left at the collar (top of the hole). Blast holes are drilled for the purposes of fragmenting the material through explosive blasting to facilitate loading and stockpiling. The holes are drilled vertically on a nominal 3 m x 3.5 m grid with a (sampling) length matching the bench height (Figure 3.4). The bench height used at W3 is 6 vertical meters and has little relationship to the grade characteristics of the geological formations.

Blast holes are sampled opportunistically as a low cost alternative to other methods such as Reverse Circulation (RC). RC drill samples, especially when the drill hole is drilled perpendicularly to the gross dip of the mineralised formation, are generally considered superior to the vertical blast holes [41]. Additionally, RC samples are generally taken at smaller intervals; increasing the edge resolution between different geological units, a desirable quality, unlike blast holes samples which tend to blur the edges between geological units. This is especially true if the geological units occur on a scale less than the blast hole interval. As a result, over a common volume, it is highly probable that RC data will be differently distributed both spatially and statistically from blast hole data [42].

Figure 3.4: Blast hole rig on pattern; cuttings piles at collar easily visible (Photo courtesy Cliffs NR

3.3 Data description

The dataset selected for this study consists of a single complete bench of ROM blast hole analyses from the W3 open pit at Windarling (Figure 3.5). At this elevation through the steeply dipping ore body the following features are noted:

1. the blast hole analyses selected describe the ROM material only, not the waste,
2. the mineralisation trends approximately East-West,
3. the bench is composed of two, discrete, parallel, mineralised BIF lithologies with internal waste,

4. the mineralisation appears as two distinct limbs with different strike orientations, which hinge about the ˜721165 X-coordinate, and
5. each limb is further broken into discrete faulted blocks some of which display rotation relative to the adjacent block and in some cases minor imbrication.

The data consist of 1594 geographic locations with geochemical analyses for the following analytes

1. Fe
2. SiO_2
3. Al_2O_3
4. S
5. P
6. Mn
7. LOI Loss on Ignition, the volatile content driven off the ore during heating beyond 950°C.

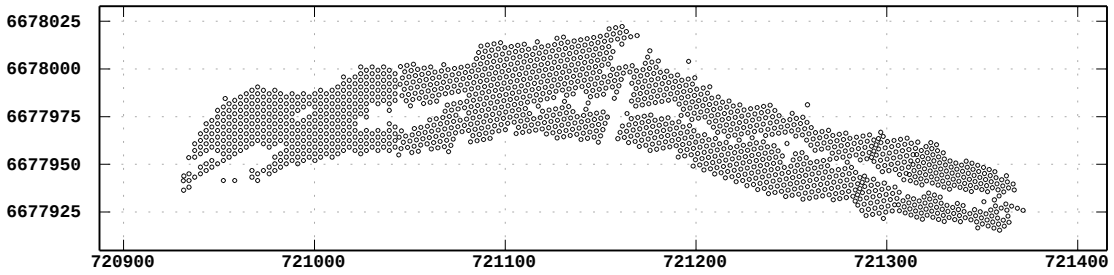


Figure 3.5: Plan view of blast hole analyses from W3 open pit

The blast hole analyses list the seven principal analytes of importance for the smelting of the ore in a blast furnace. Although this is not an exhaustive list of possible analytes these seven generally constitute the bulk of the material delivered to a furnace. The analytes (Al_2O_3, Fe, Loss on Ignition (LOI), Mn, P, S, SiO_2) are a mixture of oxides and native element abundances and therefore form a subcomposition of the total mass of the sample; in other words, the total analysed does not total a constant value (Figure 5.1). Due to this a filler variable was introduced in order to satisfy the constant sum constraint. Consequently, the filler represents the missing oxygen mass not accounted from the iron, manganese, phosphorus, and sulphur oxide minerals as well as the small concentration of other elements and oxides not reported. The reason the analyses are routinely performed this way is to account for the different iron oxide speciation, i.e. the analyses are consistent whether magnetite (Fe_3O_4) or hematite (Fe_2O_3) is considered as the customers are interested in total iron units.

Chapter 4

Implementation

The flow chart in Figure 4.1 encapsulates the major processing steps and outputs in this study. The process involved the use of four primary software packages:

1. Geovariances Isatis
2. Microsoft Office Excel and embedded Visual Basic for Applications
3. CoDaPack
4. MATLAB

Several of the steps require the application of a process to each individual observation and therefore scripting was required. A simplified process flowchart is shown in Figure 4.1 outlining the process followed.

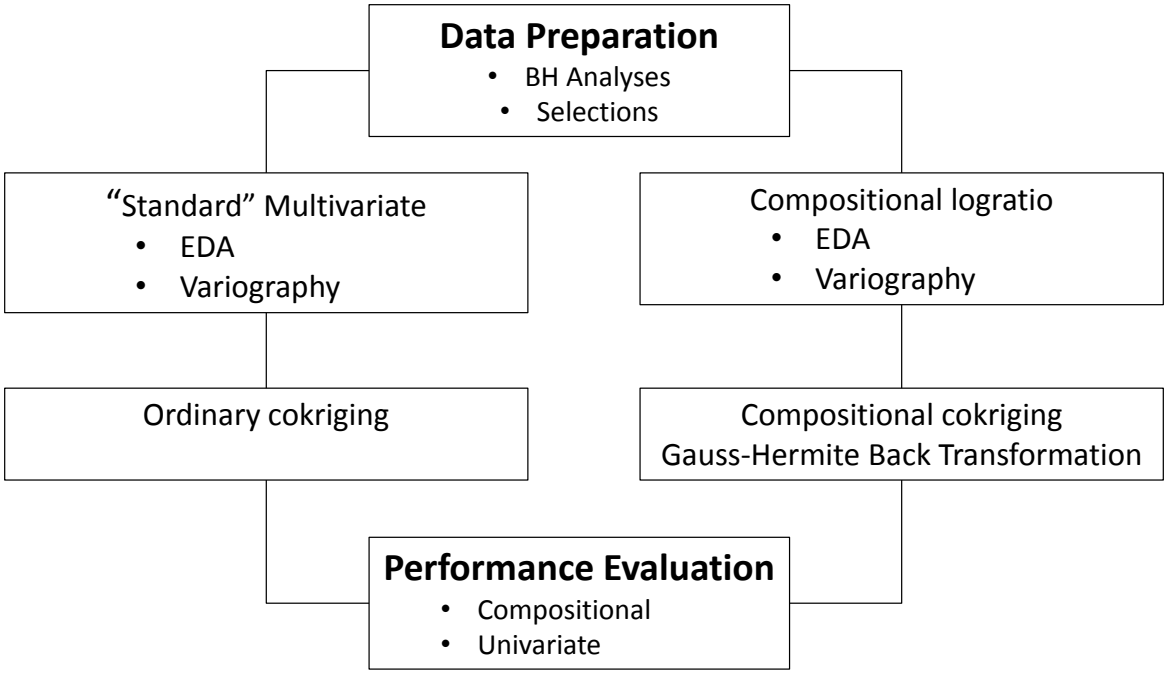


Figure 4.1: Flow chart of major process steps.

As is typical in a geostatistical study the majority of the iterative processes occurred in Isatis followed by output of the variance-covariance matrices from each cokriging calculation for input into the Gauss-Hermite (GH) Quadrature back-transformation calculation. The scripts are available in Appendix H.

4.1 Data spacing

Geostatistical techniques were originally developed to predict mineral grades ahead of the mining face from relatively sparse, highly variable data. As was discussed in Section 3.2, different stages of mining projects have different spatial data density requirements; the more advanced the project, the more dense the data. The guidelines produced by the Joint Ore Reserves Committee (JORC) established a nomenclature used by mining houses and reporting bodies such as the Australian Securities Exchange (ASX) to classify mineral resources based on operational and therefore financial risk [43]. These categories (summarised in Figure 4.2) broadly capture the decreasing risk as a function of increasing geoscientific knowledge. At the scoping level this knowledge is a combination of physical mapping, geophysical mapping and sparse drilling information. As the mining evaluation progresses the geoscientific knowledge generally stems from geostatistical evaluations of drill hole data with increasing spatial density as well as the geophysical and physical mapping information. Other practitioners of compositional geostatistics have asserted that the high spatial density of the blast hole data described here may mask bias induced by any particular estimation technique (*pers.comm.* Job, 2013). To test this assertion the initial high density complete dataset shown in Figure3.5 was sub-sampled to create alternative subsets which mimic the spatial densities ordinarily encountered during the different stages of mine development. Ideally the *real* RC exploration drilling data should be used to evaluate the technique; however the fundamental differences between the short, low spatial density, small volume RC samples and the longer, vertical, high spatial density larger volume blast hole samples outlined in Section 3.2 prohibits this and use of the sub-sampled data to avoid biasing any comparative results is necessary.

Table 4.1: Sampling selection subset spacing parameters

Dataset	X mesh	Y mesh	Number of samples
Spacing 1	-	-	1594
Spacing 2	7	7	432
Spacing 3	14	7	234
Spacing 4	14	14	137
Spacing 5	28	14	71

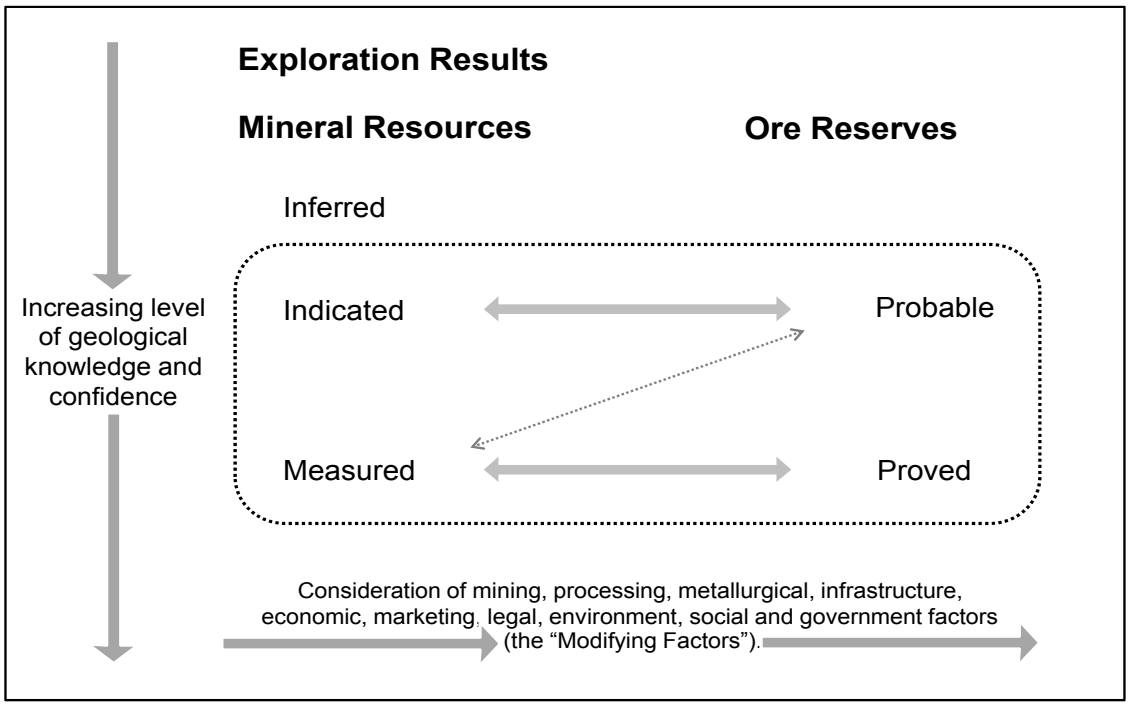


Figure 4.2: JORC(2012) classification diagram

Table 4.2: Description of subsets

Dataset	Mining Stage Description	JORC classification
Spacing 1	ROM - Grade Control - Weekly / daily Mine plan	Measured
Spacing 2	Advanced Grade Control - 90 day Forecast Mine Plan	Indicated / Measured
Spacing 3	Resource level drill out - Life of Mine Plan	Indicated
Spacing 4	Exploration Infill drilling - Conceptual Mine plan	Inferred
Spacing 5	Scoping level - No mine planning yet	Unclassified

East - West division

The data exhibit a strike direction change as a result of the structural imbrication discussed in Section 3.1.3 pivoting about 721165 m Easting. The dominant strike direction in the West block is 80°East of North increasing to 110°East of North in the East block (Figure 4.3). Strike changes of that magnitude result in changes to spatial covariance functions which should be captured and exploited to ensure estimates of surrounding locations is optimal. Due to the complexity of estimating and modelling the spatial covariance functions and the limitations imposed by the available software, the parsimonious decision was taken to create separate estimation domains rather than attempt alternative approaches such as local anisotropy or unfolding.

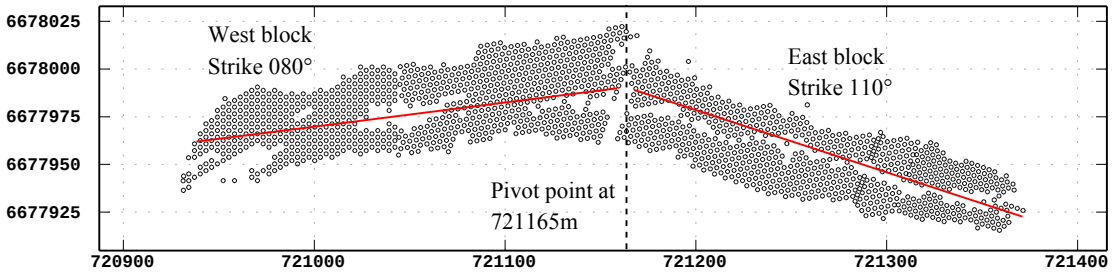


Figure 4.3: Plan view of East and West blocks showing dominant strike directions

4.2 Spatial covariance modelling

4.2.1 Constrained space variography

The auto and cross experimental semi-variograms were calculated for each subset in the constrained sample space for the seven attributes and $filler$ variable. The experimental semi-variograms were calculated using different lag criteria which mimicked the spatial selection parameters used in Section 4.1 (Table 4.3). Each set of experimental semi-variogram was modeled with a Linear Model of Coregionalisation (LMC) using a nugget and two spherical structures incorporating anisotropy. Estimates using semi-variogram models with different types of structure, ranges, and relative contributions of each structure to the overall semi-variogram model can be markedly different if all other parameters are held constant. Although the impact of differing data densities is an important aspect of this study, the effect on the variogram models was designed to be minimised. This was achieved by using the highest density (Spacing 1) data as the base case and applying the information derived from this to the lower density subsets. The nugget variance-covariance values and ranges of each spherical structure were locked and only the sill values for each spherical structure were allowed to vary to match the experimental semi-variogram values from each subset (Table 4.4).

4.2.2 Logratio variography

The same calculation parameters outlined in Table 4.4 used to calculate the logratio experimental semivariogram as were used for the untransformed (constrained space) data. The cross semi-variograms were estimated using the relationship described in Equation 2.33 which required the estimation of experimental auto semi-variograms of all logratio pairs. The Isatis software used could not perform the required cross semi-variogram calculation required by this equation as the experimental values are locked within binary parameter files. Estimating the required quantities necessitated exporting the binary parameter files into an ASCII format, importing these into Microsoft Excel, performing the arithmetic

Table 4.3: Experimental semivariogram calculation parameters

Direction	Parameter	Spacing 1	2	3	4	5
1	Calculation lag	3.5	7	14	14	21
	Tolerance on lag	0.5	0.5	0.5	0.5	0.5
	Number of lags	40	20	12	12	7
	Angular tolerance	30	30	30	30	30
	Azimuth (Limb)	110 °(East) / 080 °(West)				
2	Calculation lag	3	6	6	12	12
	Tolerance of lag	0.5	0.5	0.5	0.5	0.5
	Number of lags	20	10	10	7	7
	Angular tolerance	45	45	45	45	45
	Azimuth (Limb)	200 °(East) / 170 °(West)				

Note: Each spacing has the same anisotropy characteristics, verified from the exhaustive dataset. These were applied to all the lower density subsets.

Table 4.4: Constrained space variogram parameters and ranges

Parameter	East Block	West Block
Number of structures	3	3
Rotation	110 °	080 °
Range Spherical 1 Strike	15	18
Range Spherical 1 Across Strike	5	12
Range Spherical 2 Strike	60	70
Range Spherical 2 Across Strike	15	27

and finally re-importing the ASCII files into Isatis' binary format for later modelling. The same approach used to model the logratio semi-variograms as was employed for the constrained space semi-variograms, i.e. locking the nugget and range values determined when modelling the Spacing 1 data, and applying these to the lower density datasets; and similarly the LMC was used to model the spatial covariance. The complete set of fitted model plots are available in Appendix I and the model parameter tabulations in Appendix J.

4.3 Cokriging

4.3.1 Cokriging parameter optimisation

The kriging neighborhood parameters dictate which samples are used as the informing data for any given kriging system. The parameters control the size, shape and orienta-

tion of the search volume and the number of data within that volume. Changes to these selections can materially change the resulting estimates and must be iteratively changed and the results analysed to ensure the optimal estimates will result. The kriging neigh-

Table 4.5: Optimised kriging neighborhood parameters

Parameter	East Block	West Block
Orientation	110 °	085 °
Rotated X range	100 metres	
Rotated Y range	15 metres	
Minimum number of samples	5	
Maximum number of data	30	
Quadrant search used	No	

borhood parameters in this study were optimised using cross-validation (Section 2.10.1). The search window was set to be an ellipse (Table 4.5) and only the minimum and maximum number of samples used, and the number of quadrants were iteratively altered in the optimisation procedure. The cross-validation function delivers the following numerical statistics:

1. mean error of the true error (Equation 2.66),
2. variance of true error,
3. standardised error (item 1 scaled by item 2), and
4. variance of standardised error.

Additionally graphical representations of the error distributions are produced, including histograms of the error and scatter-plots of the estimated versus true values. The minimum and maximum number of samples used was iterated to achieve a mean error closest to zero and a tight and symmetrical distribution of errors around the mean value. As outlined in Chapter 1, the aims of the study are to compare the outputs of additive logratio techniques with the outputs of conventional estimation techniques incorporating data density changes as well. The effects of differing search parameters are well known ([10], [40]) and not part of this study, and as a result the neighborhood parameters were optimised using the Simplex variogram and Spacing 1 dataset and those parameters were applied to the other estimates. It is is acknowledged that this may result in suboptimal estimates with other variogram models and datasets, however the author deemed the risk acceptable as the impact is likely negligible.

4.3.2 Interpolation

The cokriging step was undertaken to provide estimates at each sample location under the following scenarios:

1. complete cross-validation of each sample location for the simplex sample space,
2. jack-knifing the non-overlapping data at each data density for the simplex sample space,
3. complete cross-validation of each sample location for the log transformed sample space, and
4. jack-knifing the non-overlapping data at each data density for the log transformed sample space.

The estimation runs were completed using the semi-variogram models described in Sections 4.2.1 and 4.2.2 and the optimised kriging neighborhood parameters outlined in Section 4.3.1. The simplex cross-validation and jack-knifing runs were completed using the traditional interface of the Isatis software package and associated scripting functions. The logratio sample space estimation runs were completed using the ”print complete output” switch checked in order to provide the complete kriging systems for each estimated point. Checking this option generates two files, the traditional spatially located estimate file and the kriging systems output file. The kriging systems output file had to be capped at 100 samples otherwise the file size (in excess of one gigabyte) was too large to be opened or interrogated using commonly available software and text editors. As a result runs of 100 samples were used and the results later recombined.

4.4 Back-transformation and error measures

4.4.1 Gauss-Hermite Quadrature Back-transformation

After the logratio cokriging step was completed the process described in Section 2.9 to back-transform the interpolated results from the real space to the simplex was carried out. The quadrature was performed using the MATLAB software and required the export of an estimate and the variance-covariance matrices from Isatis. The MATLAB procedure (Appendix H) generates the approximation of the interpolated value and the estimation variance associated with the estimate. These back-transformed values were matched to the true sample values to create datasets for comparison and error measurements.

4.4.2 Spatial recombination and error measures

Prior to the estimate runs the data were split into a number of independent zones to facilitate the modelling and downstream processing; before the error variables could be generated the data were recombined to create the final datasets. The end result of this process was five datasets, Spacing 1 through to Spacing 5, each containing the original

assayed analyte value as well as an estimate using OCK and the GH methodologies. In addition to these, Spacing 2 through to Spacing 5 also included an attribute identifying whether the data location is an informing point or an estimated point for the jack-knifing process. Subsequent to the data being recombined, the error measures outlined in Sections 2.10 and 2.12 were calculated for each dataset using the Isatis *Calculator* function.

Chapter 5

Characteristics of the Input Data

5.1 Exploratory Data Analysis

The characteristics of the exhaustive and each subset are described in Sections 5.1.1 and 5.1.2 below; any differences between a subset and the exhaustive set must be quantified as this could have a material impact on downstream uses of the data and comparisons with the exhaustive set. The data are examined in the raw units in the simplex; although this is known to be incomplete and in some cases spurious, the customers for the iron ore expect weight percent analyses of the data. Additionally the data are explored using the *stay* and the *move* methodologies described in Section 2.2 and Section 2.5

5.1.1 Constrained space EDA — Exhaustive

The Exhaustive dataset, *Spacing 1* shows characteristics typical of Direct Shipping Ore (DSO). These ores have left (negative) skewed distributions for *Fe* (and by construction the filler), strongly right (positive) skewed for Al_2O_3, *Mn*, SiO_2, and *S*, and less strongly right skewed for *LOI* and *P* (Figure 5.2). The descriptive statistics (Table 5.1) emphasise the high grade and relatively low contaminant levels of the ore. There are compositional relationships in this type of ore which are related to the initial precipitation of the BIF and the stoichiometry possible from that initial composition. Simply this means that as the *Fe* content increases, the SiO_2 must decrease along with Al_2O_3, and also that the maximum *Fe* content cannot exceed 72.35 weight % (pure magnetite). This *compositional bounding* is clearly noted in the Figure 5.3 as a linearity in the *Fe* vs. SiO_2 and *Fe* vs. Al_2O_3 plots beyond which no points plot. These compositional parameters constrain the possible point locations to an even more restricted sample space than the simplex theoretically allows (0 - 100). The compositional and stoichiometric constraints impose correlations between the analytes, illustrated in Figure 5.3 and quantified in the lower triangle of Table 5.1.

The correlations range between 0.95 and -0.909; with the majority (22 out of 28) between -0.299 and 0.306 which is considered low. The grouping of the data points in the scatter plots (Figure 5.3) is tight with no obvious groups emerging.

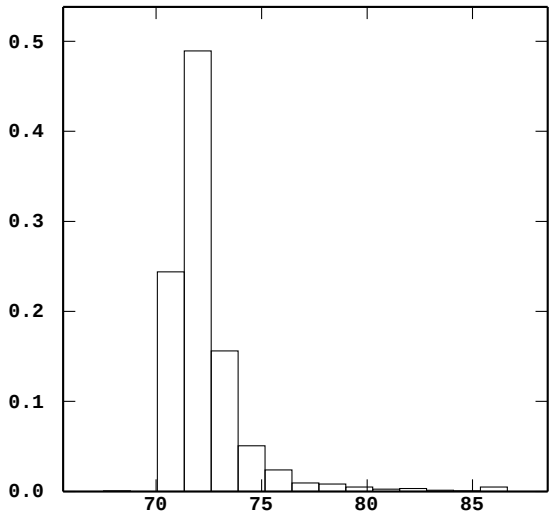


Figure 5.1: Sum of analytes — Exhaustive data

Table 5.1: Descriptive statistics of Spacing 1 dataset (n = 1594)

VARIABLE	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2	$Filler$
Min	0.080	9.940	0.520	0.003	0.016	0.001	0.340	13.368
Max	21.410	71.740	10.940	12.918	0.417	0.975	61.580	32.472
Mean	1.651	62.926	4.078	0.188	0.128	0.030	3.382	27.618
Std. Dev.	1.981	5.777	1.409	0.546	0.051	0.054	5.918	1.930
CoV	1.201	0.092	0.345	2.907	0.398	1.843	1.75	0.070
Skewness	3.559	-5.268	0.624	12.615	0.939	8.916	6.446	-3.613
Correlation	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2	
Fe	-0.819							
LOI	0.273	-0.258						
Mn	0.028	-0.099	0.177					
P	-0.107	0.147	0.298	0.072				
S	0.306	-0.181	0.273	-0.003	0.091			
SiO_2	0.658	-0.941	-0.002	-0.037	-0.249	0.062		
$Filler$	-0.806	0.950	-0.299	-0.031	0.166	-0.192	-0.909	

The mean and variance are plotted as a function of the easting coordinate (Figures 5.5 and 5.6). As discussed in Section2.6 an assessment of the absence of drift in the first two moments is necessary in order to make the decision of stationarity. When analysing the mean statistic in Figures 5.5 it is noted that the *Mn* analyte displays the greatest trend with respect to Easting; the segregation between the East and West blocks is also marked. Five of the other seven analytes also display marked differences in mean value between the East and West blocks, only the *LOI* and *P* do not. This characteristic of these two analytes

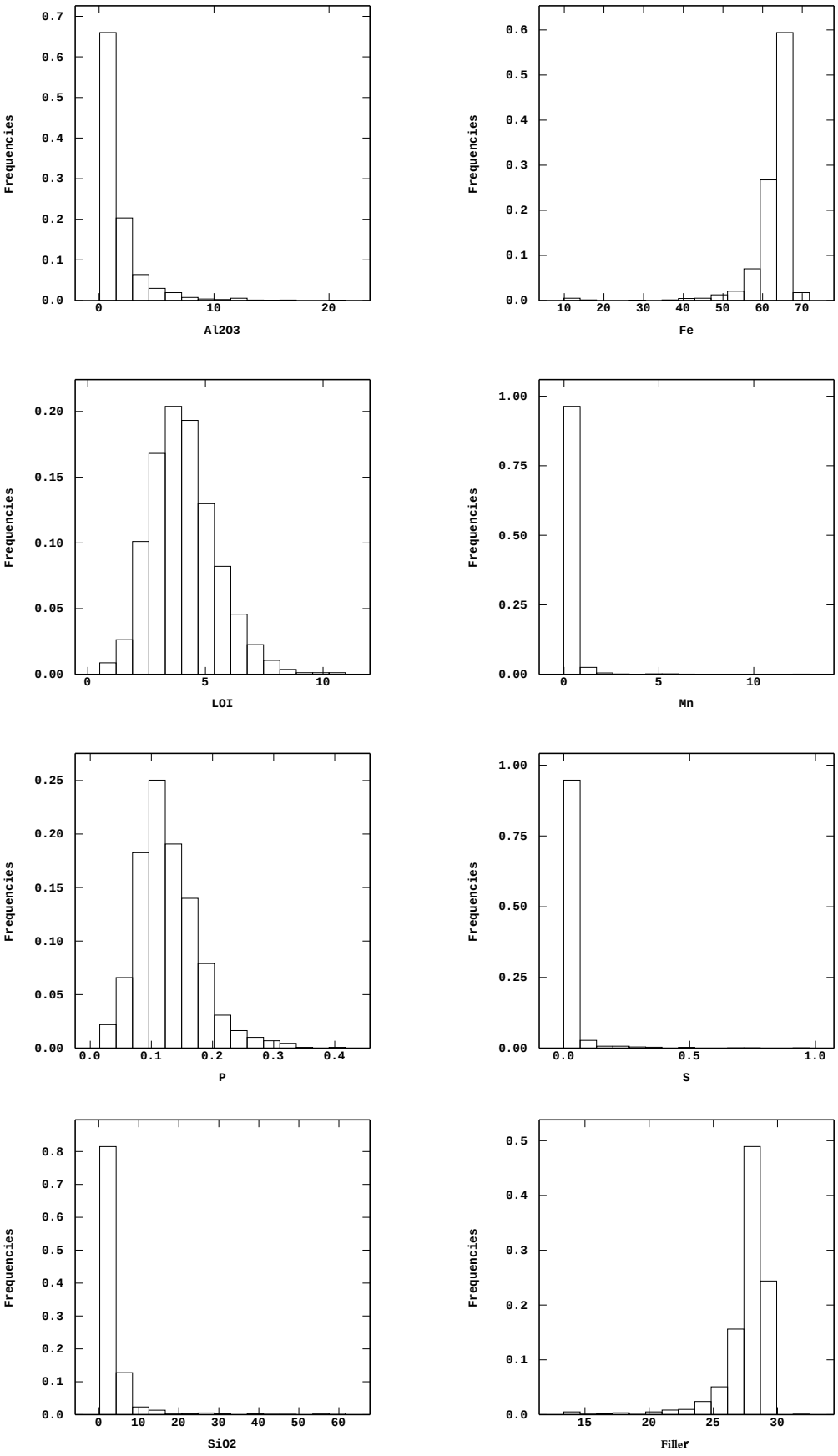


Figure 5.2: Histograms of Spacing 1 dataset

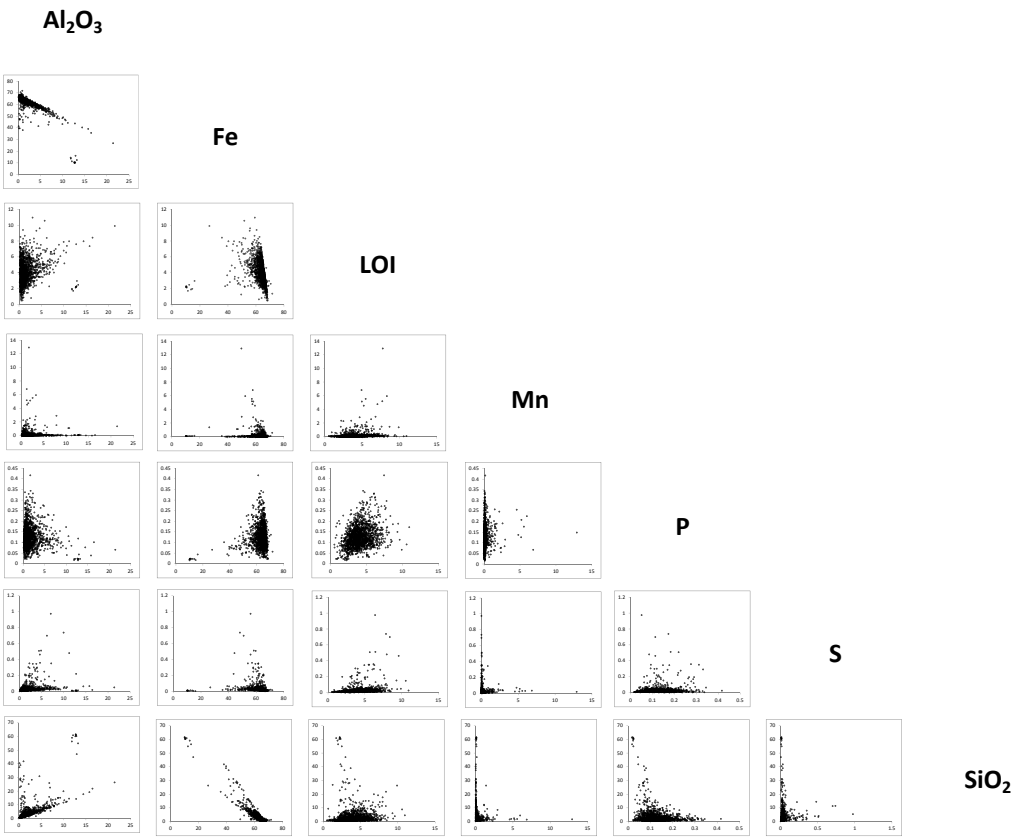


Figure 5.3: Scatterplots of Spacing 1 dataset

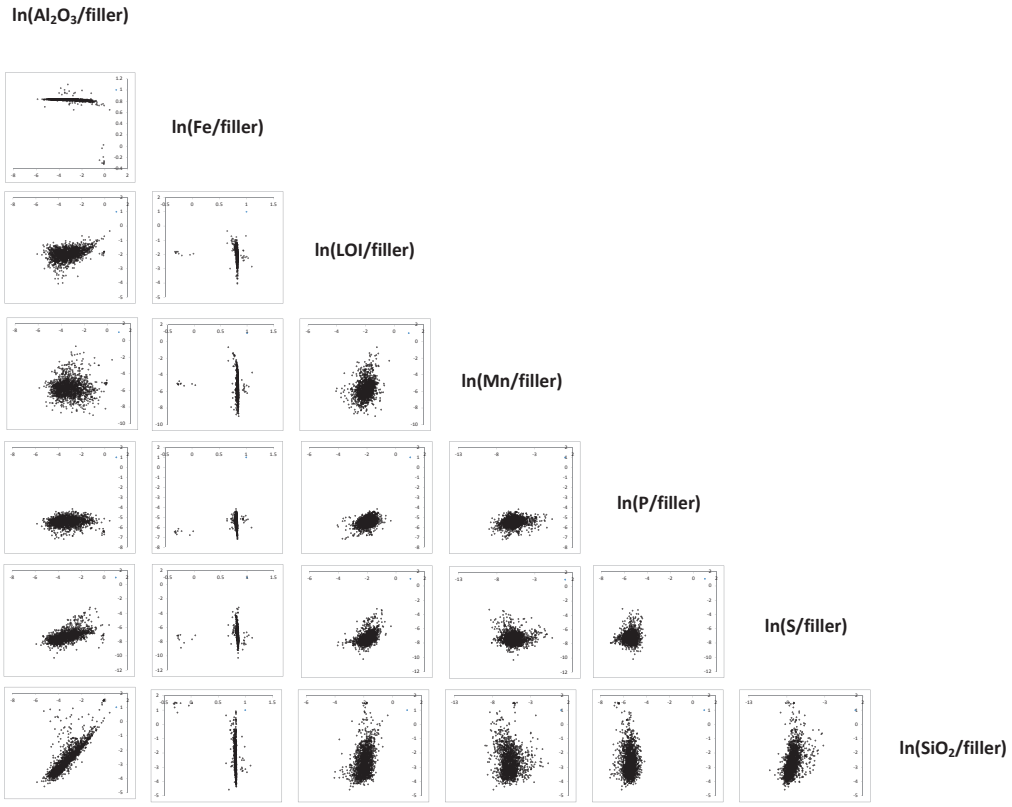


Figure 5.4: Scatterplots of Spacing 1 dataset after logratio transformation

relates to the geological processes; the distributions are not highly controlled by the BIF or upgrading mechanisms discussed in Section 3.1.4, rather it is a function of the vertical regolith processes. Regolith characteristics are the end product of the effect of climate on the earths surface and has a vertical effect, and therefore ubiquitously applied to the East - West aspect of the study area. The East and West blocks are visually discernible when analysing the variance of the 20 meter wide swathes in Figure 5.6. The *LOI* is the only analyte displaying almost constant variability as a function of the Easting.

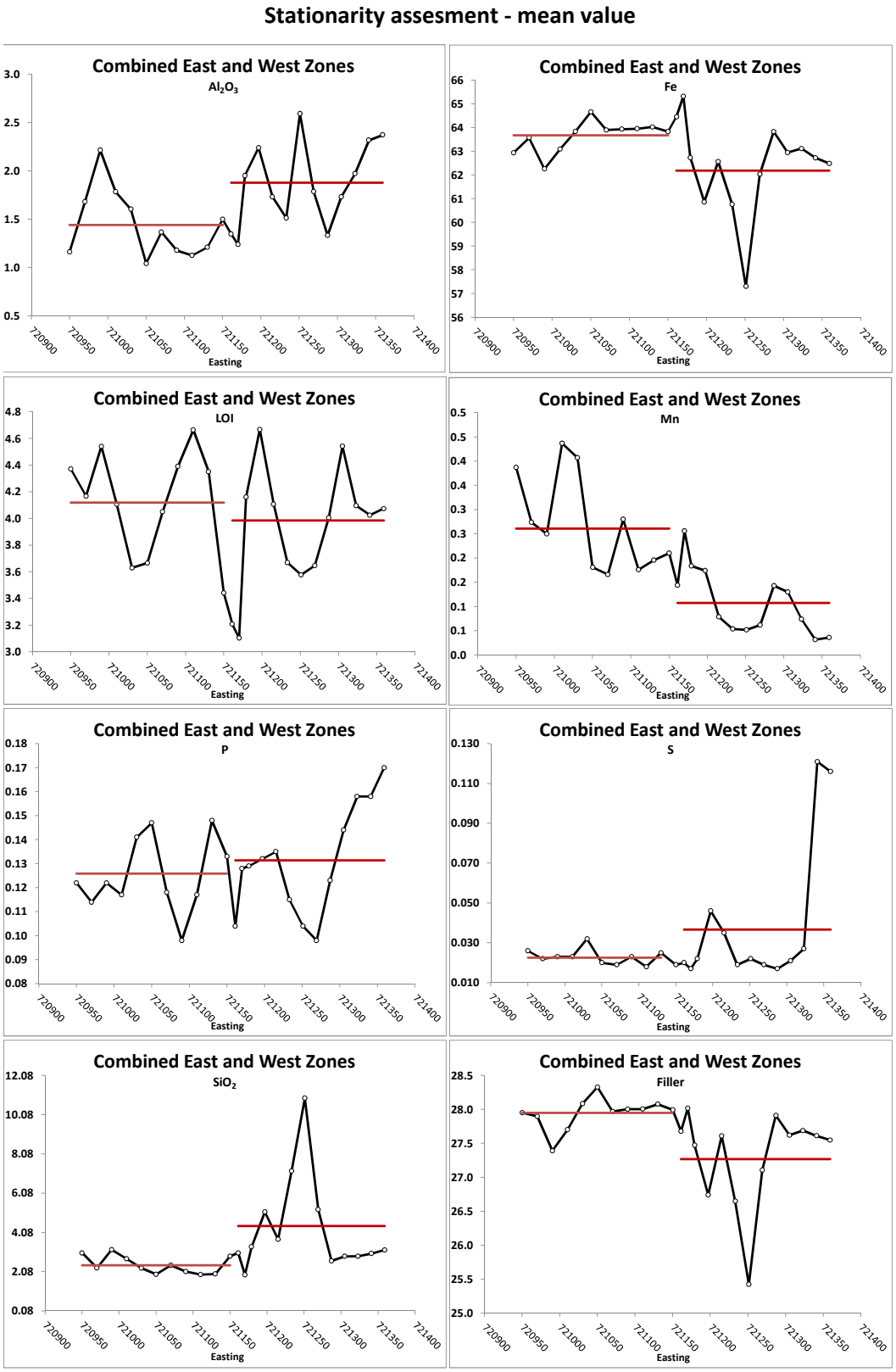


Figure 5.5: Stationarity assessment plot - mean of data within 20 m wide swathes. Red lines represent the mean and the break in the red line represents the Easting (721165m) about which the estimation domains are divided.

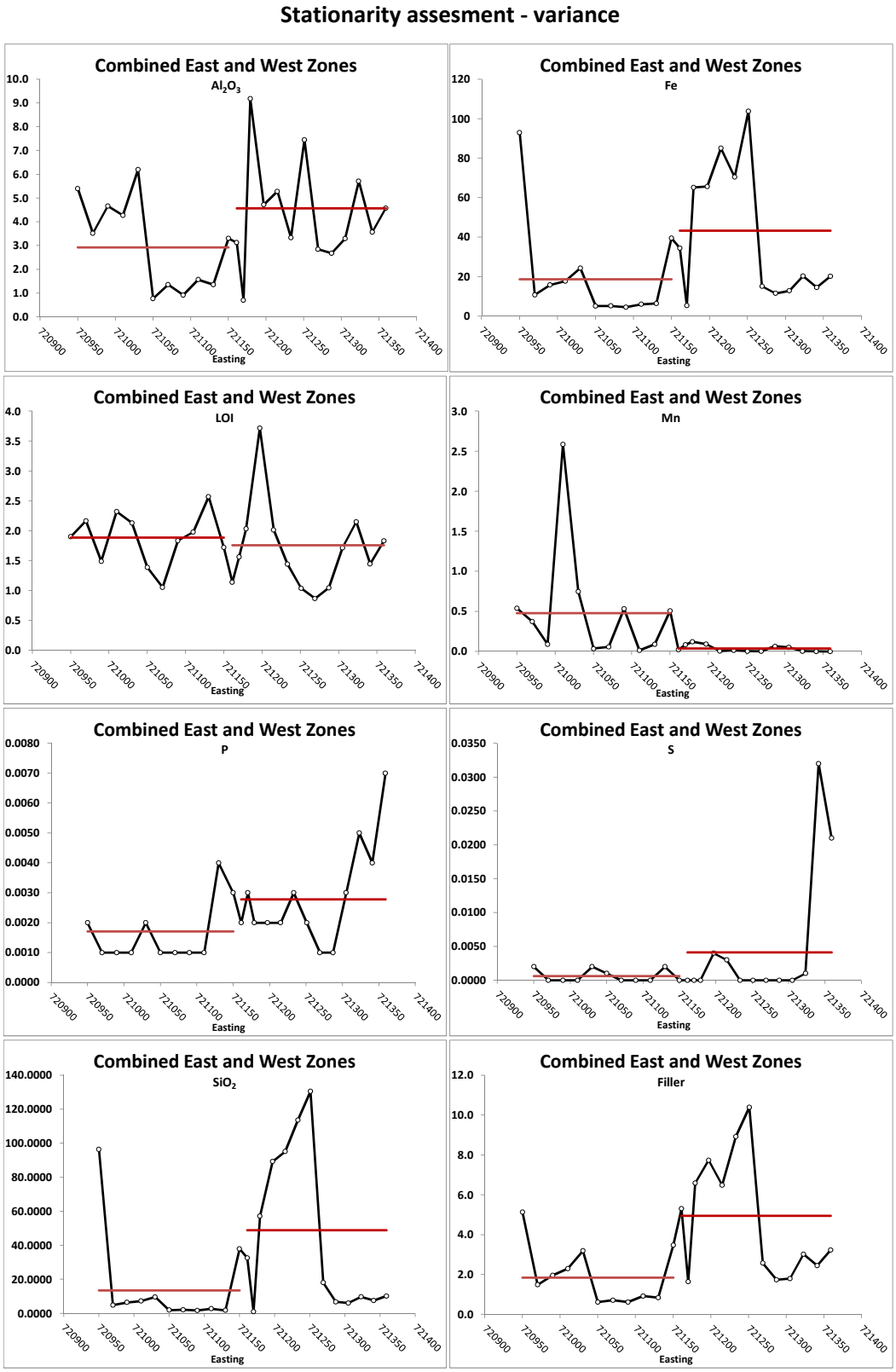


Figure 5.6: Stationarity assessment plot - variance of data within 20 m wide swathes. Red lines represent the mean and the break in the red line represents the Easting (721165m) about which the estimation domains are divided.

5.1.2 Constrained space EDA — Subsets

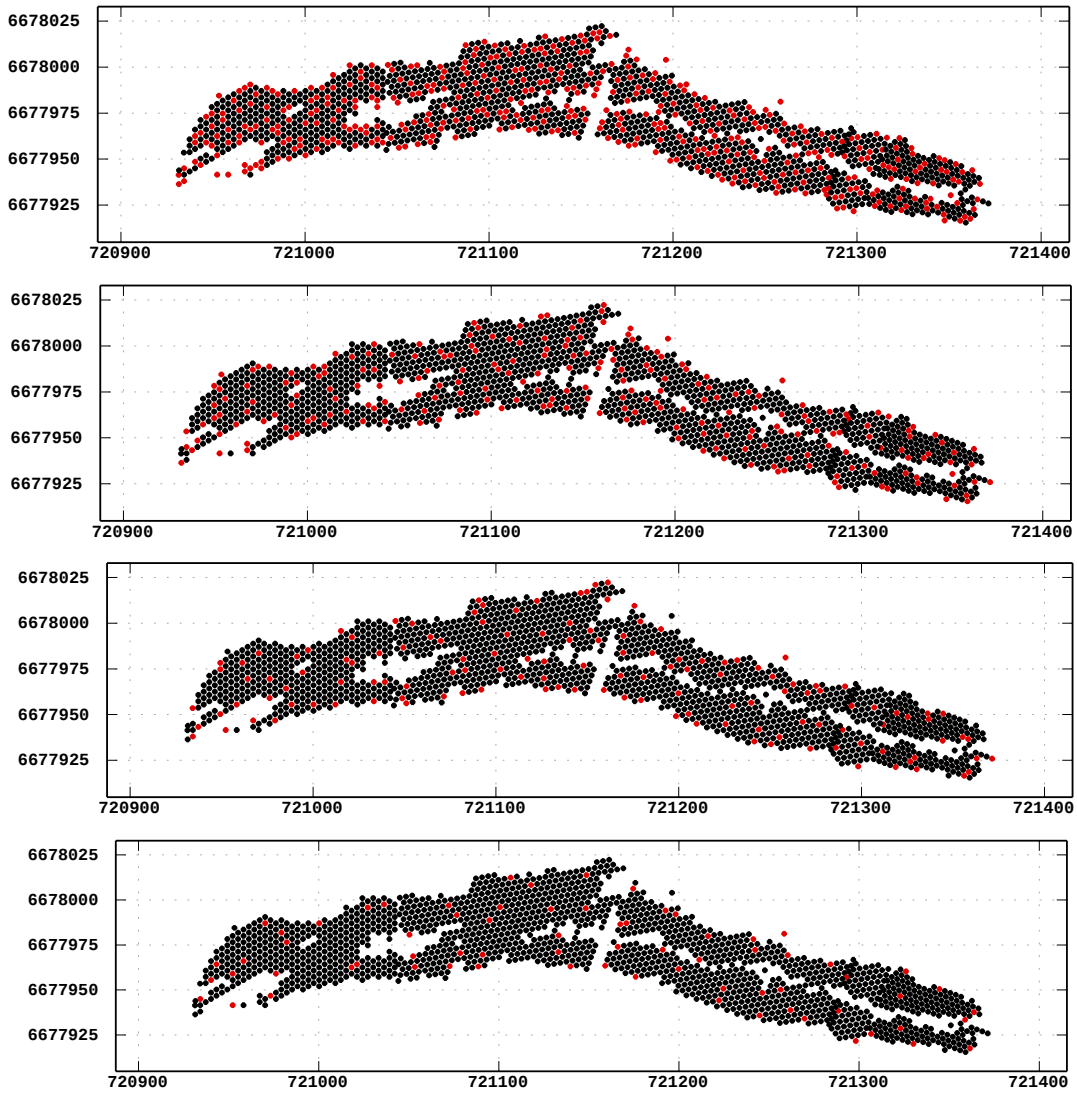


Figure 5.7: Informing locations, red points are informing data, black points are the estimation locations — Plan view

The subsets are intended to be representative of the entire bench. In order to evaluate the representativity, two tools were used; the mean percentage bias statistic and QQ plots. The mean percentage bias is calculated as

$$\text{Bias}_{element} = \frac{mean_{element\ \ Spacing\ \ 1}}{mean_{element\ \ Spacing\ \ i}} \times 100 \quad for\ \ i = 2,\ldots,5 \ \text{ and all elements.} \qquad (5.1)$$

The tabulated bias statistics in Table 5.2 contain SiO_2, Al_2O_3, and Mn values of up to 30%. These values at first appear too high for the subset to be considered representative. However, evaluation of the histograms (Appendix K) and specifically the QQ plots in Figure 5.8 shows that the biases are brought about by the presence or absence of several extreme values in the tails of the highly skewed distributions when compared with the exhaustive set. The skewness statistics in Table 5.1 are highest for these biased elements,

Table 5.2: Mean percentage bias between Spacing 1 and subsample datasets

VARIABLE	Spacing 2	Spacing 3	Spacing 4	Spacing 5
Al_2O_3	-8.42	-20.71	-31.25	-20.84
Fe	0.68	1.36	1.9	1.62
LOI	1.28	0.2	-2.23	2.48
Mn	0.53	-4.26	15.43	15.43
P	0	0	-2.34	1.56
S	-3.33	10	-10	6.67
SiO_2	-13.63	-21.58	-27.32	-30.19
$filler$	0.43	0.78	1.13	0.77

supporting this assertion. The QQ plots demonstrate that where the highest density of data exists, the distributions plot along the 45° line, which is the required outcome. *Fe*, *LOI*, and *P* have lower skewness statistics and have associated lower mean biases. Given that the subsets are sub-sampled from the master set, combined with the link between the skewness and bias, and excellent appearance of the QQ plots over the critical portion of the grade ranges, the subsets are deemed representative of the master set and therefore acceptable for use in this study. The comprehensive listing of all the subsets base statistics is available in Appendix A.

5.1.3 Spatial covariance modelling

The alr transformation of skewed, (spuriously) correlated data can advantageously result in partially or completely decorrelated data with a deskewed or even normal distribution. In fact the alr transformation has been used in a manner analogous to Gaussian anamorphosis in simulation studies to achieve the removal of spurious correlation and normalise the distribution [23]. As a result of the the pseudo-normal distributions and decreased correlations between variables (Figures 5.3 and 5.4), the transformed distributions are more tractable to model than the untransformed data. The transformed data (top of Figure 5.9) were modelled using the relationships and techniques outlined in Section 2.8.2 and Section 4.2.2 respectively, while the untransformed data (bottom of Figure 5.9) are estimated using using conventional semivariograms. Both resulting sets of experimental variograms were modelled using the Linear Model of Coregionalisation. The high spatial density of the data results in unambiguous experimental semivariograms to model, a boon in a univariate case, but this can be problematic when fitting due to the constraints imposed by the LMC. The Spacing 1 examples illustrated in Figure 5.9 show both these attributes; the shape of the semivariograms are very clear, but the fitment of some of the cross semivariograms is not ideal. Qualitatively, the fitment of the logratio semivariogram set appears superior to that of the raw space data. The remainder of the experimental semivariograms

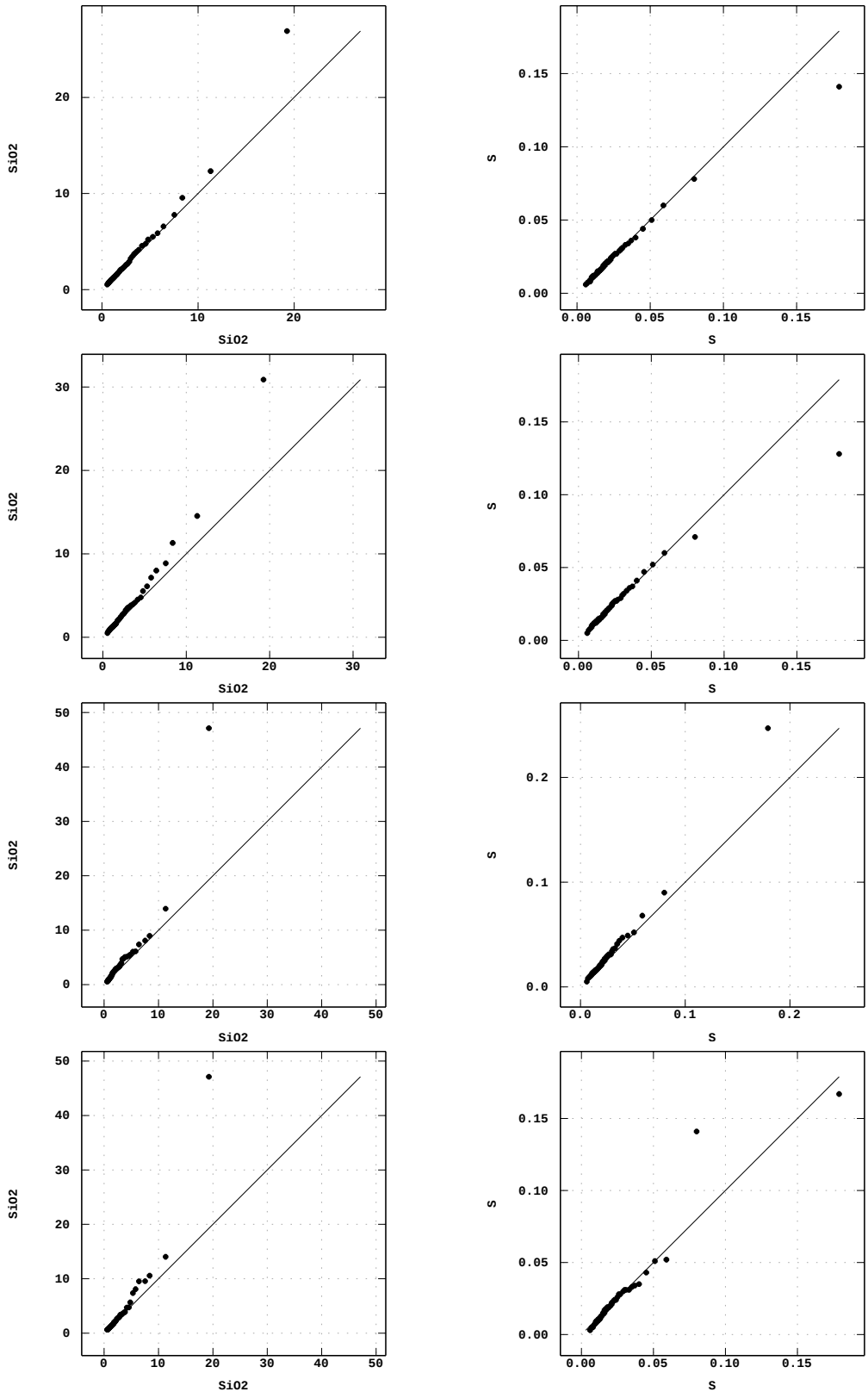


Figure 5.8: Selected comparative QQ plots of Spacing 1 (ordinate) and subsets (abscissa) in ascending order of subset spacing (2 to 5) SiO_2 on left S on right

and fitted models are available for perusal in Appendix I and tabulated in Appendix J. All the spacings but the lowest (Spacing 5) exhibit model-able structure, as both logratio and raw space are reasonably fitted by the LMC. The decision to use the Spacing 1 nugget and

ranges is coherent with current mining geostatistics practices, where use of *a priori* spatial covariance information from geologically similar or nearby deposits is commonplace. In the spacing 5 example specifically, the lack of structure demands that some subjectivity would ordinarily be required to fit an appropriate set of models in the absence of *a priori* information.

5.1.4 Compositional EDA

The implementation of the logratio approach to geostatistics in this thesis predominantly follows the work of Pawlowsky-Glahn and Olea [7]. This approach proposes that the alr is a good choice for the analysis and subsequent estimation of spatially correlated compositional data for practical reasons. These reasons include mathematical tractability and readily available software to perform the analysis and estimation as outlined in Sections 2.5 and 2.7. The exploratory data analysis for compositions generally follows a common approach independent of whether the variables are spatially correlated and considers all the possible log pairs. This effectively covers all possible combinations which could be used in the calculation of the alr. This five step approach is summarised in Pawlowsky-Glahn *et al* [8] and then applied to the W3 data. The steps are as follows:

1. compute compositional descriptive statistics,
2. centre the data set for a better visualisation of subcompositions in ternary diagrams;
3. analyse at the biplot of the data;
4. define an appropriate representation in orthonormal coordinates and computing the corresponding coordinates; and
5. represent the results in a balance-dendrogram.

The centre of the Spacing 1 dataset $\mathbf{g}_{\mathbf{Spacing1}}$ for the analytes in the following sequence is

$$\begin{aligned} \mathbf{g}_{\mathbf{Spacing1}} &= [Al_2O_3, Fe, LOI, Mn, P, S, SiO_2, filler] \\ \mathbf{g}_{\mathbf{Spacing1}} &= [0.011, 0.643, 0.039, 0.001, 0.0012, 0.0002, 0.021, 0.284]. \end{aligned} \tag{5.2}$$

The Fe, Al_2O_3, and SiO_2 ternary diagram is shown in Figure 5.10 where is is clear the data are naturally compressed into the Fe corner of the diagram. To facilitate interpretation, data are centred through perturbing the values with the inverse $\mathbf{g}^{-1}{}_{Spacing1}$ of the centre values. The two diagrams show that there is only one material type (ore) predominately sampled in the dataset with some contaminated edge effect samples. The contamination edge effect samples form the two separate groups of low iron material in the left hand diagram, while the mineralised BIF forms the linear grouping perpendicular to the Al_2O_3 and SiO_2 axis. The geochemical bounding inherent to BIF mineralogies is evident in the right hand side diagram of Figure 5.10 as an upper linear boundary highlighted in red and subdividing the centred ternary diagram into area A and area B. The boundary

Table 5.3: Spacing 1 Variation matrix Variance (upper triangle), Mean (lower triangle)

$\ln(X_i/X_j)$	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2	$Filler$
Al_2O_3		0.959	0.778	1.942	1.014	0.662	0.31	0.898
Fe	4.074		0.171	1.072	0.163	0.568	0.952	0.007
LOI	1.282	-2.792		0.979	0.209	0.441	0.789	0.158
Mn	-2.473	-6.547	-3.755		1.049	1.715	2.153	1.047
P	-2.202	-6.275	-3.484	0.272		0.657	1.11	0.167
S	-3.975	-8.049	-5.257	-1.502	-1.774		0.869	0.554
SiO_2	0.652	-3.422	-0.63	3.126	2.854	4.628		0.883
$filler$	3.255	-0.819	1.973	5.728	5.457	7.231	2.603	

Table 5.4: Sign matrix Spacing 1 dataset

Al_2O_3	Fe	LOI	Mn	P	S	SiO_2	$filler$	r	s
-1	-1	-1	1	-1	-1	-1	-1	1	7
-1	-1	-1	0	-1	1	-1	-1	1	6
1	-1	-1	0	-1	0	-1	-1	1	5
0	-1	-1	0	-1	0	1	-1	1	4
0	-1	1	0	-1	0	0	-1	1	3
0	-1	0	0	1	0	0	-1	1	2
0	1	0	0	0	0	0	-1	1	1

therefore represents a limit beyond which it is unlikely to find a BIF sample, a structure that estimation methods should honour. The variation matrix shown in Table 5.3 is a compact form of the base statistics where the mean values for each pairing are below the diagonal and the variance values above the diagonal. The variation matrix explains how the total variation is split between the logratio parts. The *total variance* (Equation 2.12) for the *Spacing 1* data equals 2.784. These three analyte combinations of $ln(SiO_2/Mn)$, $ln(S/Mn)$, and $ln(Al_2O_3/Mn)$ account for 26% of the total variance. Examination of the scatterplots of logratio transformed quantities (5.4), although not part of the standard compositional EDA work-flow outlined by Pawlowsky-Glahn and Olea [7], shows that the only highly correlated variable set after alr transformation is Al_2O_3 and SIO_2 (with the common denominator of the $filler$). The biplots (Figure 2.2) are shown in two projections to facilitate interpretation. Using the guidelines outlined in Section 2.4 the following features are noted:

- near co-linearity between Fe, P, and r and to a lesser degree LOI,
- the mutually perpendicular orientation of Mn, Al_2O_3, S, and SIO_2 to one another, and the near collinear group identified above.

Table 5.5: Orthonormal basis Ψ Spacing 1 dataset

Al_2O_3	Fe	LOI	Mn	P	S	SiO_2	$filler$
-0.13	-0.13	-0.13	0.94	-0.13	-0.13	-0.13	-0.13
-0.15	-0.15	-0.15	0	-0.15	0.93	-0.15	-0.15
0.91	-0.18	-0.18	0	-0.18	0	-0.18	-0.18
0	-0.22	-0.22	0	-0.22	0	0.89	-0.22
0	-0.29	0.87	0	-0.29	0	0	-0.29
0	-0.41	0	0	0.82	0	0	-0.41
0	0.71	0	0	0	0	0	-0.71

An orthonormal basis was generated through the sequential binary partitioning as in Section 2.5. The sign matrix is shown in Table 5.4 where it is evident that the simplest solution was selected. This solution involved splitting into groups containing only one part each in a successive manner such that the first decision was to split the most different part (Mn) from the rest. The next part selected and split was S as the biplot identified this as different from the other parts and from Mn. This process continued until no further splitting was possible. The result is the decrease of the s statistic (of the first subgroup) from 7 to 1 with increasing order; coupled with the r statistic of the second subgroup remaining constant and equaling 1. The resulting orthonormal basis Ψ is shown in Table 5.5. Once the ilr transform had been applied the balance dendrogram (Figure 5.12) is produced which summarises the salient statistical measures of the isometric coordinates of the Spacing 1 dataset as outlined in Section 2.5.3. The diagram illustrates that the coordinate associated with Mn has the highest variance as well as the largest range; S, Al_2O_3 and SiO_2 have similar large variances and ranges when compared with the remaining four parts. The variation matrix shows that the logratios with Mn in the denominator account for a large portion of the total variation contained within the data. The biplot supports this conclusion as the links connecting the Mn, S, Al_2O_3 and SiO_2 vertices to any other part are long and the rays are approximately perpendicular to the others. The ternary and centred ternary diagrams do not contain discrete groups of observations; rather the data form a single ellipse almost perpendicular to the Al_2O_3 - SiO_2 continuum. These interpretations support the conclusion garnered from the scatter plots in Section 5.1.1 which suggest a stationary domain.

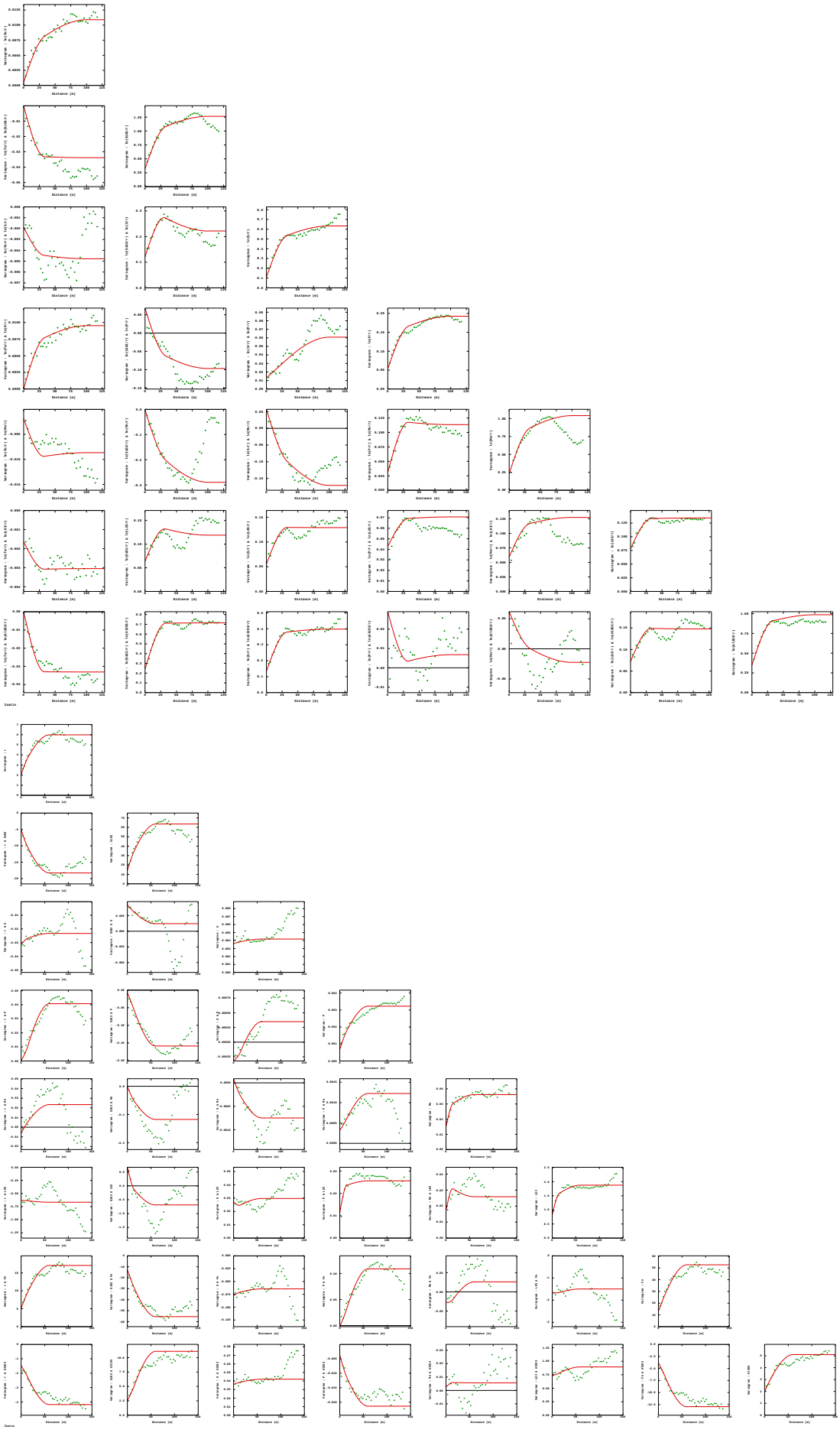


Figure 5.9: Logratio LMC East zone 1 spacing (top), LMC of raw data (bottom)

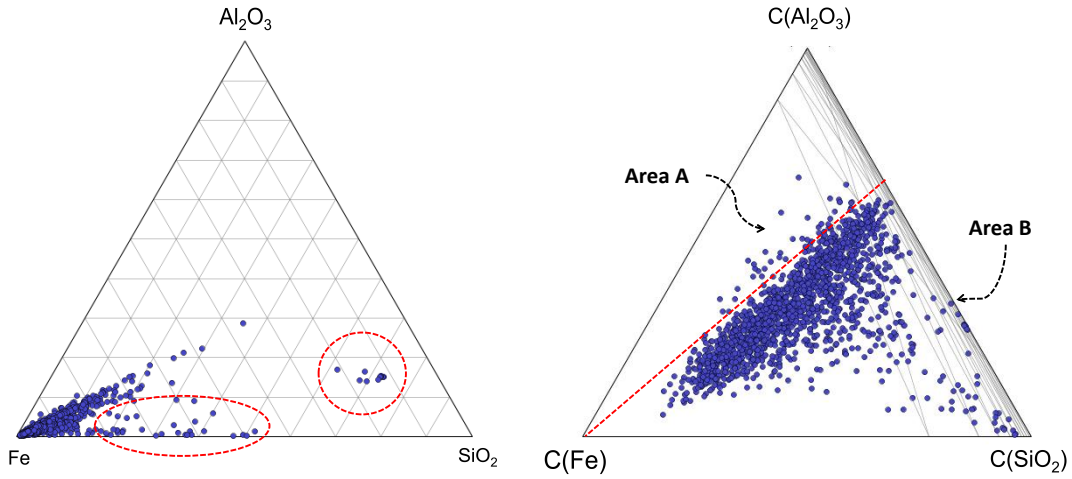


Figure 5.10: Spacing 1 Fe, SiO_2, and Al_2O_3 ternary diagram (left) and centred (right)

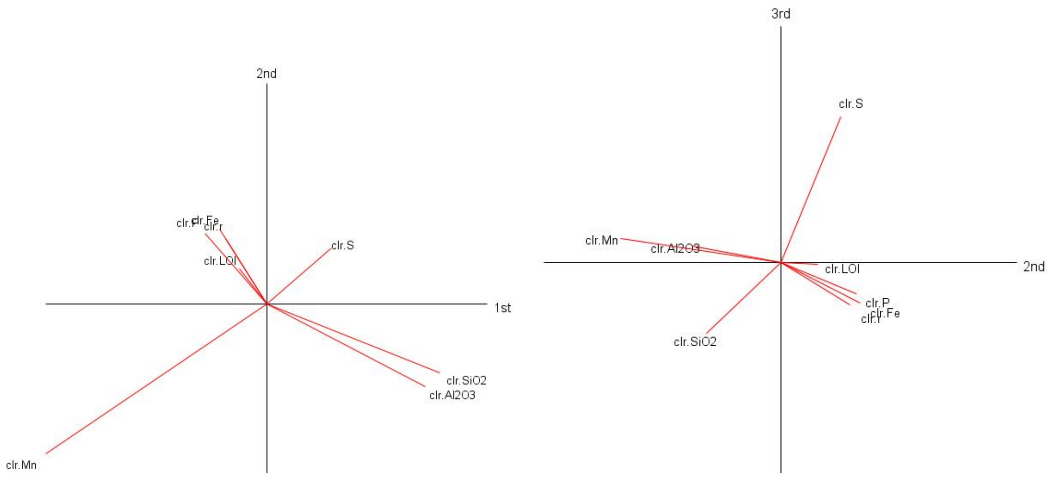


Figure 5.11: Biplot Spacing 1 XY projection (left) and YZ (right)

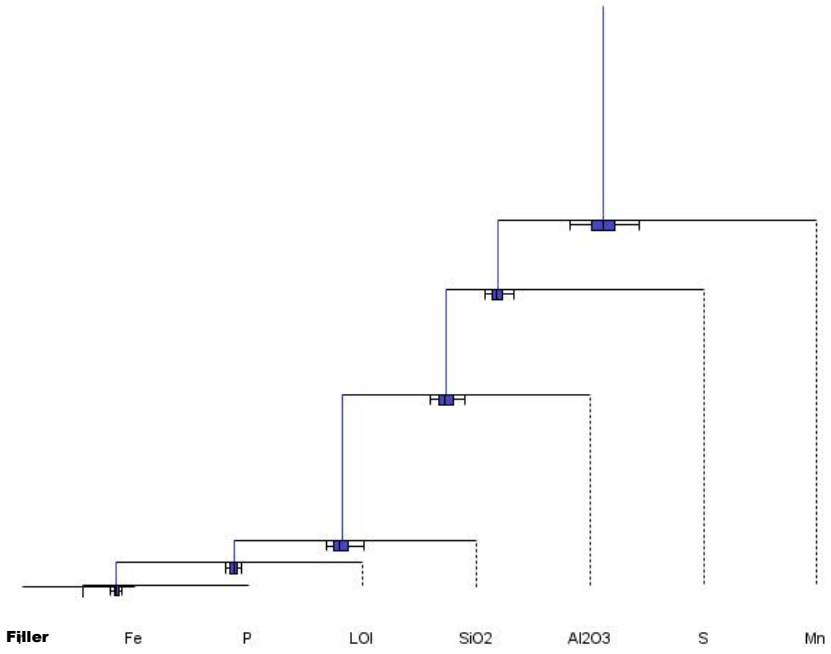


Figure 5.12: Balance-dendrogram

Chapter 6

Estimation Results

Given the eight variables, the three distributions (assayed results, OCK, and Gauss-Hermite), and five spatial configurations used in this study, it is not feasible to describe each result individually. Two principal analytes were selected for in-depth analysis and one correlation set of geologically correlated variables:

1. Fe as it is the dominant part in the total composition, the principal analyte of interest economically,
2. Mn as it is one of the least dominant parts and was identified as the major contributor to the compositional variability in the logratios, and
3. the SiO_2 versus Fe scatter-plot as it is the pairing with the highest linear correlation and described a critical geological compositional relationship.

In general only the Exhaustive dataset (Spacing 1) and the sparsest dataset (Spacing 5) are described in detail throughout this section; this decision was taken for brevity and as these two datasets represent the end-members of the data density scenarios. All the other results are available in Appendix A.

6.1 Estimated distributions

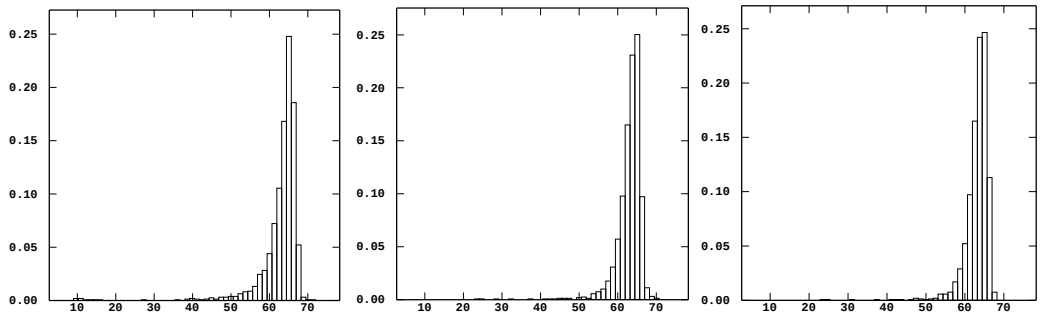
6.1.1 Cross validation estimates (Spacing 1)

The Spacing 1 data were cross-validated resulting in 3 distributions for comparison, those of the sample, OCK and the Gauss-Hermite back-transformed datasets. The basic statistics for each of the true values and estimates per analyte are shown in Table 6.1. The mean statistic of each analyte for the exhaustive dataset are very similar to the mean estimates resulting from the OCK estimates. The mean estimates of each analyte from the Gauss-Hermite Quadrature back-transformed logratio estimates, hereinafter referred to as ”GH”,

Table 6.1: Spacing 1 Cross-validation statistics (n = 1594)

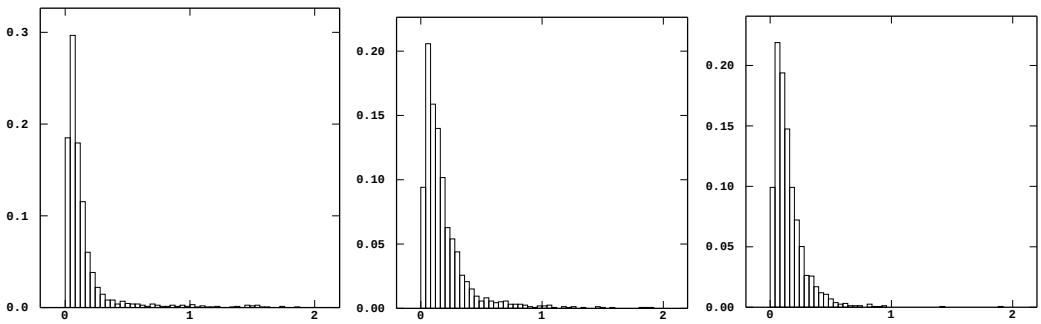
Attribute	Min.	Mean	Std. Dev.	Skew.
Al_2O_3	0.080	1.651	1.981	3.559
Al_2O_3 - OCK	-0.167	1.620	1.156	3.035
Al_2O_3 - GH	0.250	1.583	1.143	3.696
Fe	9.940	62.926	5.777	-5.268
Fe - OCK	23.036	63.007	3.495	-4.330
Fe - GH	24.069	63.150	3.254	-4.442
LOI	0.520	4.078	1.409	0.624
LOI - OCK	1.483	4.078	0.902	0.286
LOI - GH	1.264	4.076	0.934	0.282
Mn	0.003	0.188	0.546	12.615
Mn - OCK	-0.033	0.186	0.203	3.410
Mn - GH	0.006	0.156	0.154	5.435
P	0.016	0.128	0.051	0.939
P - OCK	0.010	0.128	0.039	0.973
P - GH	0.030	0.129	0.039	0.961
S	0.001	0.030	0.054	8.900
S - OCK	-0.003	0.030	0.031	3.668
S - GH	0.001	0.028	0.029	5.489
SiO_2	0.340	3.382	5.918	6.446
SiO_2 - OCK	-2.007	3.310	3.726	4.586
SiO_2 - GH	0.614	3.093	2.946	4.065
$filler$	13.368	27.618	1.930	-3.613
$filler$ - OCK	16.138	27.642	1.197	-3.131
$filler$ - GH	21.519	27.785	0.979	-1.267

are also very similar to the true values. In all cases the OCK mean is closer to the true value than the GH mean. The difference between the *Fe* true value and the GH estimate is 0.224 percent while the true value to OCK difference is 0.081 percent. Similarly the *Mn* differences are 0.032 percent and 0.002 percent for GH to OCK respectively. Both of the interpolated datasets exhibit the smoothing effect of cokriging, evidenced by the standard deviations of the estimates being lower than the true values and as is the skewness. In the majority of cases (5 out of 8 analytes) the skewness of the GH estimates is closer to the true value than that of OCK. Graphically, the true distribution (Figure 5.2) is more peaked than the estimated distributions (Figures 6.1 and 6.2) but the effect is muted and best examined using the selected QQ plots for *Fe* and *Mn* in Figures 6.3 and 6.4 respectively. The complete suite of histograms are shown in Appendix K, the OCK QQ plots in Appendix C, and the GH QQ plots in Appendix E. These figures demonstrate that the left skewed tail of the *Fe* distribution and the right skewed tail of the *Mn* distribution are not well reproduced by either interpolation technique; a typical result of the linear combination process. As the kriging algorithm assigns weights to the informing data, it is



(a) True values (b) OCK estimate (c) GH estimate

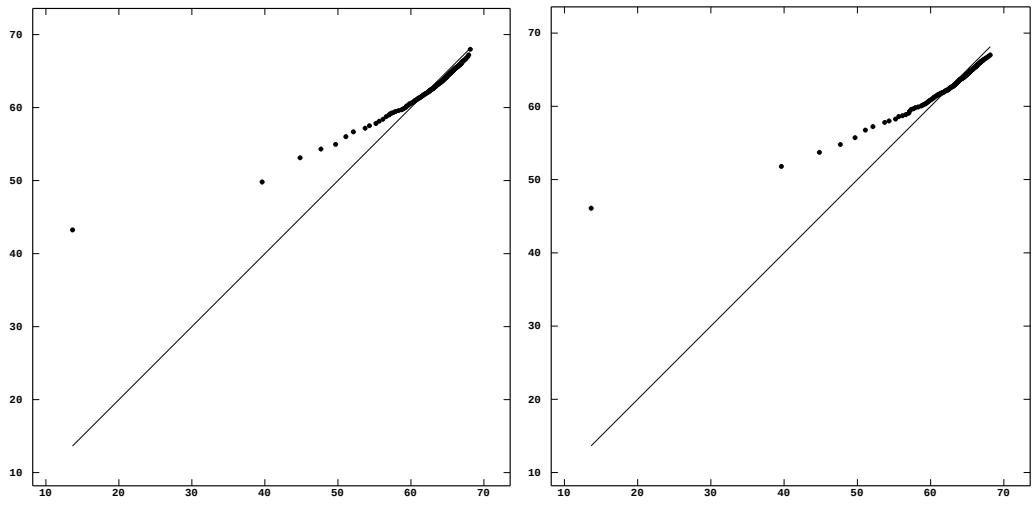
Figure 6.1: Spacing 1 *Fe* Histograms



(a) True values (b) OCK estimate (c) GH estimate

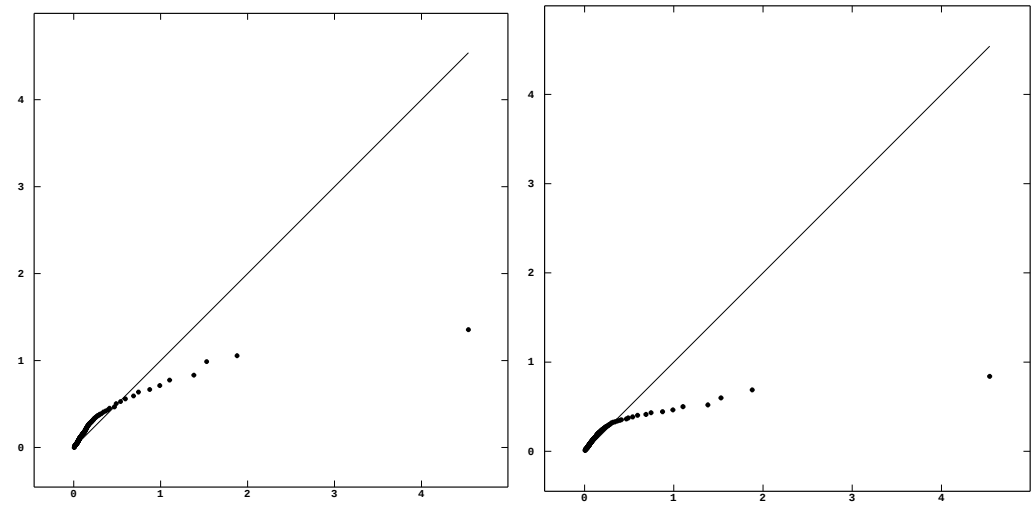
Figure 6.2: Spacing 1 *Mn* Histograms

only in the relatively rare circumstance where weights greater than or equal to 1 (coupled with negative weights to ensure unbiasedness) are assigned, or the minimum number of samples is set to 1 (effectively nearest neighbour technique), that the input maximum and minimum values are reproduced or exceeded [10]. Thus in general, the range of kriging estimates is wholly contained within the range of the input data, with the exception of negative estimates which are noted (Table 6.1). Given that the search neighbourhoods and data locations are constant between the OCK and GH estimation passes, it is highly probable that the same location in the GH kriging process step computed a result lower than the lowest input datum. However, the Gauss-Hermite quadrature step maintains the back-transformed estimate to within the range of the input data. The scatter-plots (Figure 6.5) show that the correlation is better reproduced using the GH methodology. The OCK methodology can enforce a highly linear relationship between two variables (Figure 6.5), however, the nuance of the dataset with the contaminated edge samples is visibly better reproduced using the GH methodology. This is most likely a direct consequence of the pseudo-normal distribution of the logratio variables which will, by design, be better reproduced by the cokriging algorithm which is optimal in Gaussian sample spaces (See Section 2.8 for discussion on this point). The sum of the analytes including the filler variable is of prime importance; by construction the informing dataset sums to 100% in all locations. The GH cross-validated results also sum to 100% in all locations, an expected and constructed result. The OCK results do not, although the mean value 100.00001 is very similar to the constructed sample mean value. Of further interest is the sum of an-



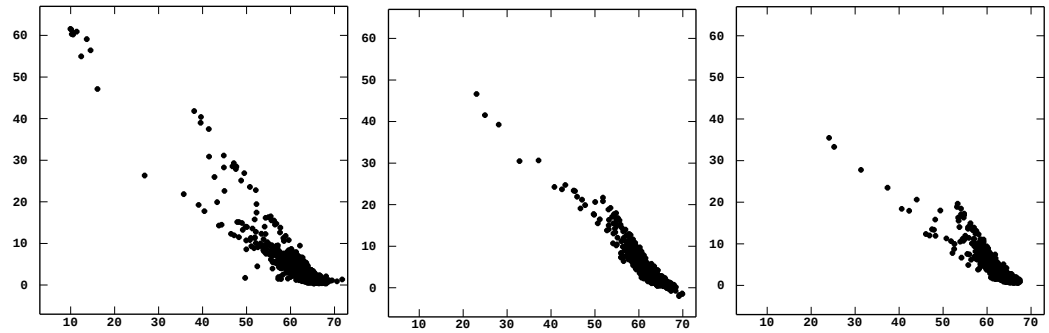
(a) OCK (b) GH

Figure 6.3: Spacing 1 *Fe* QQ plots - True Value on abscissa, estimates on ordinate



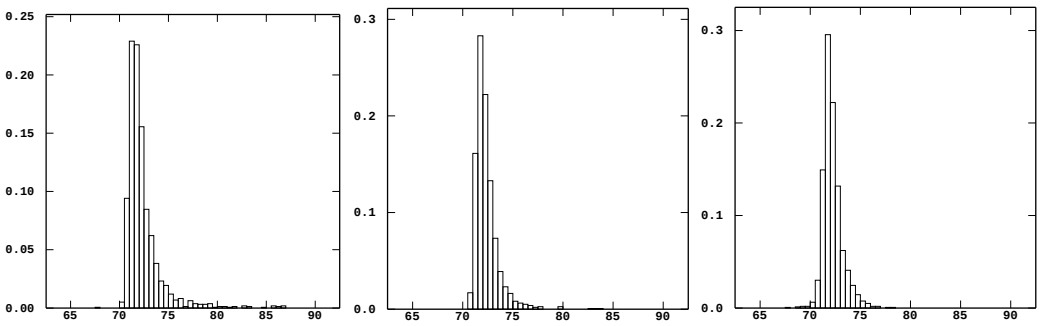
(a) OCK (b) GH

Figure 6.4: Spacing 1 *Mn* QQ plots - True Value on abscissa, estimates on ordinate



(a) True values (b) OCK estimate (c) GH estimate

Figure 6.5: Scatter plot Spacing 1 *Fe* (abscissa) versus SiO_2 (ordinate)



(a) True values (b) OCK estimate (c) GH estimate

Figure 6.6: Histogram Spacing 1 *sum_var*

alytes without the influence of the filler variable. The quantity calculated and analysed *sum_var* is the following:

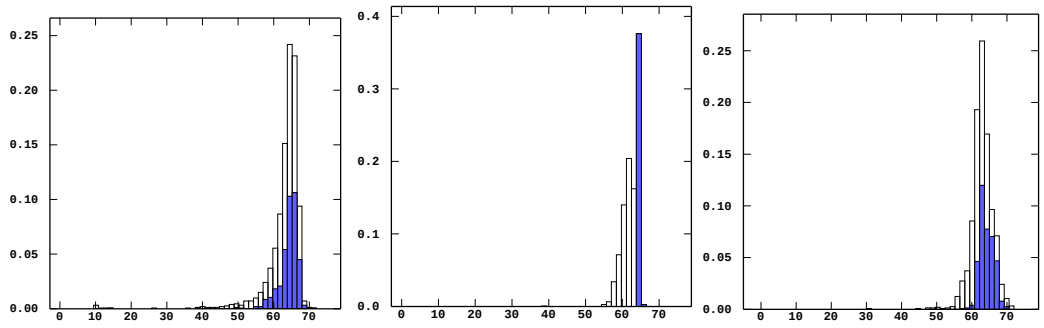
$$sum\_var = Fe + SiO_2 + Al_2O_3 + P + S + LOI + Mn. \tag{6.1}$$

The *sum_var* was calculated for each dataset and each spatial density. The results for Spacing 1 are shown in Table 6.2. The statistics illustrate that although the mean value is well reproduced by both estimation techniques, the OCK mean estimate is superior to the GH mean estimate as it is closer to the true value, albeit by a small margin. However, the mean does not completely describe the performance as the minimum value is better reproduced by the GH technique, the maximum value is better reproduced by the OCK which also reproduces the variability of the data better. Neither estimated distributions display variability close to that of the input dataset; the reduction is typical of linear estimators. The histograms in Figure 6.6 for the three distributions display different morphology, the

Table 6.2: Statistics of *sum_var* variable for Spacing 1

Attribute	n	Minimum	Maximum	Mean	Std. Dev.
True values	1594	67.528	86.632	72.382	1.93
OCK	1594	70.719	83.862	72.358	1.197
GH	1594	67.738	78.481	72.215	0.979

main difference between the two estimated distributions is the reproduction of the lower end of the distribution. The OCK results do not adequately reproduce the very low values, a typical result for a linear estimator; the non-linear nature of the GH approach assists with this aspect. It is postulated that, similarly to the distributions of individual analytes, the pseudo-normal logratio distributions assists with the reproduction of skewed tails due to the optimality of the cokriging estimator with normally distributed data.



(a) True values (b) OCK estimate (c) GH estimate

Figure 6.7: Spacing 5 *Fe* Histograms

6.1.2 Jack knifed distributions (Spacing 5)

As noted in the opening paragraph of Chapter 6, only the sparsest data density results are examined in detail. The exhaustive dataset and the sparsest dataset represent end-members; only where necessary to validate a hypothesis are the other jack-knifed estimates explicitly referred to. The complete results of the jack-knife estimates are available in Appendix B (OCK statistics), Appendix C (OCK QQ plots), Appendix D (GH statistics), Appendix E (GH QQ plots), Appendix F (Scatter-plots), and Appendix K (Histograms of sample data and estimates). The estimated distributions based on spacing 5

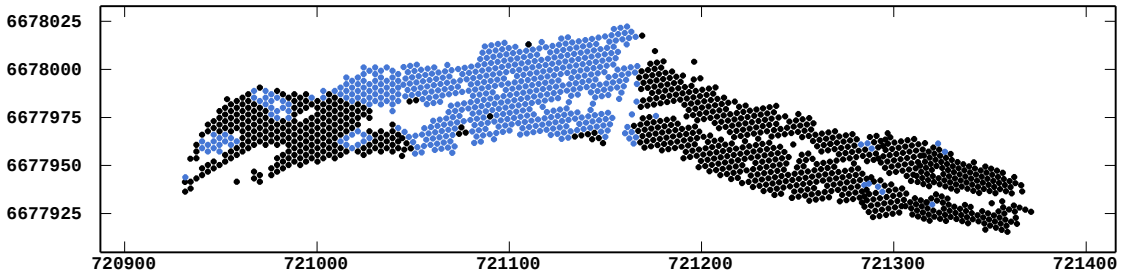
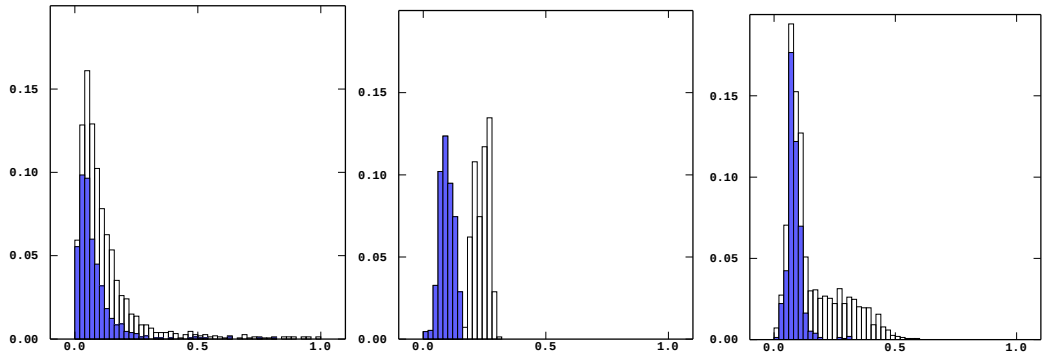


Figure 6.8: Jack-knifing Basemap - *Fe* - Spacing 5 - Blue represents the right hand bimodal peak selection of the OCK

(Figures 6.7 and 6.9) show deviation from the true distribution (Figure 5.2) and the effects of being model driven, rather than data driven as the estimates are in the Spacing 1 scenario. The distribution of the *Fe* OCK estimates has a narrower spread and a bimodality not noted in either the true values or the GH estimates. The bimodality is clearly visible in the QQ plots (Figure 6.11) as the inflection in the curve. This effect is emphasised in the *Mn* distributions (Figure 6.9) where the OCK estimate is clearly bimodal, a property the true value data do not share and which is not evident in the GH estimates. Examination of the data showed that the bimodality is caused by the use of two estimation domains with differing variogram models and orientations that were later recombined. The estimation locations corresponding to the left hand peak are shown highlighted in blue across all three histograms (Figure 6.9) and the Jack-knifing basemap for Spacing 5 data (Figure



(a) True values (b) OCK estimate (c) GH estimate

Figure 6.9: Spacing 5 *Mn* Histograms

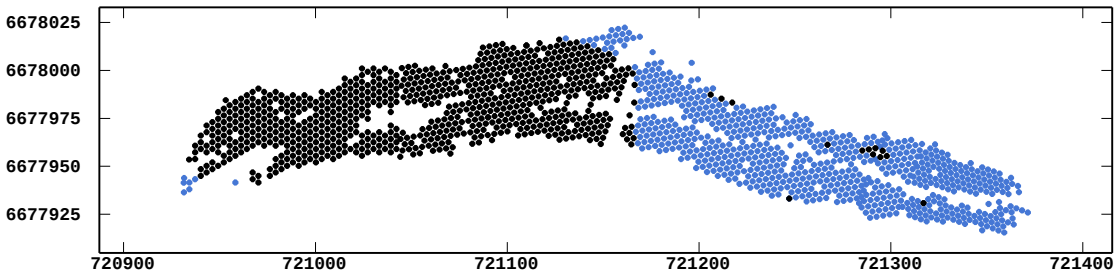
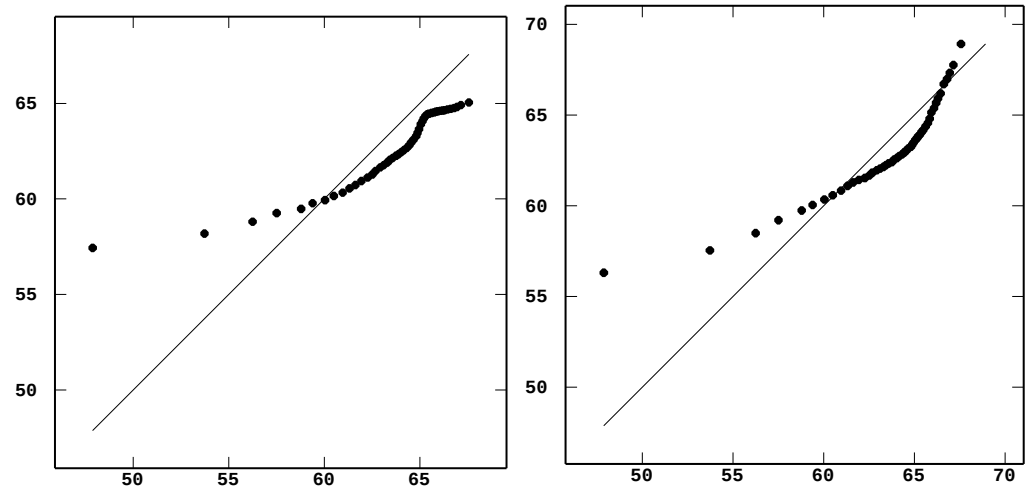


Figure 6.10: Jack-knifing Basemap - *Mn* - Spacing 5 - Blue represents the left hand bimodal peak selection of the OCK

6.10). It is hypothesised that these effects are most pronounced in the *Mn* estimates due to the following:

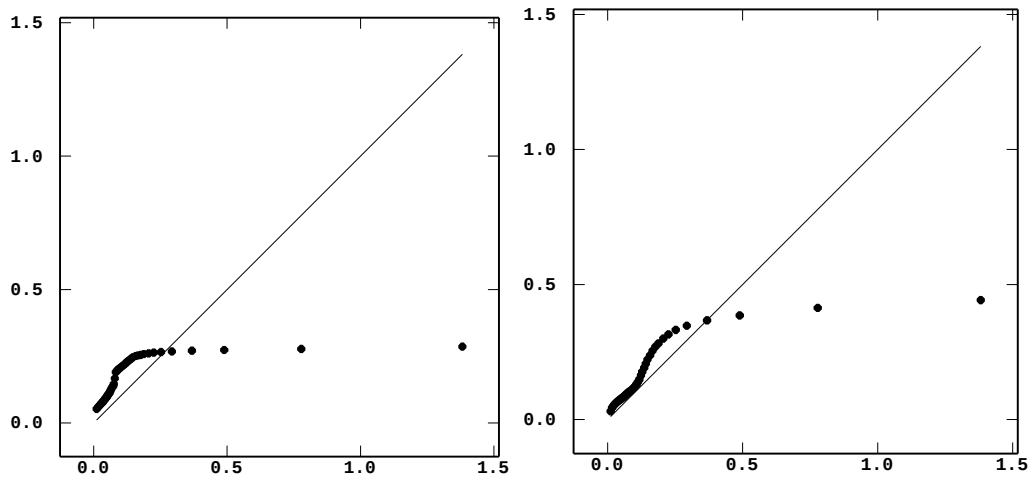
1. the highly skewed nature of the input data,
2. the distinct difference between the *Mn* mean values in between the East and West blocks, and
3. the very low data density of the informing dataset.

An important feature of the true value histogram is the exponential appearing drop-off of frequencies with increasing grade; this is partially reproduced in the GH estimates and the greater similarity between the True value and GH estimates is discerned in the QQ plots (Figure 6.12). This appears to support the proposition that the psuedo-normal additive logratio distribution is beneficial with input data that are highly skewed. Figures 6.15 and 6.16 show the consistency in the morphology of the the Spacing 1 and Spacing 5 estimated distributions, even as the informing data density decreases. The scatter plots of the *Fe* and SiO_2 shown in Figure 6.13 demonstrate the model driven nature of the estimates at this low informing data density. The linear nature of the OCK estimates and the enforcement of the correlation modeled through the LMC is easy to discern visually. This is juxtaposed with the curvilinear form of the relationship between the GH estimates which is reproducing the strict compositional nature imposed by the GH back-transform.



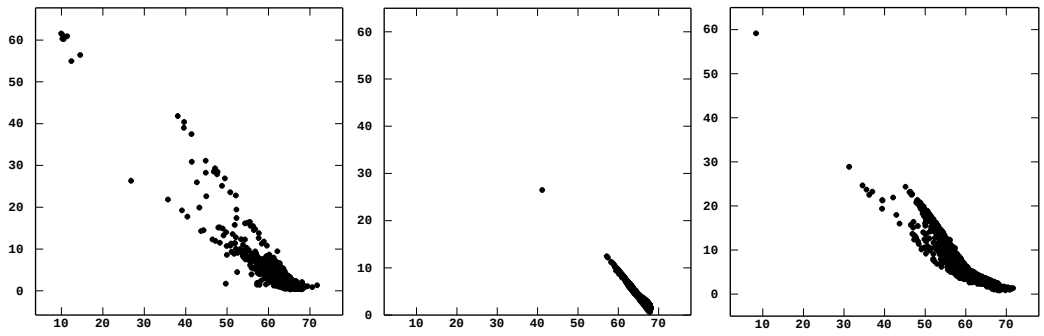
(a) OCK (b) GH

Figure 6.11: Spacing 5 *Fe* QQ plots - True Value on abscissa, estimates on ordinate



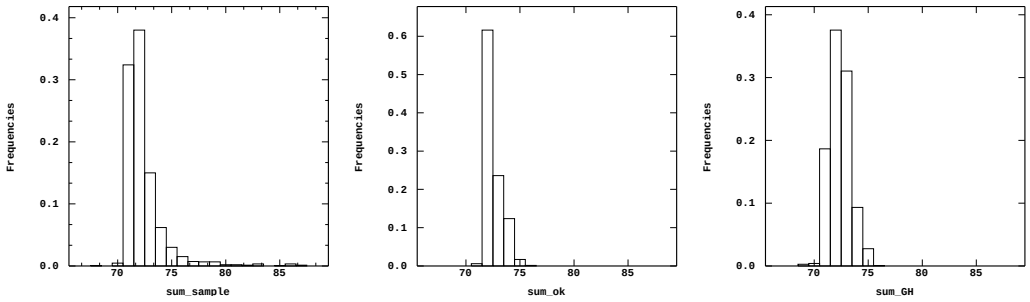
(a) OCK (b) GH

Figure 6.12: Spacing 5 *Mn* QQ plots - True Value on abscissa, estimates on ordinate



(a) True values (b) OCK estimate (c) GH estimate

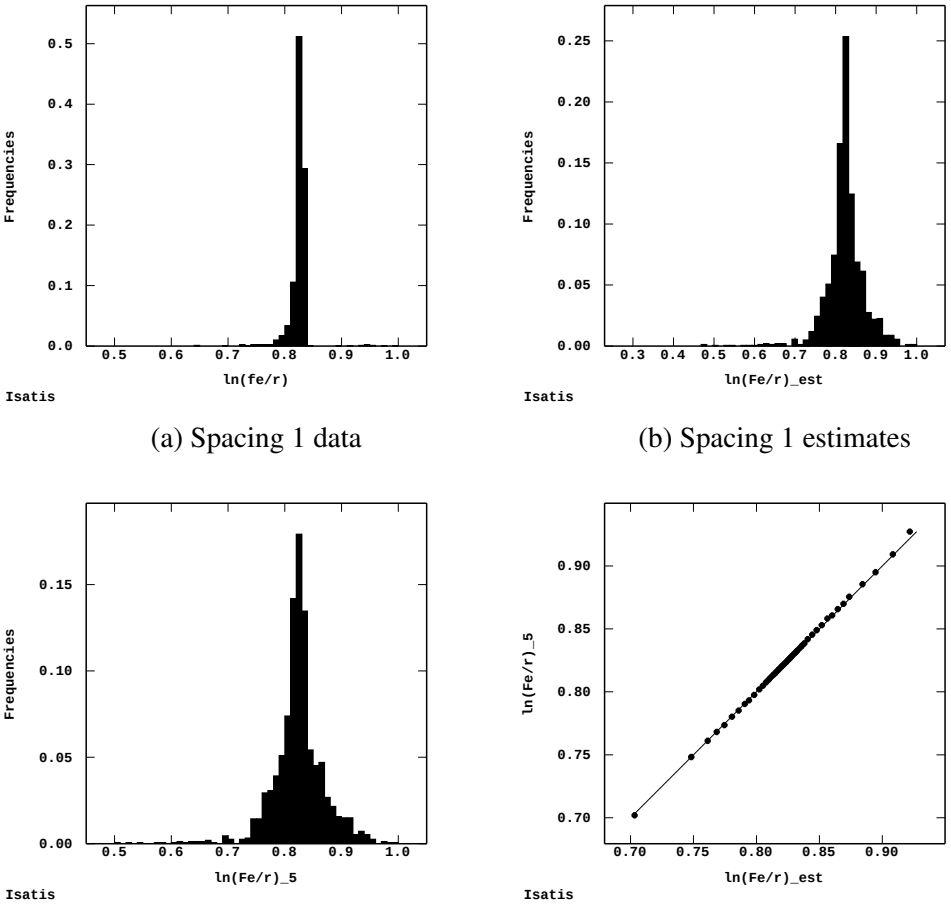
Figure 6.13: Scatter plot Spacing 5 *Fe* versus SiO_2



(a) True values (b) OCK estimate (c) GH estimate

Figure 6.14: Histogram Spacing 5 *sum_var*

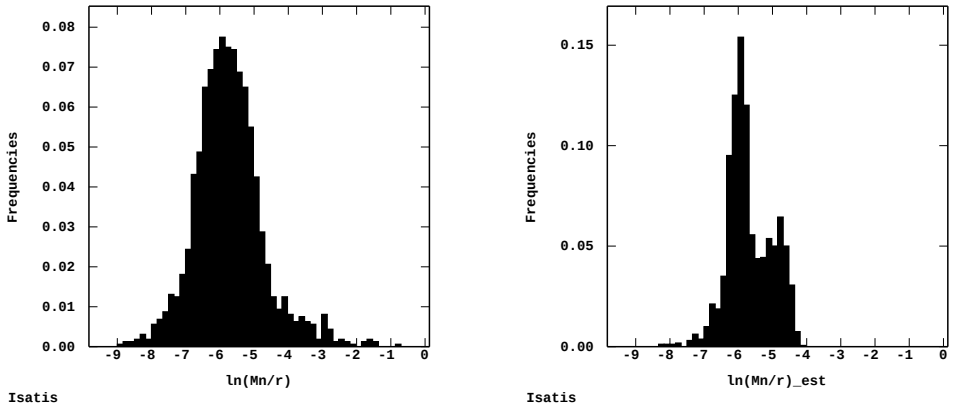
Neither is a perfect reproduction of the input data at this data density but the relationship between the GH estimates carries more compositional reality than the strict linearity of the OCK estimate.



(a) Spacing 1 data (b) Spacing 1 estimates

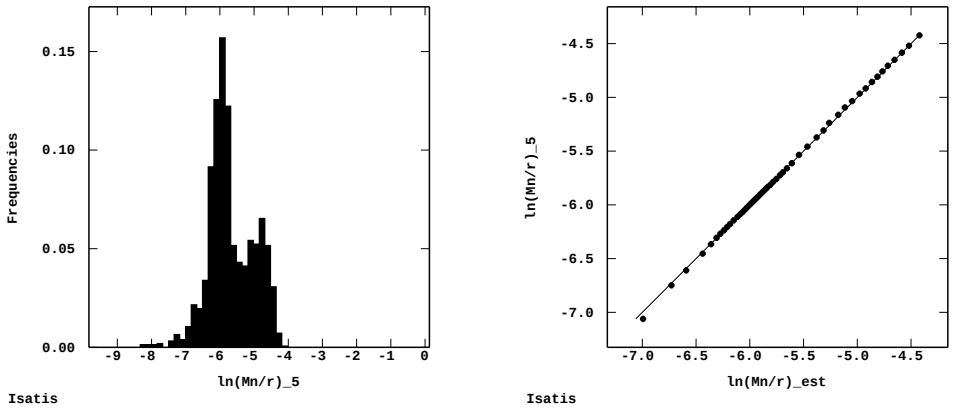
(c) Spacing 5 estimates (d) QQ plot Spacing 1 vs 5 estimates

Figure 6.15: alr cokriging outcomes: Fe



(a) Spacing 1 data

(b) Spacing 1 estimates



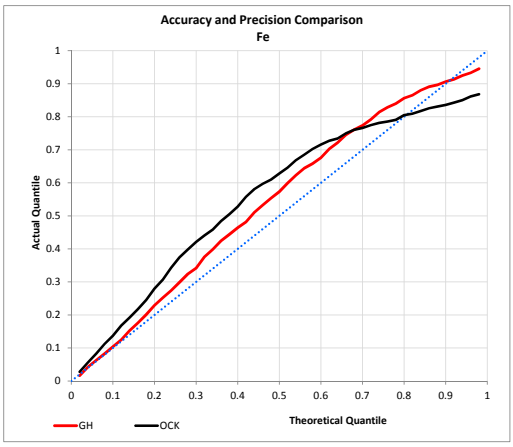
(c) Spacing 5 estimates

(d) QQ plot Spacing 1 vs 5 estimates

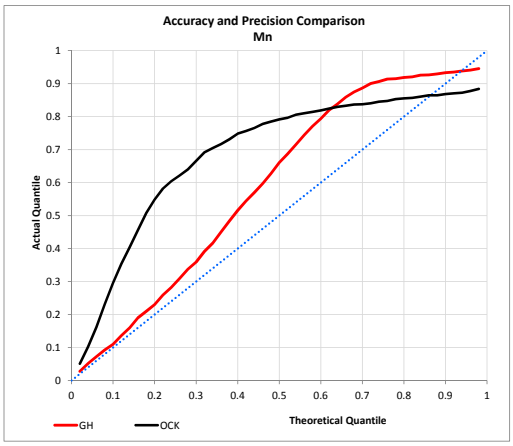
Figure 6.16: alr cokriging outcomes: Mn

6.2 Accuracy and Precision comparison

In Sections 6.1 the shape and statistics of the distributions of estimated analytes and univariate errors have been discussed. To further compare the methods the accuracy and precision of the univariate results are calculated and compared using the techniques described in Section 2.11. As in Section 6.3.2 the *accuracy plots* shown in Figure 6.17 and 6.18 the *Fe* and *Mn* distributions are examined, firstly for the exhaustive data and then for the Spacing 5 dataset. Overall the GH curve both for the *Fe* and *Mn* exhaustive sets show greater accuracy than the OCK curves. This conclusion is reached through the red GH curve maintaining a positive (above the 45°line) position for a greater portion of the theoretic centile range than the black OCK curve and it lies closer to the 45°line. These conclusions are supported by the statistics in Table 6.3 where for Spacing 1 *Fe* and *Mn* the GH values are closer to 1.0 for all three criteria. The Spacing 5 data are not as definitive; the GH *Fe* curve follows the same pattern as the Spacing 1 data but the *Mn* curve does not as the OCK estimated distribution better reproduces the true values, this is upheld by the accuracy statistic, even though the goodness values for the GH estimate are better than those of OCK. When examining the Spacing 5 plot (Figure 6.18) the two estimated *Fe*

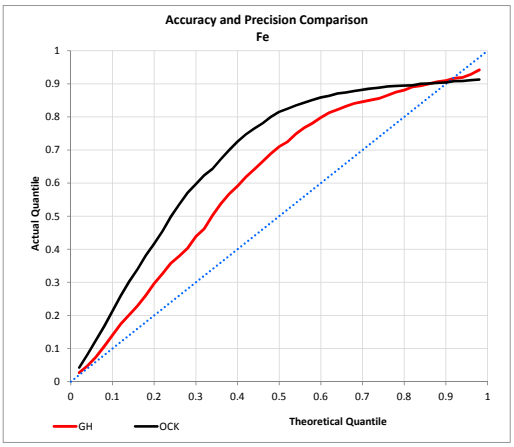


(a) *Fe* estimates

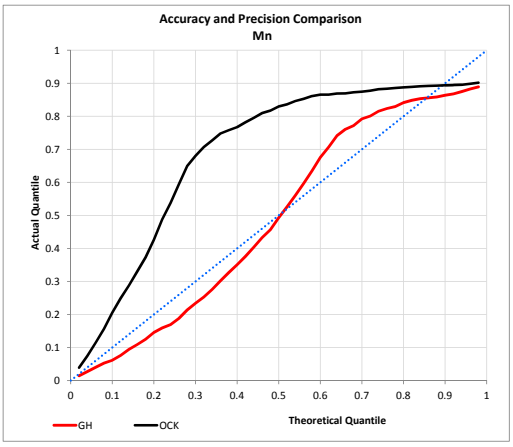


(b) *Mn* estimates

Figure 6.17: Comparative accuracy plots Spacing 1



(a) *Fe* estimates



(b) *Mn* estimates

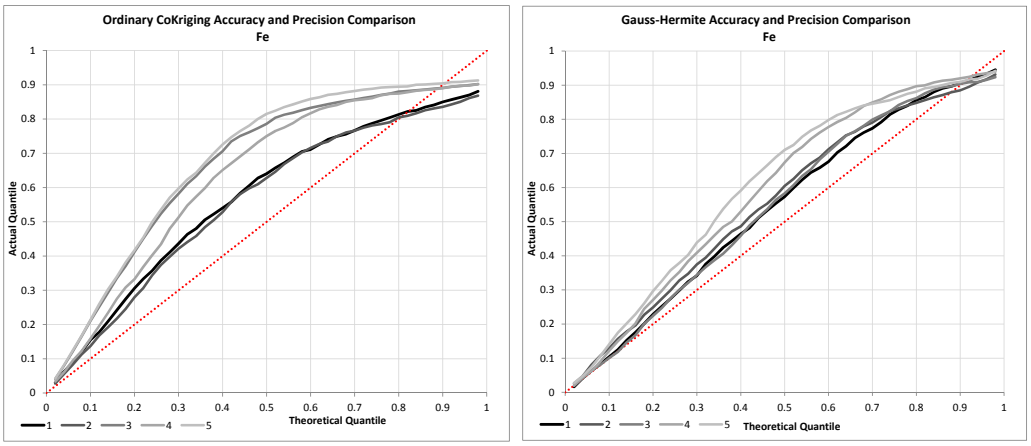
Figure 6.18: Comparative accuracy plots Spacing 5

distributions are equally accurate, shown by the two curves intersecting the 45°line at the same point but the GH estimated distribution is more precise, shown by the closeness of the red curve to the 45°line. The Spacing 5 *Mn* plot however show that although the OCK distribution is more accurate, the goodness is better for the GH distribution. This is an interesting outcome, because the goodness calculation (Equation 2.65) penalises inaccurate results with twice the weighting compared with accurate estimates. The change in accuracy and precision as a function of decreasing data spacing is illustrated in Figure 6.19. Generally the curves bow away from the 45°axis as the data density decreases, indicating poorer precision and goodness results, upheld by the results in Table 6.3. This is most likely a consequence of increase in the average distance between informing data and location of point to be estimated and the changes to the relative sills of the variograms used in the estimates. The consequence of increasing distance between the estimation location and the informing data (within the range of the variogram) will be increasing estimation variances and thus wider confidence intervals within which accuracy is preserved. The relative contribution of any informing point to the estimation variance is dependent on the

Table 6.3: Accuracy, Precision, and Goodness statistics

Element	Statistic	Spacing 1		Spacing 5	
		OCK	GH	OCK	GH
Fe	Accuracy	0.837	0.898	0.918	0.918
	Precision	0.842	0.909	0.629	0.773
	Goodness	0.902	0.951	0.808	0.883
P	Accuracy	0.327	0.980	0.755	0.000
	Precision	0.995	0.909	0.943	1.000
	Goodness	0.956	0.954	0.947	0.915
SiO_2	Accuracy	0.857	0.816	0.918	0.878
	Precision	0.725	0.891	0.543	0.880
	Goodness	0.846	0.943	0.765	0.929
Al_2O_3	Accuracy	0.837	0.776	0.816	0.857
	Precision	0.796	0.873	0.811	0.819
	Goodness	0.882	0.934	0.895	0.899
S	Accuracy	0.898	0.490	0.898	0.510
	Precision	0.557	0.976	0.610	0.983
	Goodness	0.769	0.971	0.797	0.968
Mn	Accuracy	0.898	0.939	0.898	0.347
	Precision	0.514	0.813	0.602	0.958
	Goodness	0.748	0.904	0.792	0.919
LOI	Accuracy	0.000	0.000	0.000	0.000
	Precision	1.000	1.000	1.000	1.000
	Goodness	0.938	0.907	0.865	0.770
$filler$	Accuracy	0.816	0.939	0.878	0.939
	Precision	0.847	0.851	0.803	0.658
	Goodness	0.904	0.923	0.890	0.827

relative contribution of each variogram structure to the total model. Appendix I shows the variogram models for each data density scenario; the Spacing 5 comparative matrices exhibit qualitatively better fit of the logratio model to the experimental data than the raw space model, itself most likely a result of the pseudo-normal alr distribution. It follows that for the same spatial arrangement of informing data the confidence interval for the GH estimates are likely to be comparatively narrower than the OCK. This coupled with the generally narrower univariate error distributions for GH, results in GH displaying better goodness statistics even though the accuracy as quantified by A is on occasion worse than OCK.



(a) OCK estimates (b) GH estimates

Figure 6.19: Comparative accuracy plots OCK versus GH

6.3 Univariate error measures

6.3.1 Cross-validation results (Exhaustive data)

The most desirable distribution of univariate errors is symmetrical, centred on zero and with a variance small with respect to the estimation problem being addressed. When analysing the errors generated through the application of the two estimation methodologies, the above criteria will be used. When examining the Spacing 1 univariate error (δ_u) results shown in Table 6.4, the OCK estimate is better centred on zero than the GH estimate in every instance. The largest bias occurs with the Fe and SiO_2-GH estimates, which display the reverse bias with respect to one another. The enforcement of the correlation between Fe and SiO_2 through the LMC could account for the discrepancy noted here as the fit of the LMC is usually a compromise [10] and the bias in the Fe is negative (overestimates) where the bias in the SiO_2 is positive (underestimates), matching the negative correlation between the two analytes. Regarding the range of error values, smaller is better; each technique has the larger range for four analytes. The OCK method range of errors is smaller for Al_2O_3, Mn, P, and $filler$, while the GH estimates have a smaller range for Fe, LOI, S, and SiO_2. The symmetry appears marginally better for the Fe GH error distribution (Figure 6.20) than the OCK distribution and appears considerably better for the Mn distributions (Figure 6.21) where the OCK curve is skewed to the left when compared with the GH curve. As the differences between the distributions are small, QQ plots of the error distributions were generated to identify particular ranges of values where one estimation methodology is superior to another. Figure 6.22 shows that when examining the Fe curve, the OCK error curve is superior to the GH error curve below zero error, approximately equal at the zero error point and inferior above the zero point. In contrast, the Mn distribution shows that the GH error distribution is superior from -0.017 through to 0.101 and slightly inferior outside of these bounds. This region represents the majority

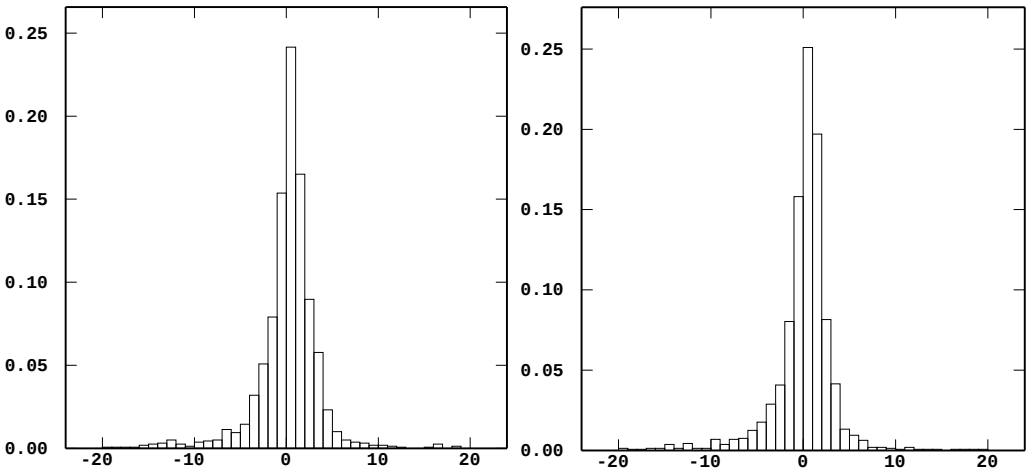
of the error range. Overall this emphasises that statistics alone cannot be used to judge the goodness of an error measure, but that other criteria and methods should complement the statistics. For this reason, the alternative complementary methods quantifying the accuracy, precision and goodness of the estimates are described in Section 2.11 and implemented in Section 6.2.

Table 6.4: Spacing 1 Univariate error δ_u (n=1594)

Attribute	Minimum	Maximum	Mean	Std. Dev.
Al_2O_3 GH	-5.624	15.069	0.068	1.611
Al_2O_3 OCK	-4.569	15.489	0.031	1.623
Fe GH	-54.121	19.747	-0.224	4.7
Fe OCK	-54.435	21.934	-0.081	4.709
LOI GH	-4.296	5.927	0.003	1.149
LOI OCK	-4.377	6.398	0.001	1.14
Mn GH	-1.693	12.115	0.032	0.512
Mn OCK	-1.404	11.927	0.002	0.516
P GH	-0.164	0.162	-0.001	0.034
P OCK	-0.155	0.134	0	0.033
S GH	-0.276	0.738	0.002	0.044
S OCK	-0.187	0.851	0	0.045
SiO_2 GH	-12.536	58.791	0.288	4.743
SiO_2 OCK	-24.011	58.59	0.072	4.746
filler GH	-15.704	3.514	-0.167	1.664
filler OCK	-14.061	5.018	-0.024	1.547

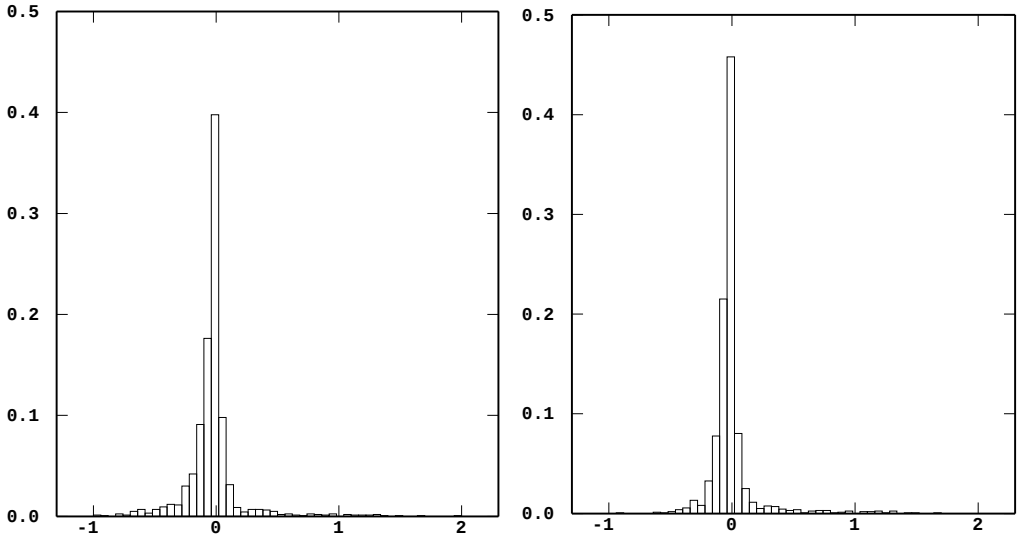
6.3.2 Jack-knife results (Low spatial density)

At the lower spatial density (as epitomised by the Spacing 5 results), the two techniques generate mean errors of similar magnitude, in fact four out of eight analytes are better estimated, in a mean error sense, by the GH methodology than by OCK (Table 6.5). The range of error values is generally smaller for the OCK estimates, and, correspondingly the standard deviation is usually smaller too. When the histograms of errors are examined (Figures 6.25 and 6.26) the same commentary used for the exhaustive cross-validated data can be used with the low informing density estimates. The *Fe* histograms are very similar, only examination of the QQ plot (Figure 6.27) shows that although the range is larger, the GH estimate outperforms the OCK estimate for the majority of the error range examined, the two coming into parity around the zero error point on the plot. The *Mn* histograms and QQ plots are different from the Spacing 1 results; the GH estimates are superior across the majority of the error grade range with a more symmetrical distribution and smaller mean



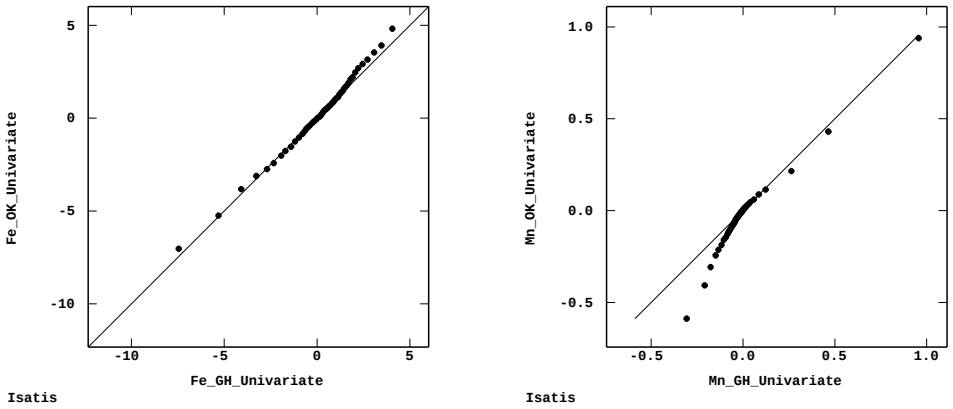
(a) OCK errors (b) GH errors

Figure 6.20: Spacing 1 *Fe* Error frequency histograms



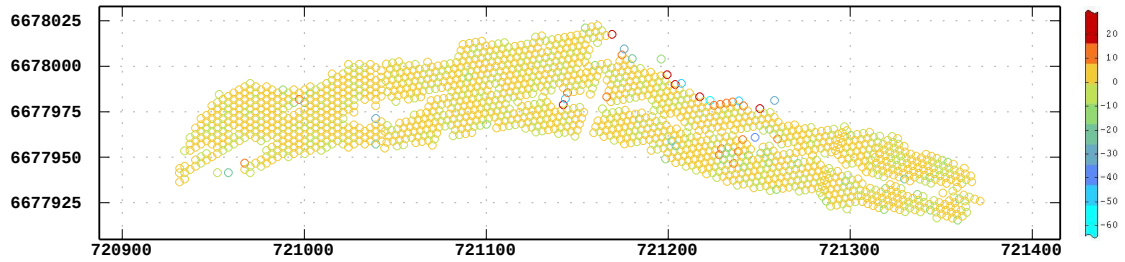
(a) OCK errors (b) GH errors

Figure 6.21: Spacing 1 *Mn* Error histograms

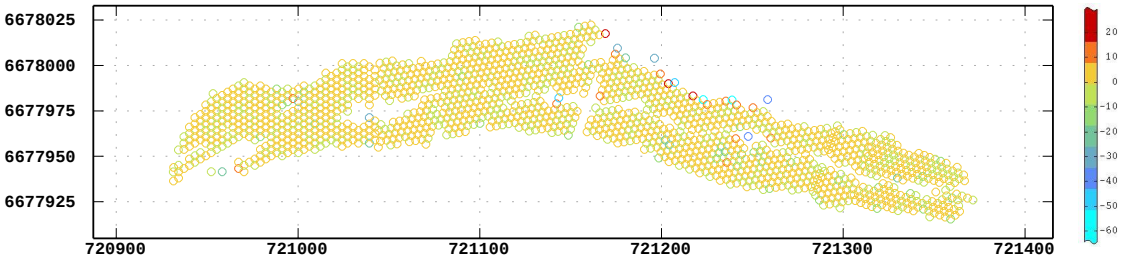


(a) *Fe* (b) *Mn*

Figure 6.22: Spacing 1 Error QQ plots - GH on abscissa and OCK on ordinate

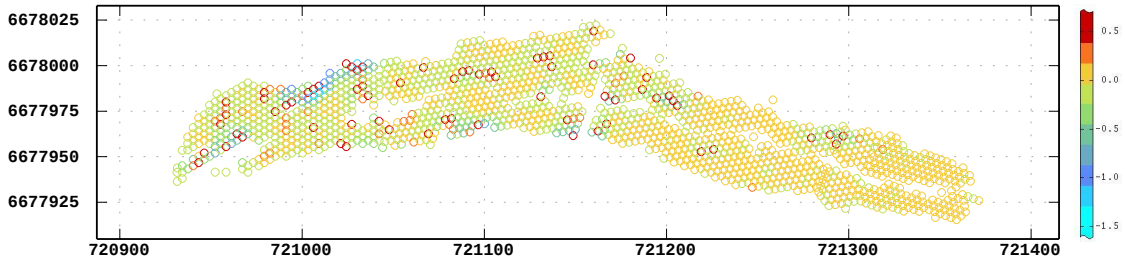


(a) OCK estimates

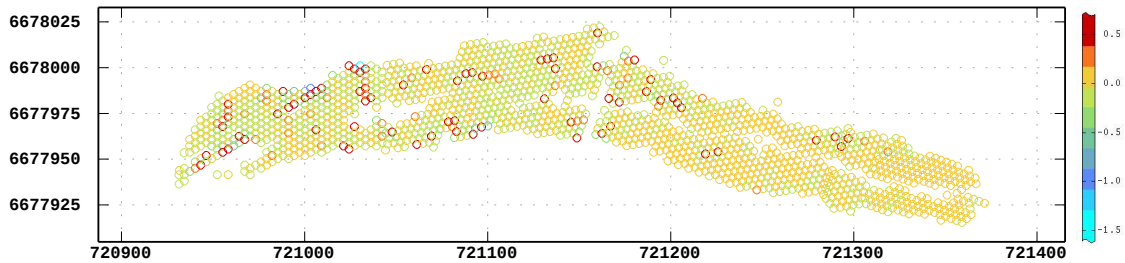


(b) GH estimates

Figure 6.23: Spacing 1 Error Basemaps *Fe*



(a) OCK estimates



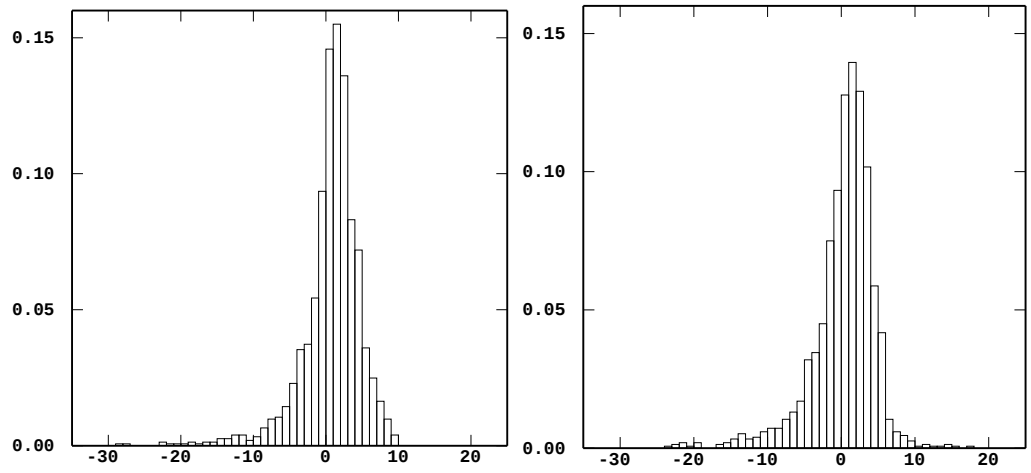
(b) GH estimates

Figure 6.24: Spacing 1 Error Basemaps *Mn*

error. The basemaps of errors given in Figures 6.28 and 6.29 show different clustering of the extreme high and low zones of errors which are different from one technique to the next. Given that the search criteria used to inform each estimate were the same for each method, it is postulated that the differences are rooted in the variogram model and/or back-transformation portions of each process. The impact of changing the data density on the mean error statistic is shown in Figure 6.30, here the close correspondence between the two different methods is noted; in general the OCK method is superior when measured this way.

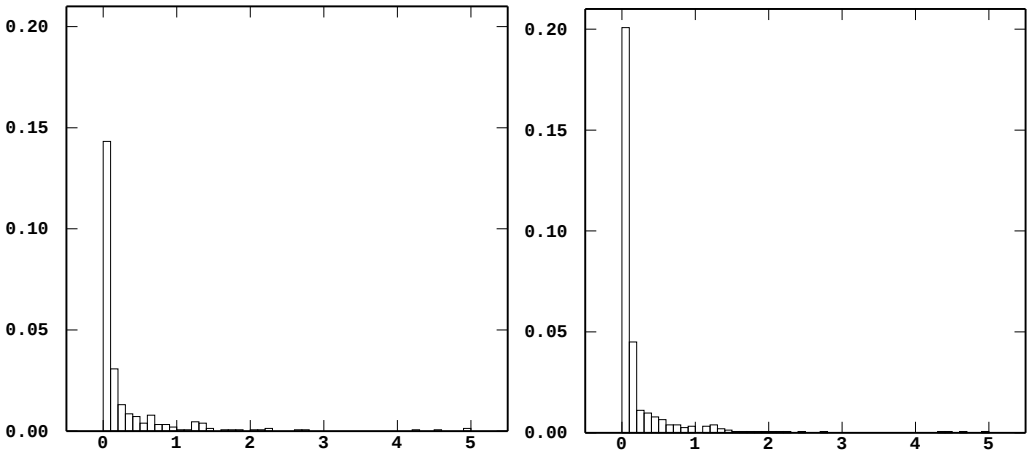
Table 6.5: Spacing 5 Univariate error δ_u (n=1534)

Attribute	Minimum	Maximum	Mean	Std.Dev.
Al_2O_3 GH	-7.55	18.01	-0.20	1.96
Al_2O_3 OCK	-3.31	17.62	-0.17	1.91
Fe GH	-52.41	20.10	0.08	5.61
Fe OCK	-52.64	9.73	0.49	5.39
LOI GH	-5.02	6.62	0.09	1.49
LOI OCK	-3.38	6.60	0.10	1.42
Mn GH	-0.47	12.35	0.03	0.54
Mn OCK	-0.26	12.65	0.02	0.55
P GH	-0.19	0.21	-0.01	0.05
P OCK	-0.15	0.24	0.00	0.05
S GH	-0.17	0.88	0.00	0.05
S OCK	-0.10	0.88	0.00	0.05
SiO_2 GH	-12.91	57.54	-0.03	5.42
SiO_2 OCK	-8.97	57.63	-0.51	5.45
$filler$ GH	-15.10	6.07	0.03	1.95
$filler$ OCK	-13.92	4.29	0.06	1.82



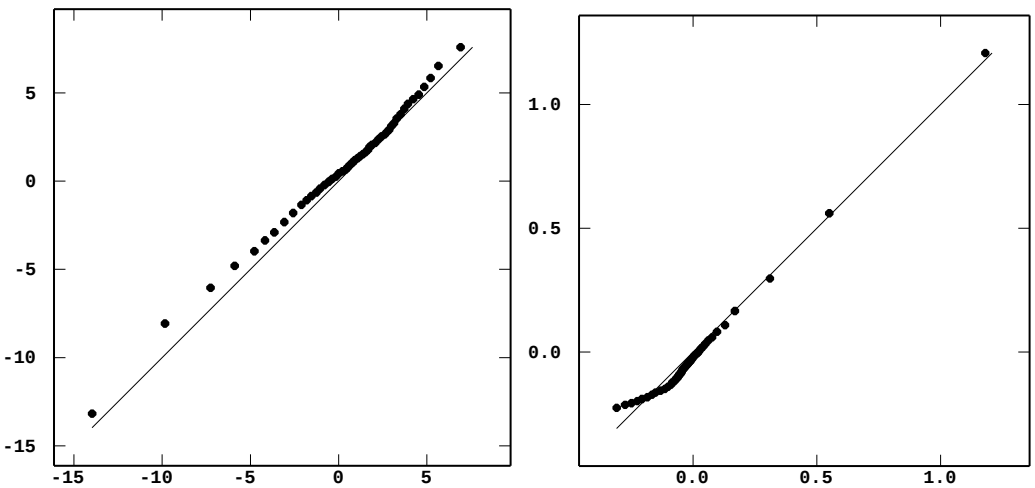
(a) OCK errors (b) GH errors

Figure 6.25: Spacing 5 *Fe* Error histograms



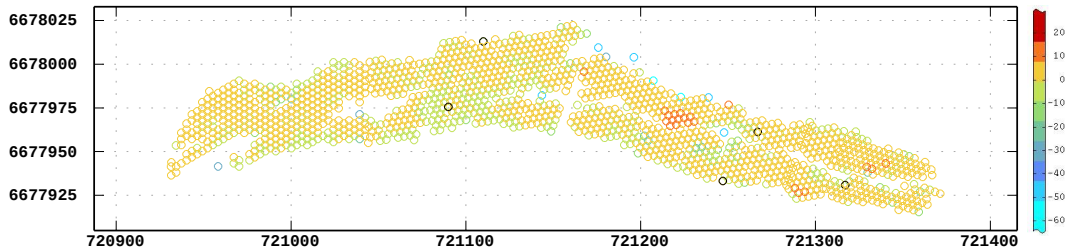
(a) OCK errors (b) GH errors

Figure 6.26: Spacing 5 *Mn* Error histograms

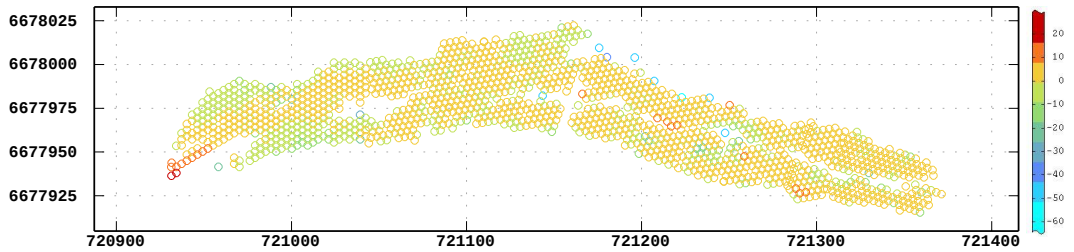


(a) *Fe* (b) *Mn*

Figure 6.27: Spacing 5 Error QQ plots - GH on abscissa and OCK on ordinate

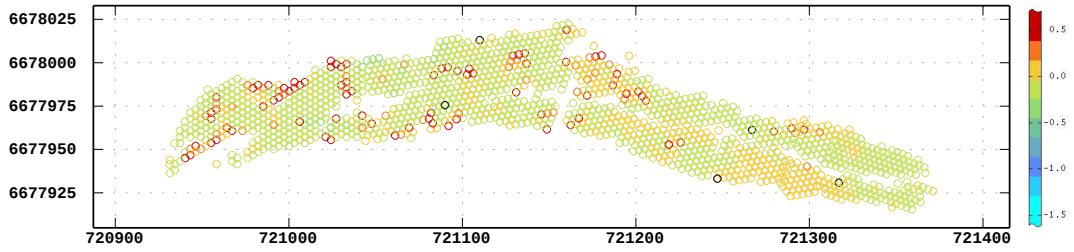


(a) OCK estimates

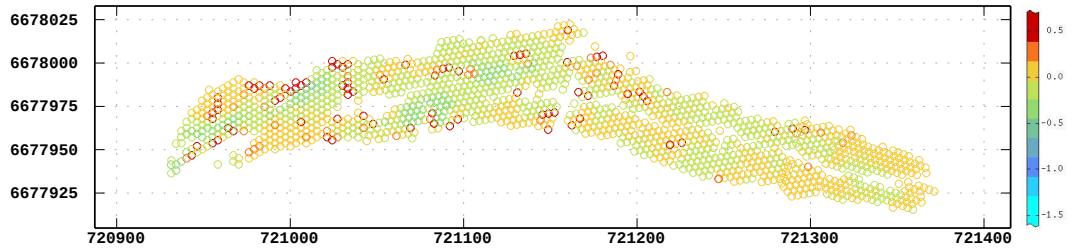


(b) GH estimates

Figure 6.28: Spacing 5 Error Basemaps *Fe*

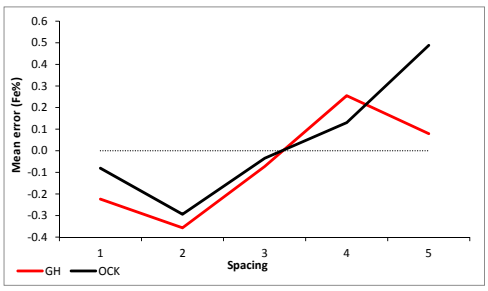


(a) OCK estimates

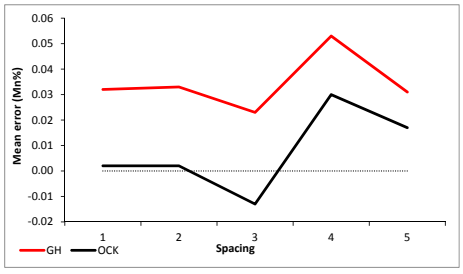


(b) GH estimates

Figure 6.29: Spacing 5 Error Basemaps *Mn*



(a) *Fe* estimates



(b) *Mn* estimates

Figure 6.30: Mean error as function of Spacing

6.4 Compositional error measures

The two compositional distance measures selected for use in this study, Euclidean (δ_e) and Aitchison (δ_a) are variations on the sum of squared differences family and are therefore always positive (Equations 2.67 and 2.68). In addition, the Aitchison error is only calculable for strictly positive components, which results in differing numbers of observations that could be calculated and used in the error calculations as several parts had negative OCK estimates.

6.4.1 Euclidean Error

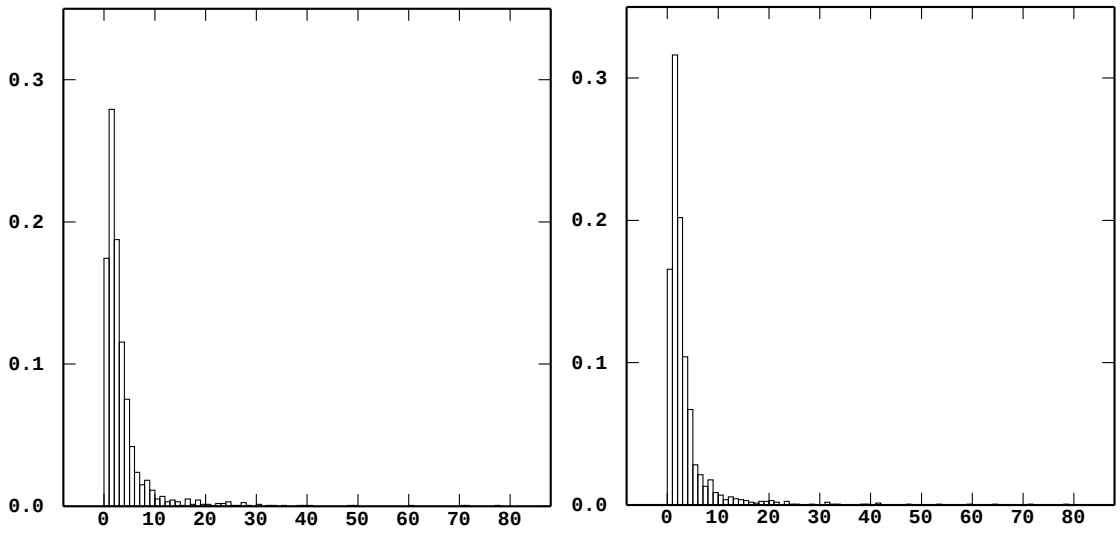
The mean Euclidean error δ_e statistics are shown in Table 6.6 where the following features are noted:

- the mean error from the GH data is closer to zero for all the spacings except Spacing 5
- the GH data are more skewed than the OCK data for all cases except Spacing 5
- in all cases except Spacing 4 the GH errors have a larger standard deviation than the OCK errors.

Table 6.6: Euclidean error statistics

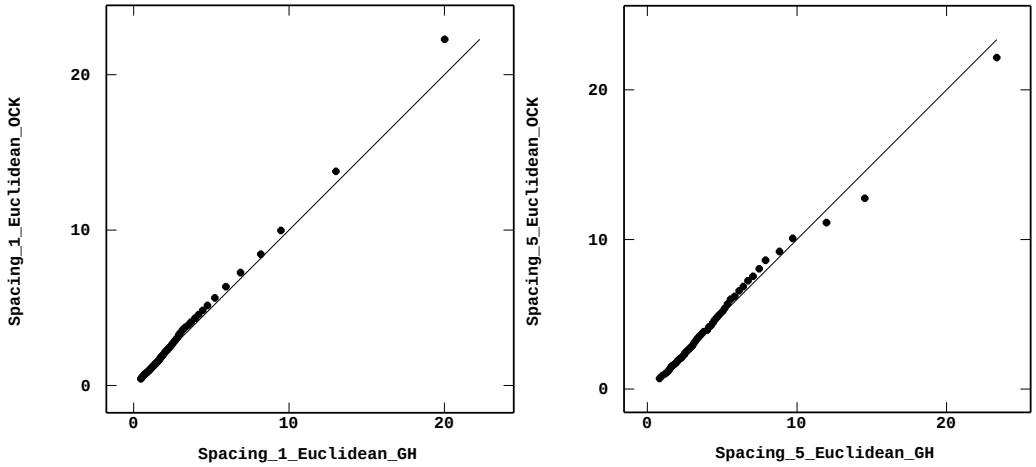
Spacing	Method	Count	Minimum	Maximum	Mean	Std. Dev.	Skewness
1	GH	1594	0.22	81.66	3.61	6.22	7.19
	OCK	1594	0.19	81.97	3.77	6.09	6.60
2	GH	1173	0.17	80.70	3.69	5.94	7.49
	OCK	1173	0.16	78.61	3.78	5.92	7.15
3	GH	1370	0.22	80.57	4.10	6.27	7.12
	OCK	1370	0.24	79.32	4.16	6.08	6.90
4	GH	1468	0.33	80.76	4.45	6.39	6.74
	OCK	1468	0.22	81.31	4.54	6.44	6.50
5	GH	1534	0.26	79.75	5.12	6.71	6.60
	OCK	1534	0.29	80.08	5.08	6.54	6.73

The histograms of the Euclidean errors for Spacing 1 and Spacing 5 (Figure 6.31) have visually similar highly skewed morphology. The distributions are also compared using QQ plots in Figure 6.32 which shows that GH is better than OCK for the majority of each distribution's range. In each scenario, the few large deviations from the true value resulted in error values that have large relative magnitude. On visual examination few differences can be discerned between the OCK and GH errors for the Spacing 1 data.



(a) OCK (b) GH

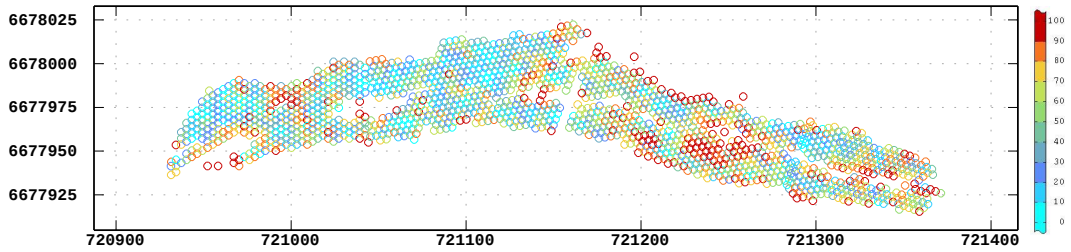
Figure 6.31: Euclidean spacing 1 error frequency histograms



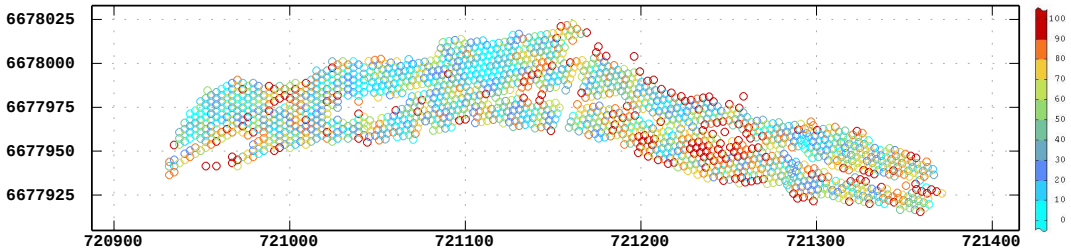
(a) Spacing 1 (b) Spacing 5

Figure 6.32: Euclidean error QQ plots - GH on abscissa and OCK on ordinate

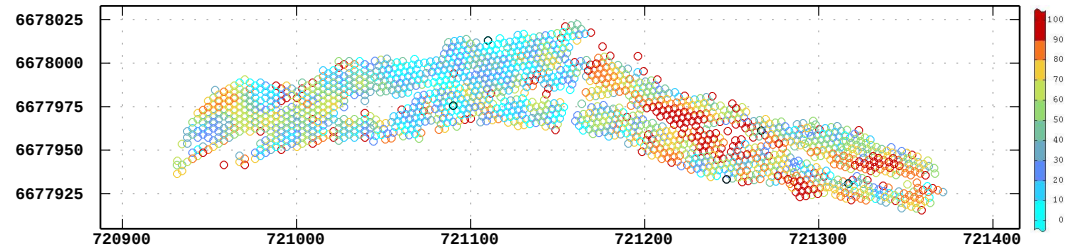
The most important difference is the greater frequency of errors at lower values for the GH data; an indicator of better performance. The Spacing 5 distributions have the same characteristic, as well as sharper drop off of the frequency of errors as a function of increasing error magnitude for the GH data. This is seemingly at odds with the quantitative tabulated data which reveal a smaller range but a higher standard deviation for the GH data. The conclusion drawn here is that the GH technique, in a mean sense, outperforms the OCK technique for the majority (4 out of 5) of the data density scenarios examined here (Figure 6.36). The spatial distribution of the errors is examined in Figure 6.33. The large and small magnitude errors for both estimation methodologies are not distributed through the estimation zone with random placement; rather the large error values tend to be on the edges where the informing data is likely to be sub-optimal and the possibility of diluted (and therefore different from surrounding) grades are expected. In contrast, the low error values cluster in the center of the estimation zone where continuity is high and



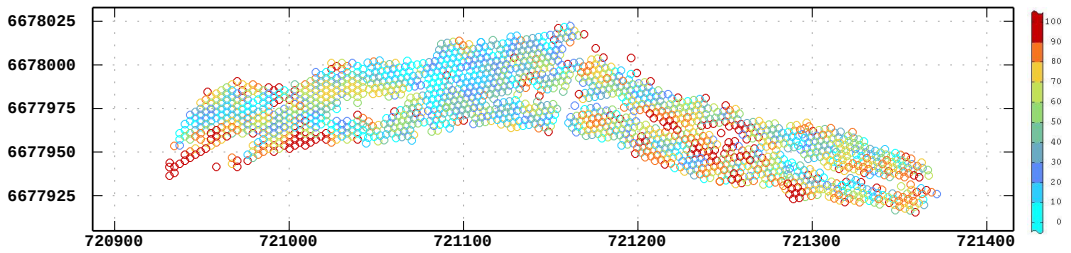
(a) Spacing 1 OCK



(b) Spacing 1 GH



(c) Spacing 5 OCK



(d) Spacing 5 GH

Figure 6.33: Euclidean compositional error basemaps

the informing data will always satisfy the optimal estimation search parameters. There is a visually greater proportion of large magnitude errors on the east side, this is evident for either method, but more marked for the OCK estimates, especially Spacing 5. The conclusion drawn from examining the spatial error plots is that the GH method produces better results than OCK using this measure.

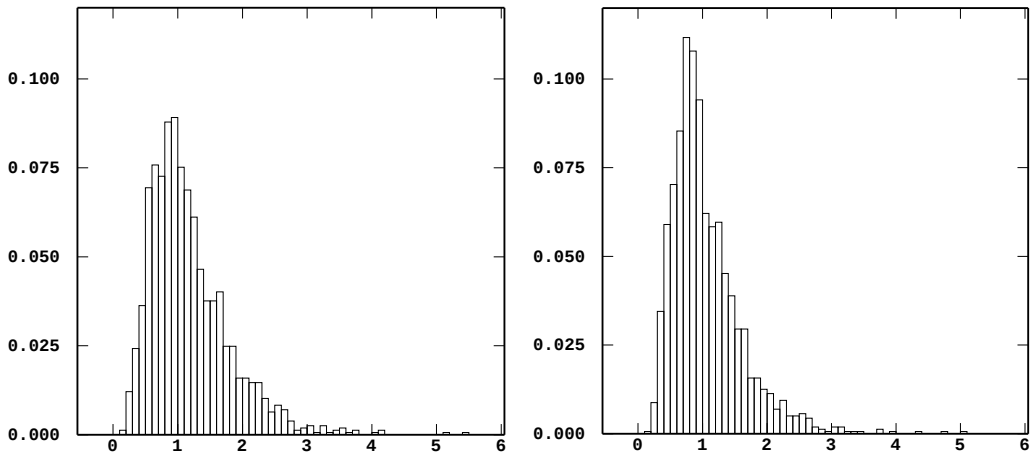
6.4.2 Aitchison Error

The histogram of GH errors for the Spacing 1 data (Figure 6.34) is more peaked about the low error portion of the abscissa than the OCK data, which in conjunction with the lower mean value (Table 6.7) indicates that when measured with the Aitchison error, the GH estimates are superior to the OCK estimates. This conclusion is supported by the QQ plots in Figure 6.35 which show the deviation of the plot away from the 45°line toward the OCK (ordinate) axis. This deviation indicates that error is always greater for the OCK estimates than for the GH estimates. The tabulated statistics (Table 6.7) also show that for Spacing 1, the range and standard deviation of the errors for GH data are smaller than those of the OCK errors. Examination of the statistics and plots of the lower spatial density data confirms that this trend continues as the data density decreases. Of specific interest is the trend plots (Figure 6.36) which show that the GH estimates outperform the OCK estimates at all data spacings, with the sole exception of the Spacing 5 Euclidean measure where the OCK estimate is marginally better. The spatial distribution of the errors

Table 6.7: Aitchison error statistics

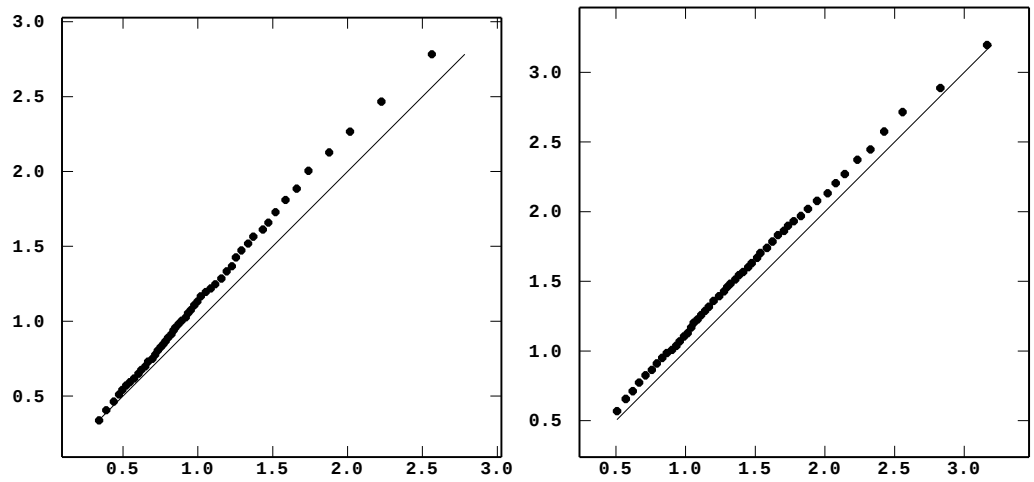
Spacing	Method	Count	Minimum	Maximum	Mean	Std. Dev.	Skewness
1	GH	1594	0.19	5.07	1.05	0.55	1.77
	OCK	1570	0.18	5.44	1.18	0.62	1.53
2	GH	1173	0.24	5.03	1.15	0.56	1.74
	OCK	1133	0.13	6.74	1.33	0.63	1.73
3	GH	1370	0.22	5.01	1.24	0.58	1.45
	OCK	1346	0.15	4.72	1.43	0.64	1.00
4	GH	1468	0.29	4.98	1.36	0.61	1.16
	OCK	1432	0.21	5.05	1.49	0.67	1.15
5	GH	1534	0.18	4.82	1.45	0.66	1.11
	OCK	1524	0.36	4.92	1.58	0.66	0.97

is examined in Figure 6.37, where the negative estimates are easily visible in the OCK plots as black circles. The placement of the negative estimates and therefore incalculable Aitchison errors is explained by the edge positioning of the estimate locations, where the search ellipse is poorly informed on one side and well on another, leading to large screen effects for deleterious (such as Al_2O_3) data with high grades. Besides the erroneous (and impossible) OCK estimates the trends within the estimation zone do not show significant spatial variation from one another. For both methods the high error values are commonly



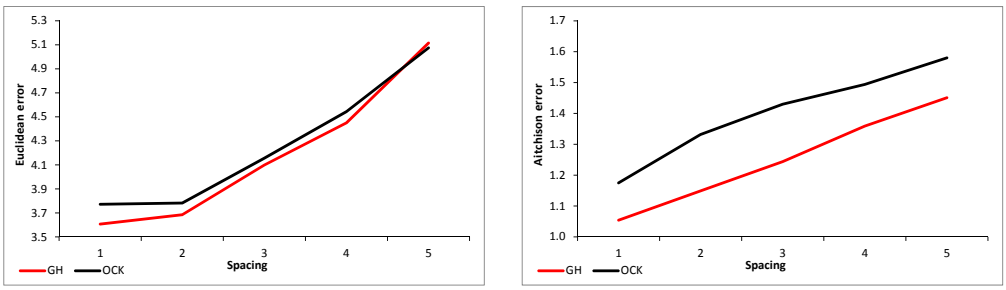
(a) OCK (b) GH

Figure 6.34: Aitchison spacing 1 error frequency histograms



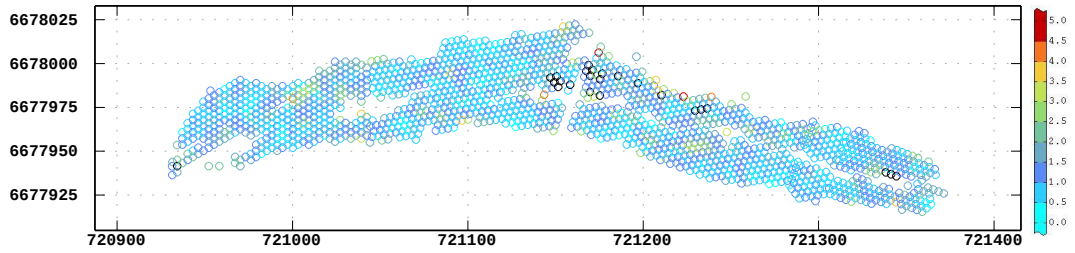
(a) Spacing 1 (b) Spacing 5

Figure 6.35: Aitchison error QQ plots - GH on abscissa, OCK on ordinate

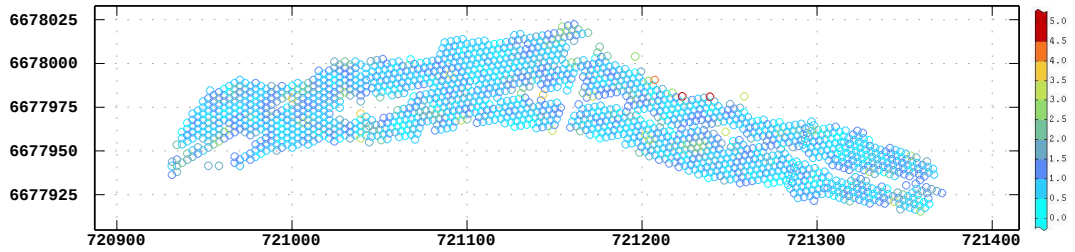


(a) Euclidean (b) Aitchison

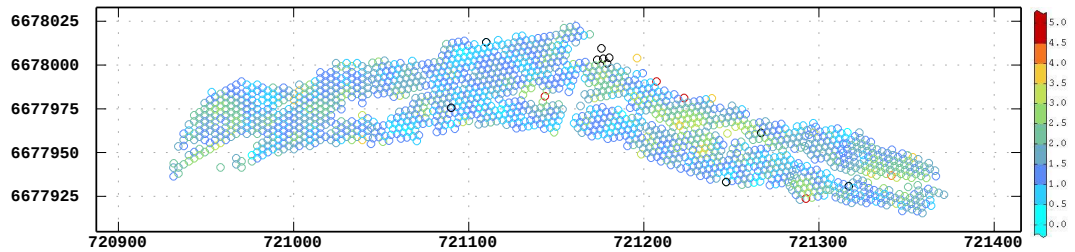
Figure 6.36: Mean compositional error as function of Spacing



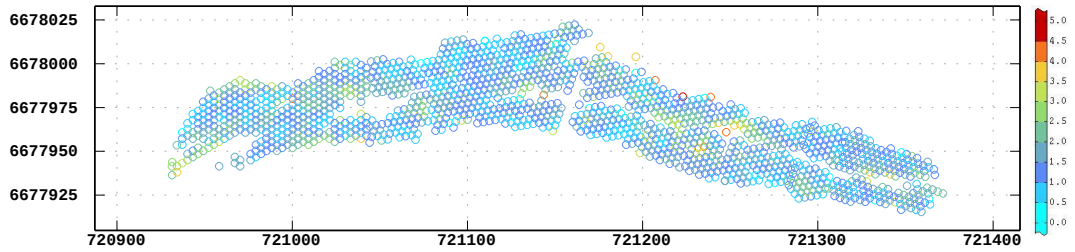
(a) Spacing 1 OK



(b) Spacing 1 GH



(c) Spacing 5 OK



(d) Spacing 5 GH

Figure 6.37: Aitchison compositional error basemaps

located on the edges of the zones where the search ellipse is sub-optimally informed and the possibility of edge dilution is elevated. This commentary is true of the exhaustive and lower data density scenarios.

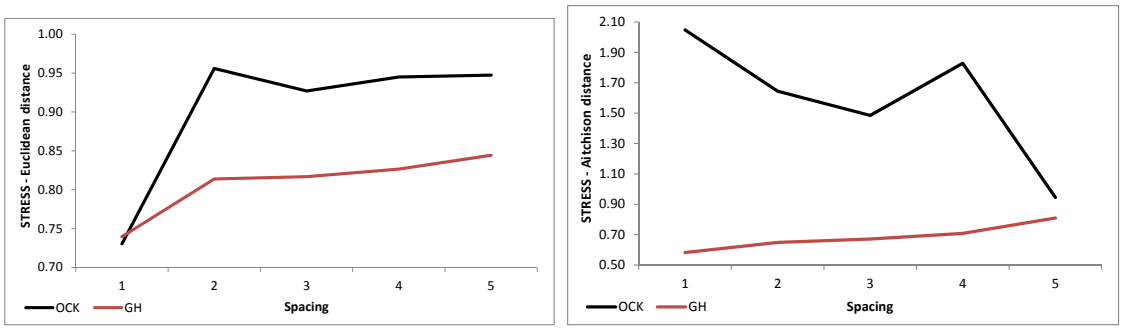
6.5 STRESS

The *STRESS* is used to measure loss of information when substituting estimates for the original observations; it is useful as it is a single number and therefore easy to compare one dataset against another [44]. Simplistically, as it is a sum of squared residual method, an unbiased, but smoothed estimate is likely to perform worse than a slightly biased estimate with variability closer to the original dataset. This is the result we note in Table

Table 6.8: STRESS as a function of spacing

Spacing	ordinary cokriging		Gauss-Hermite Quadrature	
	Euclidean	Aitchison	Euclidean	Aitchison
1	0.731	2.048	0.740	0.583
2	0.956	1.645	0.814	0.649
3	0.927	1.485	0.817	0.672
4	0.945	1.828	0.827	0.709
5	0.947	0.945	0.844	0.810

6.8 and shown graphically in Figure 6.38. With the exception of the Spacing 1 OCK estimate which (marginally) retains more of the true values information than the GH estimate when compared using the STRESS statistic and the Euclidean error measure, all other data spacings and error methods return results favoring the GH estimates.

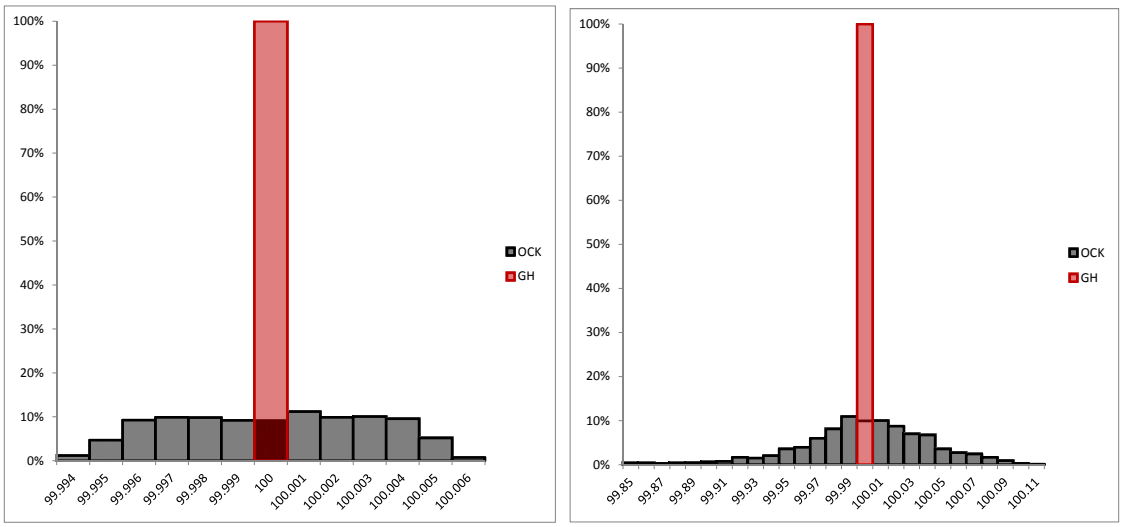


(a) Euclidean (b) Aitchison

Figure 6.38: STRESS statistic as function of Spacing

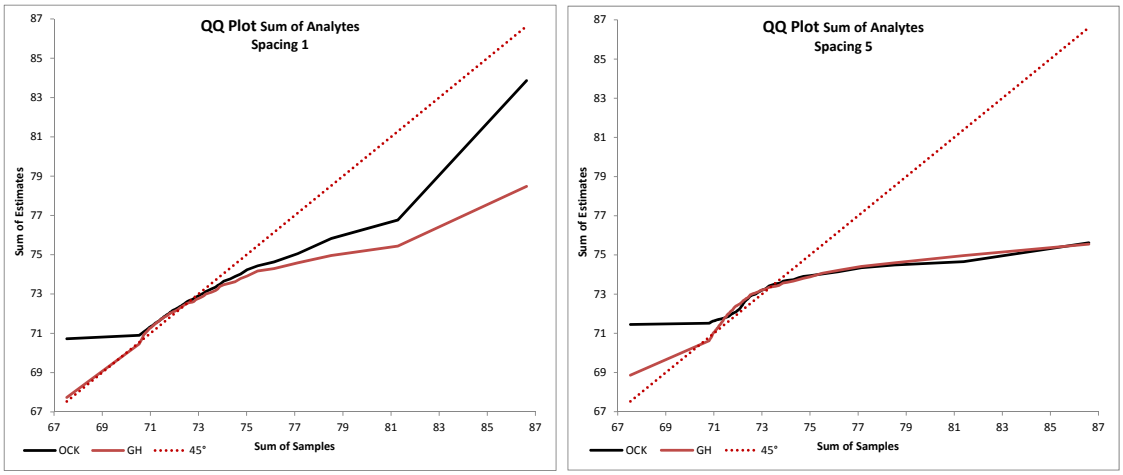
6.6 Sum of Analytes

Compositional data techniques are formulated to address the specific issue of estimating quantities which total a known and constant value. The logratio transformation and the Gauss-Hermite Quadrature back-transformation techniques used in this study result in estimates that honour this constraint by design (See Section 2.5 and 2.9). The OCK algorithm does not have this constraint and thus the resulting estimates will rarely total the target. As this is a comparative study, the $filler$ variable was calculated and estimated for both algorithms used, even though this is not common practice when using OCK. The advantage of this experimental design is that it allows comprehensive comparison of this aspect of the two techniques. The OCK and GH datasets were examined firstly by examining the histograms of the totals of all eight parts of the composition i.e. including the $filler$ and then only the seven parts measured and available in the original sample dataset. Figure 6.39 depicts the histogram totals and shows the spread of the totals for the OCK estimates juxtaposed against the single spike of the GH estimates. The Spacing 1 OCK



(a) Spacing 1 (b) Spacing 5

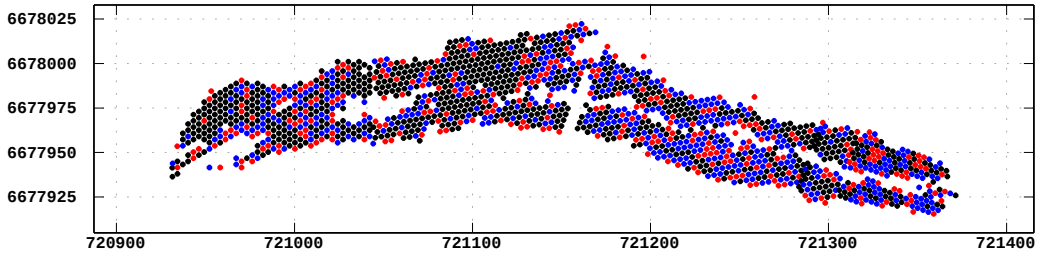
Figure 6.39: Histogram of 8 part totals



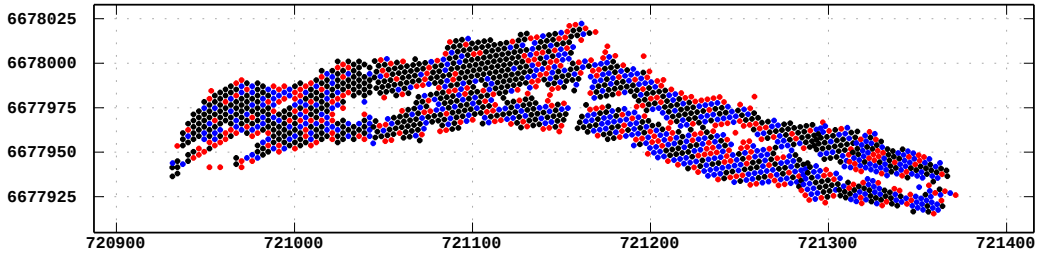
(a) Spacing 1 (b) Spacing 5

Figure 6.40: QQ plots of analyte sums

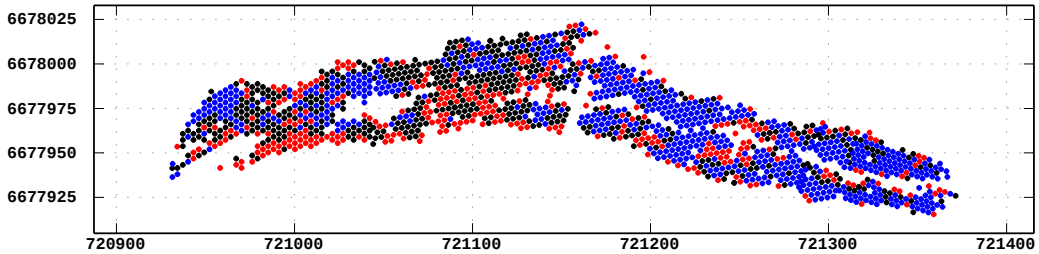
estimates range between 99.994 and 100.006 percent, small when viewed in an operational iron ore mining context. This range of values about the required 100 percent mark increases as informing data density decreases, for example the Spacing 5 dataset ranges between 99.85 and 100.11 percent, a twenty-fold increase over the range of the Spacing 1 estimates. Comparisons which exclude the *filler* variable are important in a practical sense; as the iron ore mining process focuses attention naturally on the analytes that are assayed only. To this end, the totals of the analytes and estimates excluding the filler were calculated at each data location and the results compared using two techniques. Firstly QQ plots of the OCK and GH distributions were created using the informing data distribution on the abscissa and the estimate on the ordinate axes (Figure 6.40), and secondly, the *difference* between the estimated and sample total were plotted on basemaps to ascertain if spatial patterns were evident (Figure 6.41). The spatial patterns of the differences follow very similar trends for both techniques; the differences appear distributed along



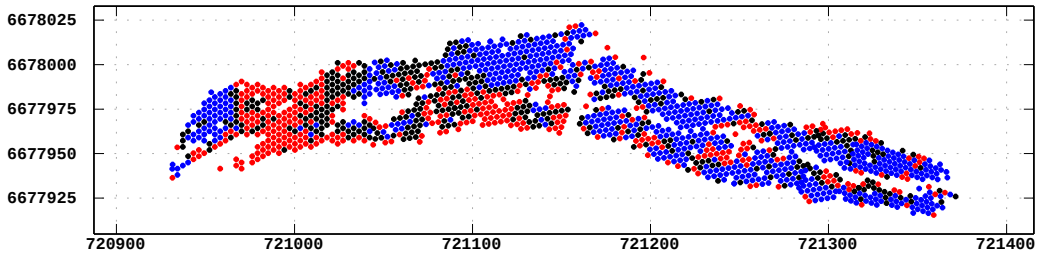
(a)



(b)



(c)



(d)

Figure 6.41: Basemaps of analyte total differences. The point colours represent the difference of the estimated analyte totals from the sample analyte total. Black - estimates total within 1 weight percent of the sample analyte totals, Blue - estimates total are less than the sample by more than half a weight percent, Red - estimates total are greater than the sample by more than half a weight percent. (a) Spacing 1 - OCK
, (b) Spacing 1 - GH, (c) Spacing 5 - OCK, (d) Spacing 5 - GH

the edges of the sampled areas; which fits with the intuitive notion that this is where the errors are expected due to edge effect and the position of errors noted with other measures or techniques (See Sections 6.3 and 6.4). The similarities in the patterns suggests that this is data driven rather than model or method driven.

Chapter 7

Conclusions

The study describes the comparison of geostatistical estimates at data locations of an exhaustive set of compositional geochemical data with the true sampled value. The body of work seeks to honour the *constant sum* constraint implicit to the whole rock geochemical analyses used as the base dataset through the application of geostatistical methods specifically kriging, a technique which ordinarily does not honour this constraint. Central to the properties of compositional geochemical data and therefore the geostatistical techniques are the correlations, spurious and real, which are a feature of the largest seaborne commodity, iron ore, which is the focus of this study. Honouring these properties is achieved through the application of logratio transformations *prior* to the application of the geostatistical methods followed by a numerical integration step (i.e. GH) to restore the resulting estimates to the compositional sample space. The results of the logratio process flow are measured against those derived from the conventional implementation of linear geostatistical methods (i.e. OCK) on compositional data. The results are compared using three error measures, one univariate and two compositional, as well as two other criteria which gauge the holistic performance of each method, namely the local accuracy and precision. Peer reviewed results of a pilot study were presented at two international conferences, where other practitioners suggested that the high spatial density of the exhaustive dataset may mask potential bias induced by non-linear logratio transformation. This concern is tested in this study through successive sub-sampling of the exhaustive dataset and using the geostatistical implementation of jack-knifing. The sub-sampling is designed to mimic the data density, spacing and orientation of information likely available at various stages of mining project evaluations.

7.1 Significant findings

Exploratory data analysis

The exhaustive dataset is a horizontal section through a steeply dipping BIF horizon which displays minor structural imbrication. Examining the descriptive statistics of the constrained sample data generates results typical of ore grade BIF where:

- the Fe component dominates,
- Fe, SiO_2, and Al_2O_3 are (potentially spuriously) highly correlated, and
- all the distributions with the exception of LOI are highly skewed.

Edge dilution effects are noted within the data, these are explained by the inability of six meter vertical blast hole samples to accurately delineate the BIF to country rock (ore/waste) interface. The subsets are evaluated for representativity of the exhaustive set through mean bias quantification and QQ plot visualisation. The mean bias results appear to be excessive for some analytes in specific subsets but combined examination of the statistics with the QQ plots showed that these were caused by few samples in the tails of the distribution and were not material. The subsets are considered to be representative and thus meaningful conclusions can be drawn from comparisons with the exhaustive dataset. The exhaustive data are also explored through the application of compositional techniques. Similarly to the constrained techniques, these also comprise numerical and visual elements from which the following items are noted:

- the only highly correlated logratio transformed variable set are SiO_2 and Al_2O_3, all other correlations are spurious,
- Mn when paired with SiO_2, Al_2O_3, and S explains the largest portion of the variability contained within the data,
- the centered ternary diagrams exhibit a continuous shape without different groupings which is interpreted to be indicative of a stationary domain.

The most important conclusion drawn from the exploratory data analysis steps is the decision that the assumption that the data are stationary is reasonable and therefore that the application of geostatistical methods is both permissible and the results likely robust. The interpretation that Mn is the key variable for the variability of the data is an important second conclusion. The spatial covariances are estimated and modeled using conventional techniques; a primary advantage of the alr method. The approach followed is designed to eliminate differences in *interpreted* modelling decisions typical of natural resource estimation studies; decisions which would arise as a function of the spatial density changes. The cokriging estimation parameters are optimised on the Spacing 1 data and the constrained sample space data and then held constant to eliminate another source

of possible variability. Using these parameters, datasets, and spatial covariance models the cross-validation and jack-knifing routines were run and validated. The alr compositional estimates are back-transformed to the constrained sample space using numerical integration via Gauss-Hermite quadrature and validated.

Analysis of results

The highest density dataset is reproduced approximately equally well by either interpolation method when the univariate mean error is considered however the GH method better reproduces the tails of the distributions. This is attributed to the approximately Gaussian form of the alr transformed histograms which will be better reproduced when cokriging is the estimation technique. The GH methodology also better reproduces the compositional bounding inherent to this lithology as noted in the scatterplots, an expected but significant result. The accuracy and precision testing is particularly interesting as the GH-derived estimates and confidence limits were in almost all cases superior to the OCK -derived results, for the high and lower data densities. In the majority of cases the goodness statistic was higher for GH than for OCK. The compositional error measures favour the compositional approach, an expected result, both for the Euclidean and Aitchison measures. With decreasing data density, the results in general increasingly support the use of the GH methodology over the conventional OCK. This statement is supported by the following points:

1. there exists bimodality in the OCK distributions not present in the true value distributions,
2. the mean univariate error statistic is similar, however the GH histogram is generally superior as it is more symmetrical and centred on zero,
3. the values of the GH correlation statistics are more similar to the true values than OCK,
4. the morphology of the GH scatterplots more closely represents the true values than the OCK,
5. the accuracy and precision curves are generally superior for the GH distributions,
6. the mean compositional errors are smaller in nine out of ten cases for the GH estimates,
7. the STRESS statistic indicates that less information is lost when using the GH methodology, and
8. the *total* sum of analytes is only reproduced perfectly by the GH technique.

Experience by other practitioners is that compositional measures will demonstrate superior performance for compositional estimation methods over methods which ignore the compositional nature of the input data (*pers.comm.* Tolosana-Delgado, 2012) and this

study generally supports that assertion. The performance of the two techniques compared here are similar for the highest density data with the exception of the total sums. However, data densities less than the exhaustive favour the GH method. The reasons could be derived from a number of items, the most likely being that the pseudo-Gaussian distribution of the alr transformed variables will be better reproduced by the Gaussian-optimal cokriging algorithm than the skewed distributions of the analytes in the constrained data space. It is hypothesised that as the data density decreases, the pseudo-Gaussian alr distribution is less likely to decay than the skewed untransformed data distributions. The accuracy and precision study in particular adds weight to the argument for the GH method. In essence, a practitioner using the alr-GH method would decrease the mine planning risk through more accurate and generally precise estimates with quantifiable confidence limits over the conventional OCK methodology, especially at the low data densities typical of early stage mineral project evaluation.

7.2 Final remarks

The application to production iron ore blast hole cuttings of the logratio method of data transformation coupled with the Gauss-Hermite Quadrature method of back-transforming the data from $\mathbb{R}^D$ to the D-simplex S^D produces acceptable results. The method is comparable with conventional methods at high data densities and superior at lower data densities. Practitioners at early stages of iron ore project evaluation would benefit from using the techniques and as such has been shown to be a viable alternative in the geostatistics toolbox.

Appendix A

Jack knife: Informing comparative statistics

Table A.1: Descriptive statistics - Cross Validation sample data

Variable	Min	Q25	Mean	Q50	Q75	Std. Dev.	Skew
Spacing 2 (n = 432)							
Al_2O_3	0.120	0.58	1.790	1.00	2.11	2.27	3.72
Fe	10.100	62.07	62.495	64.29	65.48	7.27	-5.19
LOI	0.670	3.01	4.026	3.91	4.87	1.41	0.72
Mn	0.003	0.05	0.187	0.09	0.16	0.45	8.07
P	0.016	0.09	0.128	0.12	0.16	0.05	0.74
S	0.003	0.01	0.031	0.02	0.03	0.07	9.19
SiO_2	0.440	1.10	3.843	1.85	3.76	7.57	5.84
$Filler$	13.368	27.23	27.499	28.17	28.59	2.26	-3.88
Spacing 3 (n = 234)							
Al_2O_3	0.170	0.71	1.993	1.15	2.21	2.36	2.62
Fe	10.240	61.72	62.070	64.14	65.30	7.71	-4.74
LOI	0.540	2.98	4.070	3.87	5.02	1.54	0.88
Mn	0.005	0.05	0.196	0.08	0.16	0.47	7.67
P	0.022	0.09	0.128	0.12	0.16	0.06	0.98
S	0.003	0.01	0.027	0.02	0.03	0.03	5.08
SiO_2	0.340	1.21	4.112	2.04	3.85	7.86	5.60
$Filler$	13.368	27.18	27.403	28.10	28.58	2.37	-3.25
Spacing 4 (n = 137)							
Al_2O_3	0.250	0.70	2.167	1.25	2.56	2.39	2.54
Fe	10.240	61.30	61.728	63.57	65.48	8.15	-4.70
LOI	0.770	3.12	4.169	4.08	5.04	1.35	0.34
Mn	0.004	0.05	0.159	0.07	0.14	0.43	9.30
P	0.022	0.10	0.131	0.13	0.17	0.05	0.39
S	0.003	0.01	0.033	0.02	0.03	0.05	5.76
SiO_2	0.470	1.21	4.306	2.38	4.69	8.21	5.55
$Filler$	13.368	26.84	27.307	27.90	28.61	2.25	-3.61
Spacing 5 (n = 71)							
Al_2O_3	0.150	0.62	1.995	1.07	2.50	2.42	2.69
Fe	13.650	61.63	61.906	64.18	65.59	8.72	-4.45
LOI	0.540	3.15	3.977	4.03	4.86	1.22	-0.16
Mn	0.008	0.05	0.159	0.08	0.15	0.24	3.91
P	0.022	0.09	0.126	0.12	0.17	0.05	0.24
S	0.003	0.01	0.028	0.02	0.03	0.04	4.77
SiO_2	0.540	1.09	4.403	1.94	3.63	8.88	4.89
$Filler$	13.368	27.22	27.406	28.18	28.68	2.38	-3.50

Table A.2: Correlation matrices - Cross Validation sample data

Spacing 2 (n = 432)	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2
Fe	-0.85						
LOI	0.18	-0.11					
Mn	0.04	-0.06	0.16				
P	-0.20	0.24	0.29	0.07			
S	0.30	-0.14	0.28	-0.03	0.06		
SiO_2	0.73	-0.96	-0.09	-0.04	-0.31	0.04	
$Filler$	-0.85	0.95	-0.19	-0.02	0.25	-0.18	-0.92
Spacing 3 (n = 234)							
Fe	-0.88						
LOI	0.18	-0.14					
Mn	-0.01	-0.03	0.11				
P	-0.23	0.23	0.35	0.07			
S	0.18	-0.06	0.22	0.02	0.12		
SiO_2	0.80	-0.97	-0.04	-0.06	-0.31	-0.02	
$Filler$	-0.89	0.94	-0.28	0.02	0.23	-0.09	-0.91
Spacing 4 (n = 137)							
Fe	-0.90						
LOI	0.16	-0.09					
Mn	-0.05	-0.03	0.20				
P	-0.23	0.31	0.07	0.03			
S	0.34	-0.17	0.37	0.05	-0.03		
SiO_2	0.83	-0.98	-0.09	-0.05	-0.34	0.05	
$Filler$	-0.91	0.95	-0.19	0.01	0.30	-0.21	-0.92
Spacing 5 (n = 71)							
Fe	-0.91						
LOI	-0.02	0.01					
Mn	-0.15	0.08	0.06				
P	-0.19	0.24	0.45	0.08			
S	0.12	-0.01	0.13	-0.13	0.16		
SiO_2	0.87	-0.99	-0.13	-0.11	-0.31	-0.04	
$Filler$	-0.91	0.92	-0.08	0.13	0.21	-0.02	-0.92

Table A.3: Descriptive statistics - Jack knife estimate location sample data

Variable	Min	Q25	Mean	Q50	Q75	Std. Dev.	Skew
Spacing 2 (n = 1168)							
Al_2O_3	0.080	0.53	1.594	0.95	1.87	1.86	3.36
Fe	9.940	62.24	63.097	64.42	65.69	5.09	-4.82
LOI	0.520	3.14	4.099	3.98	4.93	1.41	0.59
Mn	0.004	0.05	0.188	0.08	0.15	0.58	13.26
P	0.017	0.09	0.127	0.12	0.16	0.05	1.02
S	0.001	0.01	0.029	0.02	0.03	0.05	7.08
SiO_2	0.340	1.05	3.200	1.75	3.31	5.15	6.38
$Filler$	13.444	27.28	27.666	28.17	28.67	1.79	-3.28
Spacing 3 (n = 1366)							
Al_2O_3	0.080	0.52	1.588	0.93	1.86	1.90	3.81
Fe	9.940	62.26	63.082	64.44	65.68	5.35	-5.22
LOI	0.520	3.11	4.081	3.98	4.89	1.38	0.56
Mn	0.003	0.05	0.186	0.09	0.15	0.56	13.07
P	0.016	0.09	0.127	0.12	0.16	0.05	0.92
S	0.001	0.01	0.030	0.02	0.03	0.06	8.72
SiO_2	0.380	1.04	3.247	1.76	3.30	5.50	6.52
$Filler$	13.397	27.29	27.658	28.18	28.66	1.84	-3.66
Spacing 4 (n = 1463)							
Al_2O_3	0.080	0.53	1.598	0.94	1.85	1.93	3.72
Fe	9.940	62.36	63.047	64.45	65.67	5.48	-5.21
LOI	0.520	3.08	4.071	3.95	4.90	1.41	0.65
Mn	0.003	0.05	0.191	0.09	0.15	0.56	12.71
P	0.016	0.09	0.127	0.12	0.16	0.05	0.99
S	0.001	0.01	0.029	0.02	0.03	0.05	9.22
SiO_2	0.340	1.06	3.286	1.74	3.34	5.64	6.47
$Filler$	13.397	27.35	27.650	28.18	28.64	1.89	-3.60
Spacing 5 (n = 1529)							
Al_2O_3	0.080	0.55	1.631	0.95	1.94	1.95	3.62
Fe	9.940	62.23	62.982	64.40	65.66	5.59	-5.25
LOI	0.520	3.08	4.084	3.96	4.91	1.42	0.64
Mn	0.003	0.05	0.189	0.09	0.15	0.56	12.50
P	0.016	0.09	0.128	0.12	0.16	0.05	0.97
S	0.001	0.01	0.030	0.02	0.03	0.06	8.97
SiO_2	0.340	1.06	3.325	1.76	3.37	5.73	6.53
$Filler$	13.397	27.26	27.631	28.17	28.64	1.90	-3.61

Table A.4: Correlation matrices - Jack knife estimate location sample data

Spacing 2 (n = 1168)	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2
Fe	-0.80						
LOI	0.32	-0.34					
Mn	0.02	-0.12	0.18				
P	-0.06	0.10	0.30	0.07			
S	0.32	-0.21	0.28	0.01	0.11		
SiO_2	0.61	-0.92	0.05	-0.04	-0.22	0.08	
$Filler$	-0.78	0.95	-0.35	-0.04	0.13	-0.20	-0.91
Spacing 3 (n = 1366)							
Fe	-0.80						
LOI	0.30	-0.29					
Mn	0.04	-0.12	0.19				
P	-0.08	0.13	0.28	0.07			
S	0.34	-0.21	0.29	-0.01	0.09		
SiO_2	0.62	-0.93	0.01	-0.03	-0.23	0.08	
$Filler$	-0.78	0.95	-0.31	-0.04	0.15	-0.21	-0.91
Spacing 4 (n = 1463)							
Fe	-0.81						
LOI	0.29	-0.28					
Mn	0.04	-0.11	0.18				
P	-0.10	0.13	0.31	0.08			
S	0.30	-0.18	0.27	-0.01	0.10		
SiO_2	0.63	-0.93	0.01	-0.04	-0.24	0.06	
$Filler$	-0.79	0.95	-0.31	-0.04	0.15	-0.19	-0.91
Spacing 5 (n = 1529)							
Fe	-0.81						
LOI	0.29	-0.28					
Mn	0.03	-0.11	0.18				
P	-0.10	0.14	0.29	0.07			
S	0.32	-0.19	0.28	0.00	0.09		
SiO_2	0.64	-0.94	0.01	-0.04	-0.25	0.07	
$Filler$	-0.80	0.95	-0.31	-0.04	0.16	-0.20	-0.91

Appendix B

Descriptive statistics - OCK estimates

Table B.1: OCK estimates - descriptive statistics

Variable	Min	Q25	Mean	Q50	Q75	Std. Dev.	Skew
Spacing 2 (n = 1168)							
Al_2O_3	0.299	0.917	1.48	1.334	1.80	0.84	2.93
Fe	26.328	62.475	63.39	63.873	64.89	2.58	-5.02
LOI	1.547	3.518	4.09	4.038	4.60	0.75	0.20
Mn	-0.019	0.074	0.19	0.150	0.26	0.14	1.18
P	0.014	0.102	0.13	0.122	0.15	0.04	0.91
S	-0.008	0.015	0.03	0.021	0.03	0.04	5.38
SiO_2	-4.192	1.468	2.95	2.265	3.43	2.88	5.20
$Filler$	17.8	27.431	27.75	27.928	28.31	0.91	-3.09
Spacing 3 (n = 1365)							
Al_2O_3	0.184	1.052	1.63	1.456	1.99	0.87	1.66
Fe	38.858	62.370	63.12	63.638	64.62	2.25	-2.29
LOI	1.919	3.691	4.16	4.117	4.64	0.71	0.12
Mn	-0.222	0.075	0.20	0.177	0.25	0.16	1.64
P	0.041	0.101	0.13	0.125	0.16	0.04	0.95
S	0.003	0.019	0.03	0.023	0.03	0.01	1.41
SiO_2	-0.374	1.470	3.02	2.405	3.73	2.41	2.27
$Filler$	23.077	27.392	27.72	27.919	28.31	0.89	-1.43
Spacing 4 (n = 1463)							
Al_2O_3	0.78	1.308	1.80	1.658	2.12	0.75	2.81
Fe	32.732	62.015	62.92	63.268	64.07	2.48	-4.44
LOI	2.687	3.871	4.21	4.190	4.51	0.50	0.19
Mn	-0.523	0.070	0.16	0.134	0.22	0.12	2.06
P	0.019	0.104	0.13	0.129	0.15	0.03	0.17
S	-0.002	0.020	0.03	0.028	0.04	0.02	3.48
SiO_2	-6.82	2.060	3.15	2.764	3.83	2.50	5.46
$Filler$	19.97	27.287	27.61	27.721	28.10	0.74	-3.21
Spacing 5 (n = 1529)							
Al_2O_3	0.849	1.164	1.801	1.689	2.297	0.698	1.086
Fe	38.622	61.082	62.49	62.754	64.51	2.18	-1.50
LOI	2.698	3.705	3.98	4.016	4.25	0.34	-0.29
Mn	0.001	0.097	0.17	0.194	0.25	0.08	-0.09
P	0.073	0.112	0.13	0.123	0.14	0.02	0.75
S	-0.02	0.017	0.03	0.019	0.03	0.02	2.88
SiO_2	0.629	1.924	3.83	3.483	5.12	2.20	1.50
$Filler$	24.387	26.997	27.57	27.923	28.24	0.82	-1.00

Table B.2: Correlation matrices - OCK estimates

Spacing 1 (n = 1594)	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2
Fe	-0.85						
LOI	0.09	-0.16					
Mn	0.02	0.00	0.21				
P	-0.10	0.20	0.19	0.09			
S	0.30	-0.13	0.19	-0.06	0.30		
SiO_2	0.72	-0.95	-0.08	-0.13	-0.30	0.01	
$Filler$	-0.80	0.97	-0.17	0.07	0.25	-0.10	-0.94
Spacing 2 (n = 1168)							
Fe	-0.79						
LOI	0.04	-0.13					
Mn	-0.07	0.11	0.21				
P	-0.14	0.24	0.10	-0.01			
S	0.39	-0.16	0.16	-0.21	0.25		
SiO_2	0.63	-0.94	-0.13	-0.24	-0.31	0.03	
$Filler$	-0.73	0.95	-0.13	0.19	0.30	-0.14	-0.93
Spacing 3 (n = 1365)							
Fe	-0.92						
LOI	0.02	-0.09					
Mn	-0.31	0.34	-0.27				
P	-0.27	0.30	0.04	-0.03			
S	0.15	-0.01	0.07	-0.33	0.31		
SiO_2	0.86	-0.96	-0.13	-0.34	-0.34	-0.06	
$Filler$	-0.90	0.96	-0.19	0.41	0.35	0.00	-0.92
Spacing 4 (n = 1463)							
Fe	-0.90						
LOI	0.12	-0.11					
Mn	-0.06	0.03	0.29				
P	-0.11	0.22	-0.33	0.02			
S	0.21	-0.13	0.06	-0.14	0.19		
SiO_2	0.85	-0.97	-0.09	-0.15	-0.19	0.08	
$Filler$	-0.94	0.93	-0.19	0.11	0.23	-0.10	-0.90
Spacing 5 (n = 1529)							
Fe	-0.98						
LOI	0.45	-0.40					
Mn	-0.76	0.71	-0.81				
P	-0.07	0.06	0.31	-0.22			
S	0.29	-0.29	0.23	-0.34	0.62		
SiO_2	0.97	-0.99	0.34	-0.68	-0.11	0.24	
$Filler$	-0.96	0.93	-0.59	0.85	0.05	-0.22	-0.93

Appendix C

OCK estimates - QQ plots

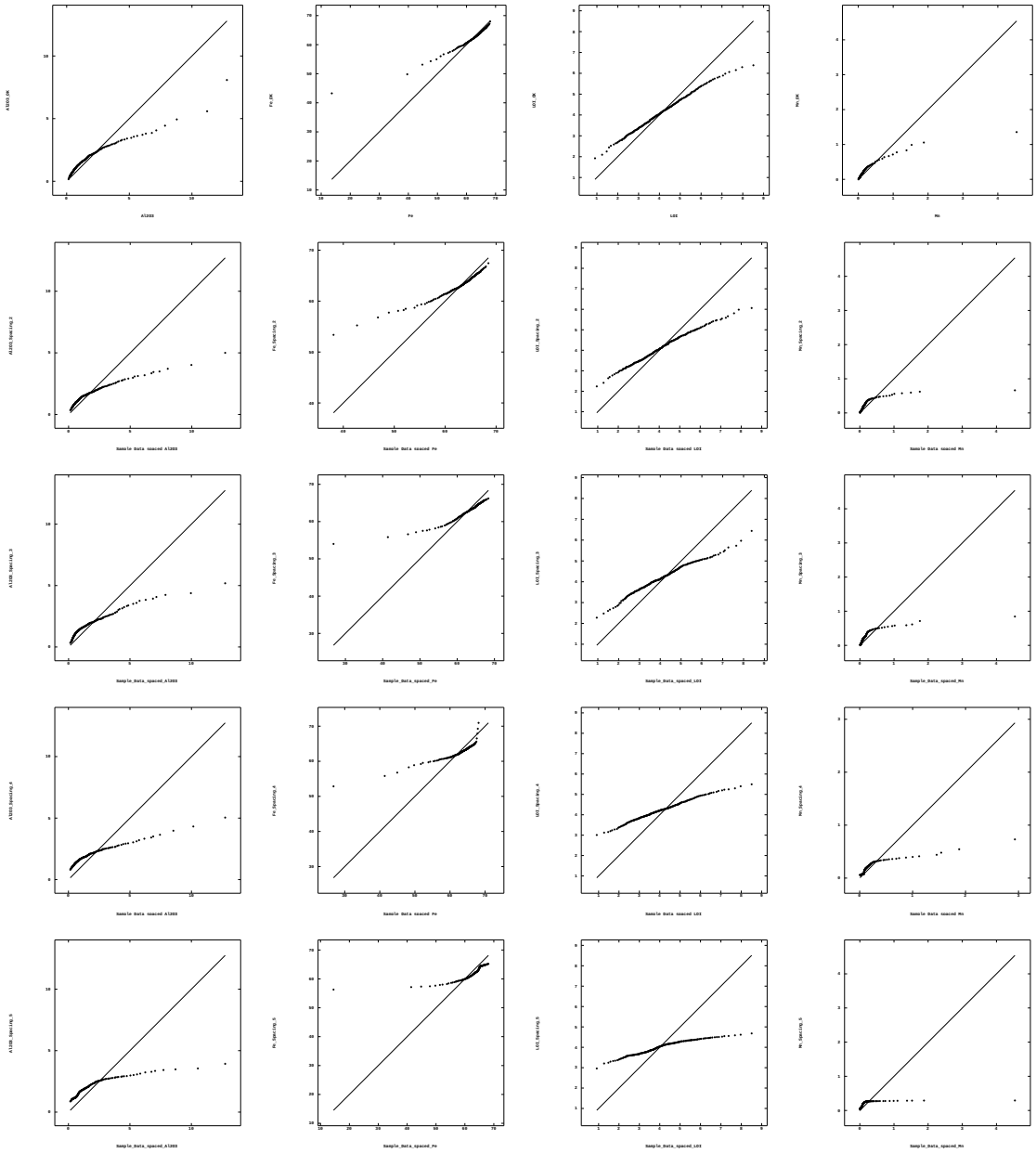


Figure C.1: QQ plots - By column Al_2O_3 far left, *Fe* near left, *LOI* near right, and *Mn* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the sample data, the ordinate is the OCK distribution.

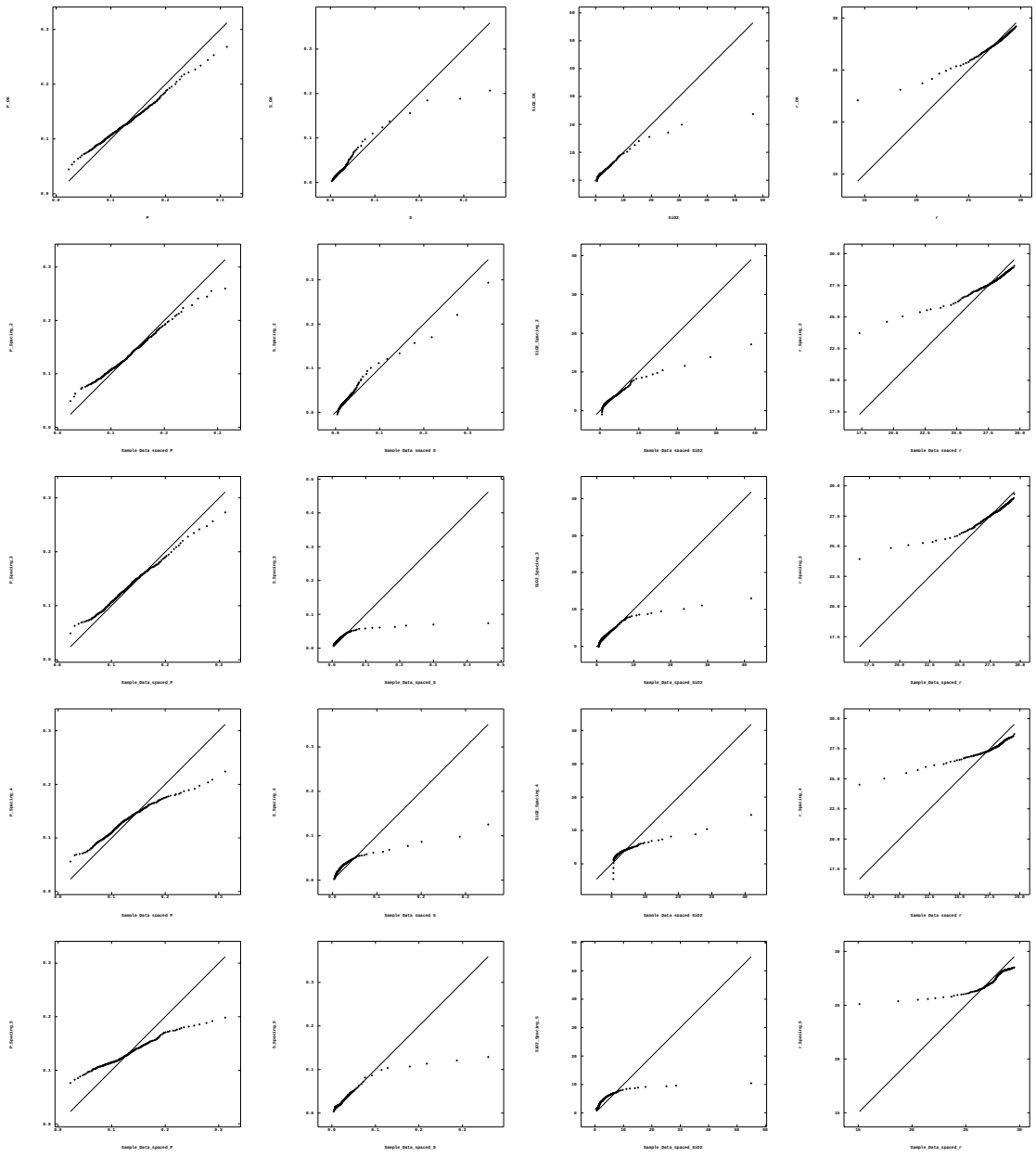


Figure C.2: OCK QQ plots - By column P far left, S near left, SiO_2 near right, and $filler$ far right in order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the sample data, the ordinate is the OCK distribution.

Appendix D

GH estimates - Descriptive statistics

Table D.1: GH estimates - descriptive statistics

Variable	Min	Q25	Mean	Q50	Q75	Std. Dev.	Skew
Spacing 2 (n = 1173)							
Al_2O_3	0.31	0.915	1.48	1.284	1.82	0.86	3.72
Fe	21.875	62.558	63.45	63.982	64.91	2.44	-6.44
LOI	1.541	3.520	4.02	3.968	4.51	0.67	0.03
Mn	0.006	0.080	0.16	0.135	0.20	0.10	1.53
P	0.039	0.102	0.13	0.123	0.15	0.05	3.46
S	0.003	0.017	0.03	0.022	0.03	0.03	6.40
SiO_2	0.59	1.586	2.93	2.278	3.40	2.42	5.04
$Filler$	22.139	27.437	27.80	27.961	28.38	0.80	-1.54
Spacing 3 (n = 1370)							
Al_2O_3	0.438	1.044	1.66	1.497	2.06	0.86	3.16
Fe	20.264	62.254	63.15	63.700	64.65	2.41	-5.29
LOI	1.184	3.622	4.10	4.050	4.56	0.77	0.07
Mn	0.012	0.073	0.16	0.131	0.22	0.12	1.82
P	0.044	0.104	0.13	0.126	0.15	0.04	0.94
S	0.009	0.018	0.03	0.022	0.03	0.02	5.67
SiO_2	0.818	1.693	3.02	2.311	3.66	2.20	4.67
$Filler$	22.206	27.337	27.75	27.887	28.34	0.96	-0.45
Spacing 4 (n = 1468)							
Al_2O_3	0.548	1.197	1.85	1.652	2.32	0.94	2.46
Fe	27.382	61.727	62.79	63.166	64.32	2.57	-5.24
LOI	2.472	3.773	4.11	4.084	4.36	0.48	0.39
Mn	0.014	0.068	0.14	0.108	0.19	0.09	1.76
P	0.036	0.100	0.13	0.127	0.16	0.04	1.21
S	0.004	0.019	0.03	0.025	0.03	0.02	2.55
SiO_2	0.953	2.060	3.34	2.968	4.22	1.99	4.52
$Filler$	24.726	27.079	27.61	27.642	28.22	0.77	-0.12
Spacing 5 (n = 1534)							
Al_2O_3	0.196	1.093	1.83	1.712	2.33	0.98	1.68
Fe	30.083	61.496	62.90	62.899	64.40	3.00	-1.42
LOI	0.366	3.539	3.99	4.223	4.40	0.99	-0.10
Mn	0.01	0.075	0.16	0.106	0.23	0.12	1.18
P	0.038	0.103	0.13	0.128	0.16	0.04	0.42
S	0.003	0.016	0.03	0.020	0.03	0.02	4.60
SiO_2	0.251	2.002	3.36	2.895	4.14	2.07	2.39
$Filler$	24.451	26.921	27.60	27.659	28.30	0.97	-0.30

Table D.2: GH estimates - correlation matrices

Spacing 1 (n = 1594)	Al_2O_3	Fe	LOI	Mn	P	S	SiO_2
Fe	-0.87						
LOI	0.15	-0.26					
Mn	-0.05	0.02	0.29				
P	-0.13	0.25	0.05	0.07			
S	0.31	-0.19	0.31	-0.09	0.15		
SiO_2	0.71	-0.92	0.01	-0.18	-0.35	0.06	
$Filler$	-0.56	0.68	-0.35	0.11	0.29	-0.22	-0.76
Spacing 2 (n = 1173)							
Fe	-0.85						
LOI	0.08	-0.17					
Mn	-0.15	0.12	0.25				
P	-0.03	0.19	-0.20	0.05			
S	0.35	-0.22	0.44	-0.20	0.17		
SiO_2	0.71	-0.92	-0.10	-0.26	-0.24	0.05	
$Filler$	-0.69	0.78	-0.13	0.25	0.27	-0.24	-0.85
Spacing 3 (n = 1370)							
Fe	-0.89						
LOI	0.13	-0.29					
Mn	-0.28	0.18	0.26				
P	-0.27	0.32	-0.06	0.05			
S	0.29	-0.16	0.26	-0.20	0.19		
SiO_2	0.86	-0.93	0.07	-0.33	-0.43	0.04	
$Filler$	-0.69	0.62	-0.40	0.23	0.42	-0.15	-0.72
Spacing 4 (n = 1468)							
Fe	-0.93						
LOI	0.42	-0.43					
Mn	-0.15	0.09	0.30				
P	-0.06	0.17	-0.38	-0.15			
S	0.38	-0.31	0.76	0.01	-0.14		
SiO_2	0.88	-0.94	0.25	-0.27	-0.27	0.15	
$Filler$	-0.61	0.46	-0.35	0.28	0.39	-0.31	-0.63
Spacing 5 (n = 1534)							
Fe	-0.92						
LOI	0.38	-0.59					
Mn	-0.25	-0.03	0.43				
P	0.13	-0.12	0.52	0.10			
S	0.42	-0.28	0.30	-0.27	0.36		
SiO_2	0.93	-0.91	0.34	-0.26	-0.09	0.21	
$Filler$	-0.52	0.38	-0.41	0.35	-0.16	-0.32	-0.59

Appendix E

GH estimates - QQ plots

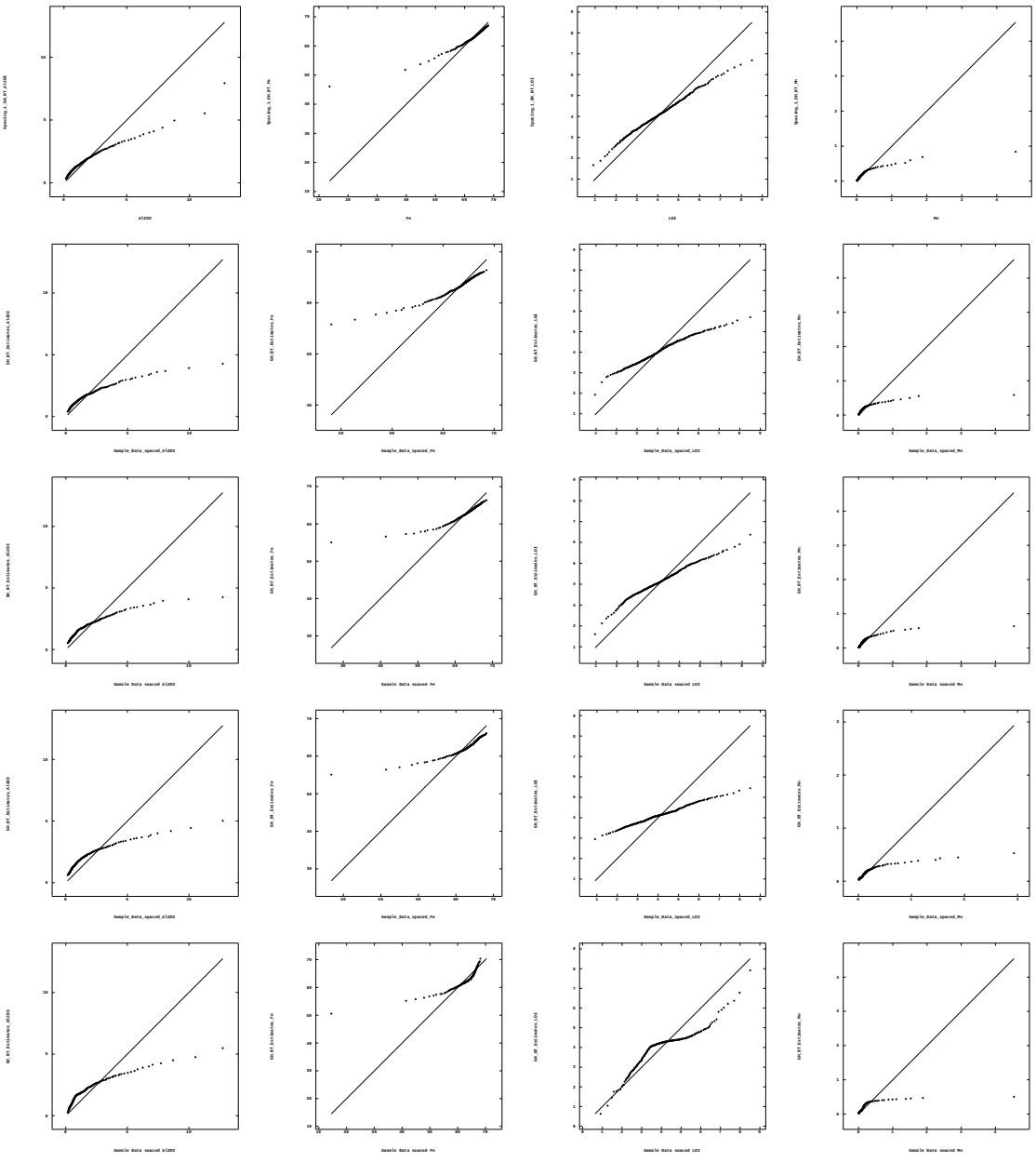


Figure E.1: QQ plots - By column Al_2O_3 far left, *Fe* near left, *LOI* near right, and *Mn* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the sample data, the ordinate is the GH distribution.

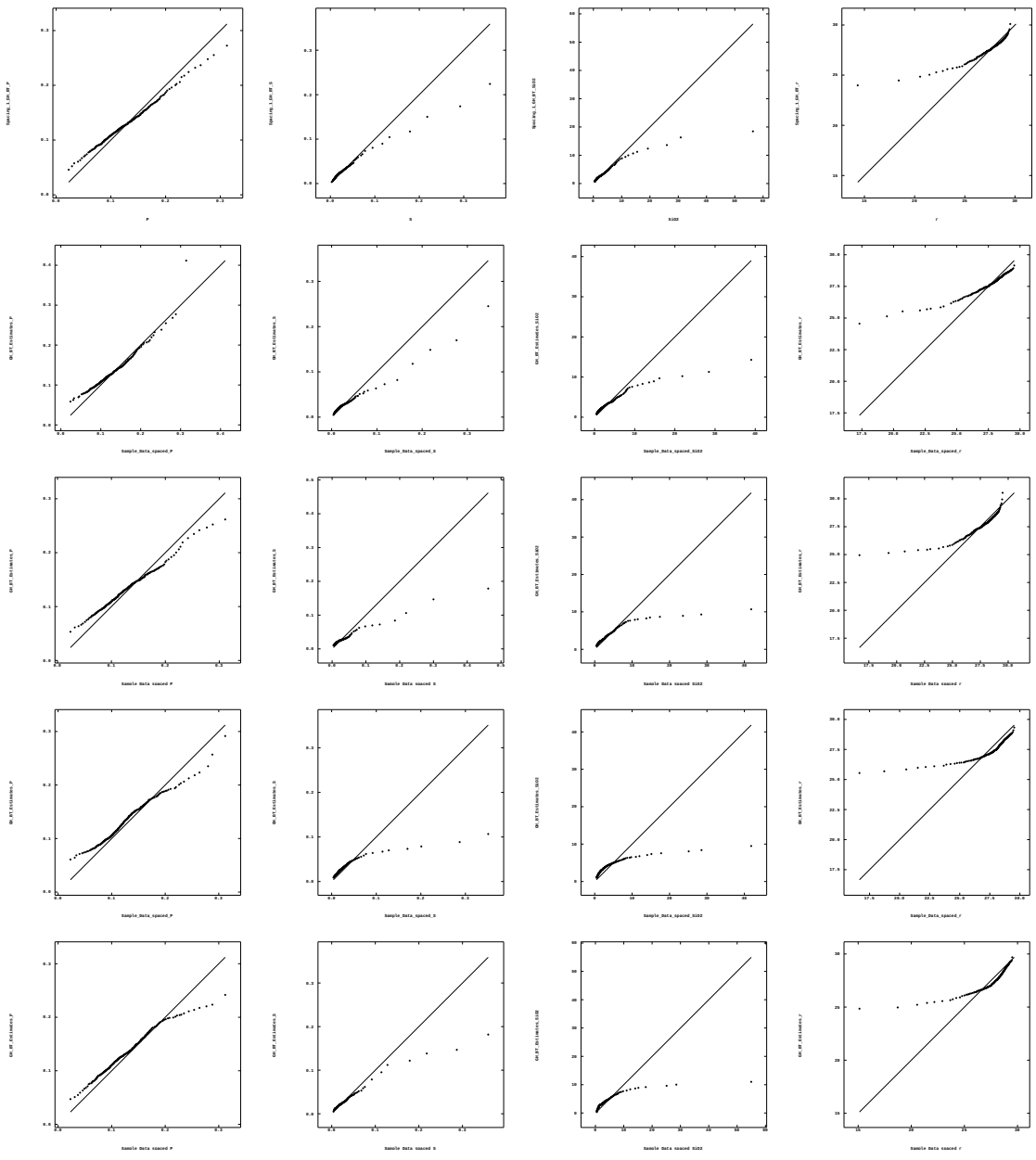


Figure E.2: GH QQ plots - By column *P* far left, *S* near left, *V* near right, and *filler* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the sample data, the ordinate is the GH distribution.

Appendix F

Scatter plots - All datasets — Constrained sample space

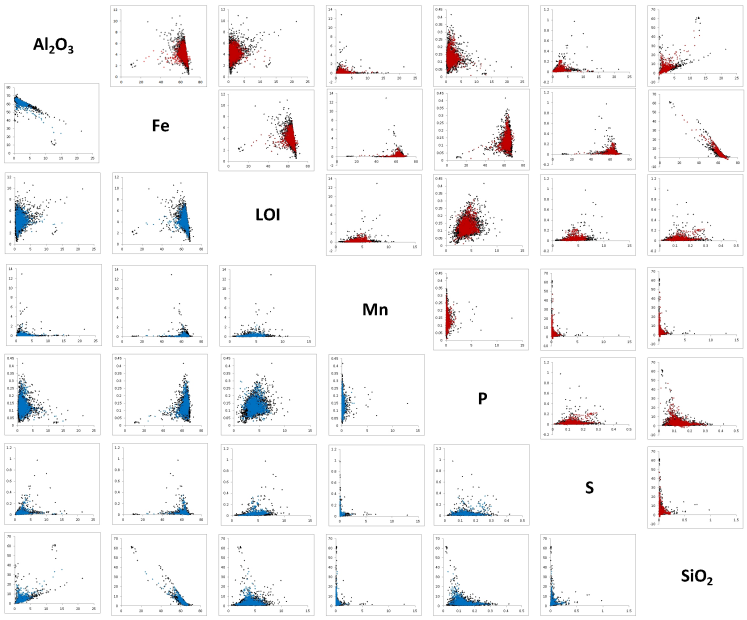


Figure F.1: Scatter plot matrix - Spacing 1. Black points are sample data, red points are OCK estimates, blue points are GH estimates

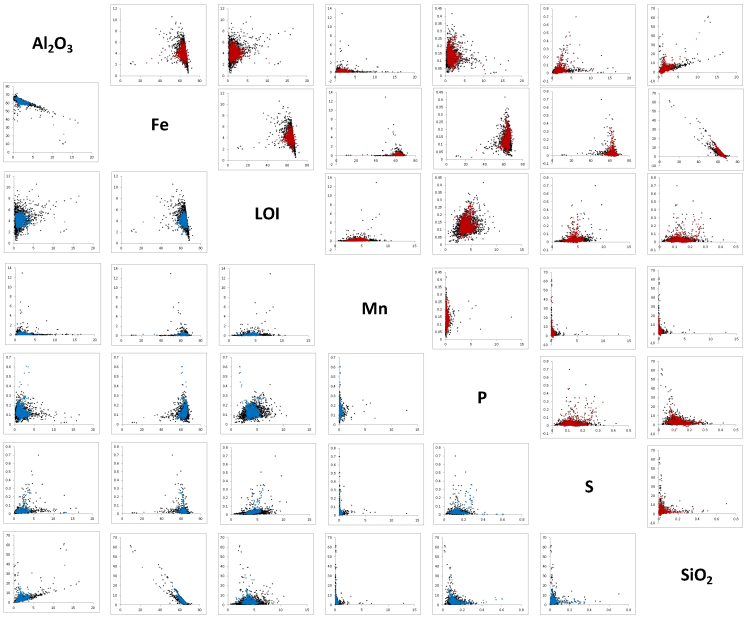


Figure F.2: Scatter plot matrix - Spacing 2. Black points are sample data, red points are OCK estimates, blue points are GH estimates

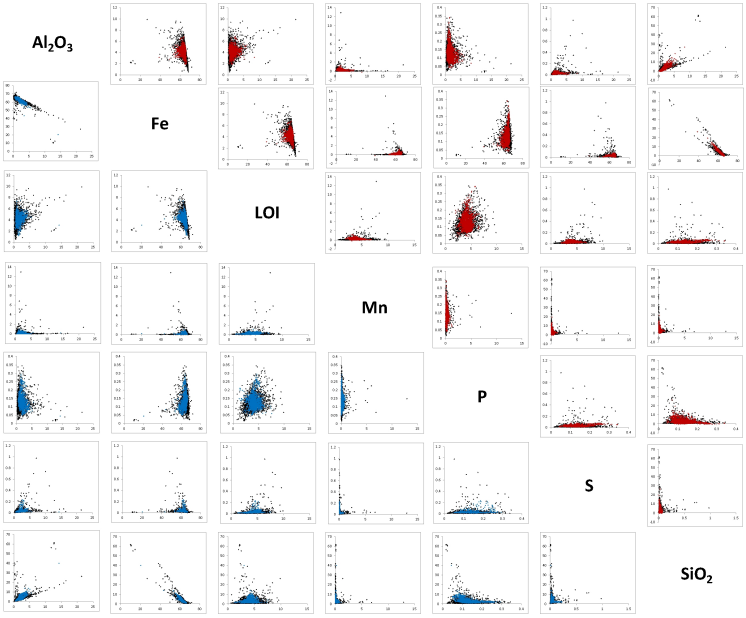


Figure F.3: Scatter plot matrix - Spacing 3. Black points are sample data, red points are OCK estimates, blue points are GH estimates

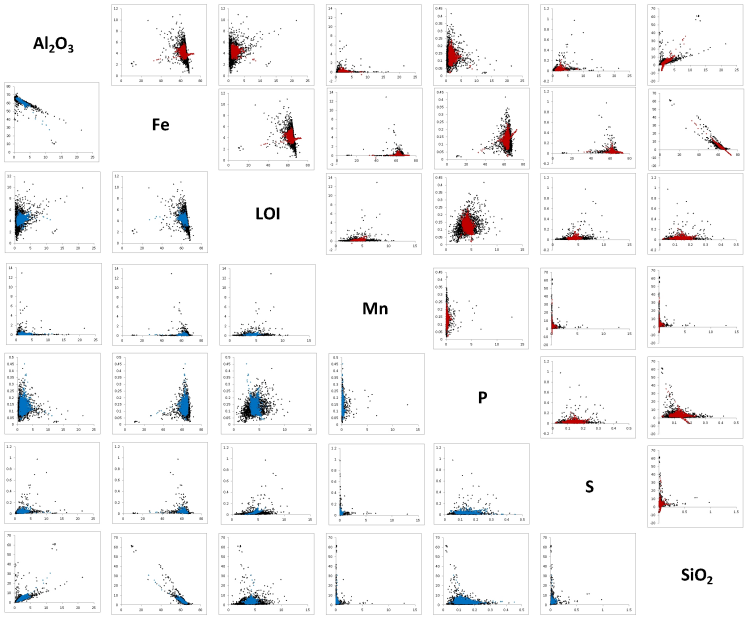


Figure F.4: Scatter plot matrix - Spacing 4. Black points are sample data, red points are OCK estimates, blue points are GH estimates

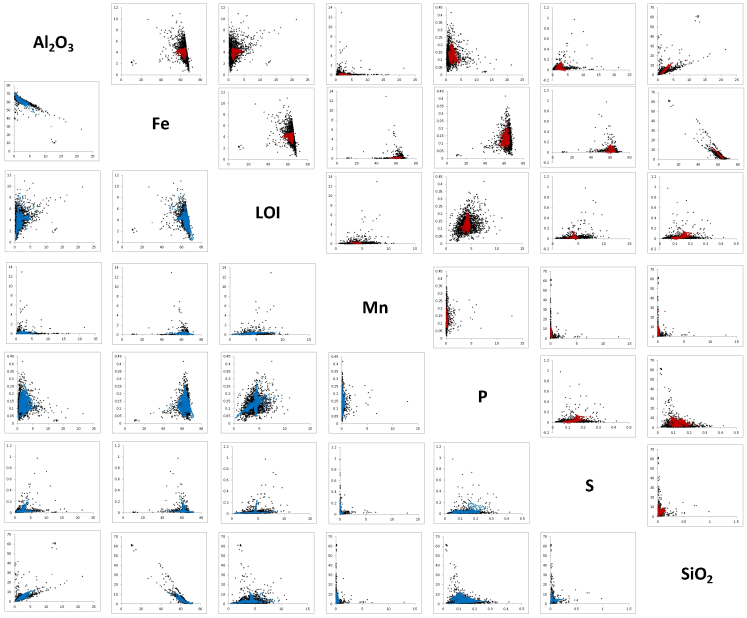
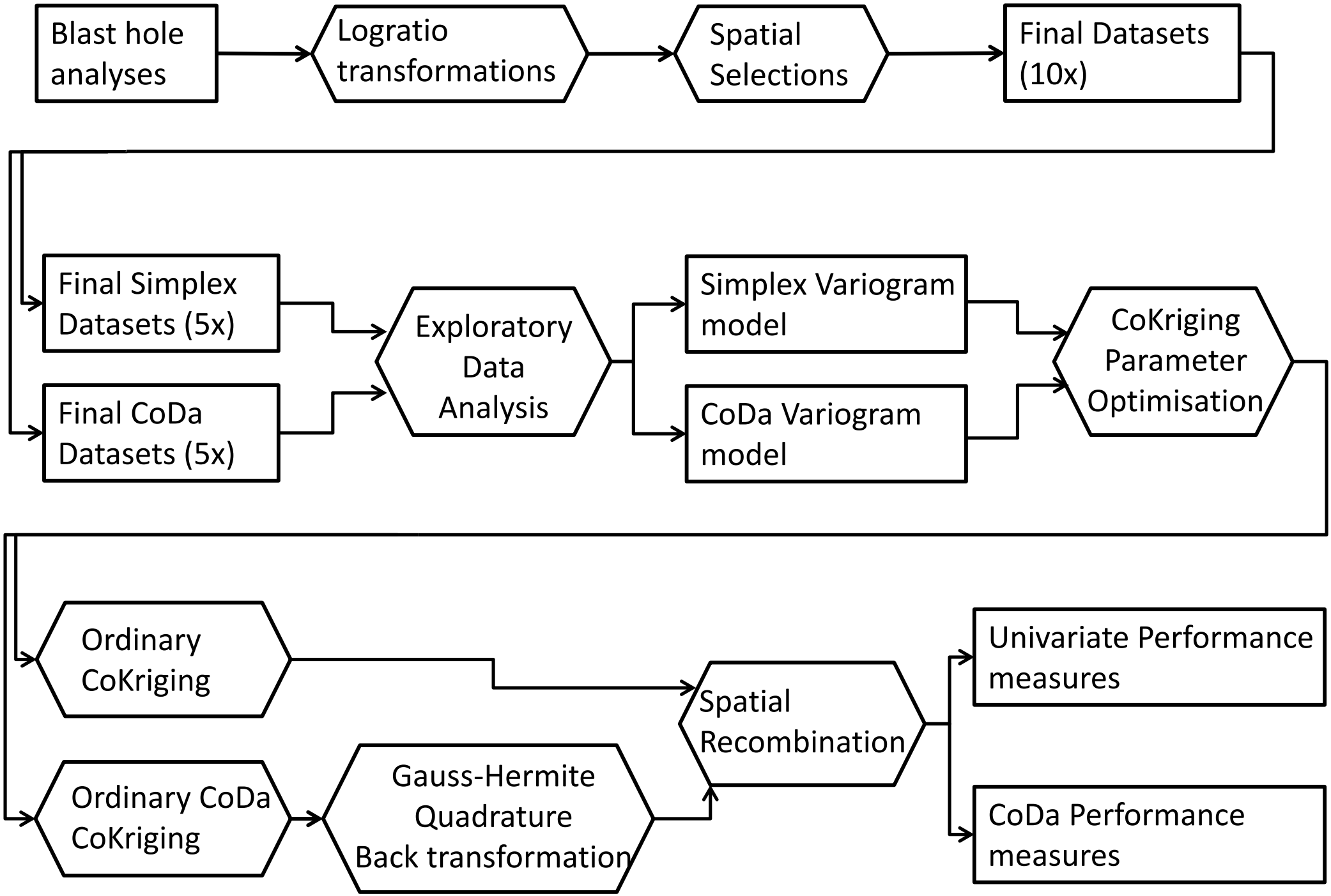


Figure F.5: Scatter plot matrix - Spacing 5. Black points are sample data, red points are OCK estimates, blue points are GH estimates

Appendix G

Detailed Implementation Flowsheet


Blast hole analyses
Logratio transformations
Spatial Selections
Final Datasets (10x)
Final Simplex Datasets (5x)
Final CoDa Datasets (5x)
Exploratory Data Analysis
Simplex Variogram model
CoDa Variogram model
CoKriging Parameter Optimisation
Ordinary CoKriging
Ordinary CoDa CoKriging
Gauss-Hermite Quadrature Back transformation
Spatial Recombination
Univariate Performance measures
CoDa Performance measures

Appendix H

Script File code

Compositional_lr_GH_backtrans_loop_input.m

```
% Read in the Excel files in the directory Specified
% Make sure this only contains the Excel files and nothing else
list=ls('C:\Users\ccward\Documents\MATLAB\East_all');
NL=size(list,1);
Final_Output=[];
for nl=3:NL
[num,txt,raw]=xlsread(['C:\Users\ccward\Documents\MATLAB\East_all\' list(nl,:)]);
[m,n]=size(raw);
% Nv = number of variables
% Nq = number of quadrature
Nv=7;
Nq=3;
Nf=0;
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Location')==1 && strcmp(cell2mat(raw(i+1,1)),'Location')==1 &&
strcmp(cell2mat(raw(i+2,1)),'Location')==1 && strcmp(cell2mat(raw(i+4,1)),'Successfully')==0
      Nf=Nf+1;
   end
end
count=0;
% reading the locations
Loc=zeros(Nf,3);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Location')==1 && strcmp(cell2mat(raw(i+1,1)),'Location')==1 &&
strcmp(cell2mat(raw(i+2,1)),'Location')==1 && strcmp(cell2mat(raw(i+4,1)),'Successfully')==0
      count=count+1;
      for j=1:3
         temp=cell2mat(raw(i+j-1,6));
         mt=size(temp,2);
         Loc(count,j)=str2double(temp(1:mt-1));
      end
   end
end
count=0;
%for each location read the number of samples usde in the krging system
Sample_Num=zeros(Nf,1);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Number')==1 && strcmp(cell2mat(raw(i,2)),'of')==1 &&
strcmp(cell2mat(raw(i,3)),'active')==1 && strcmp(cell2mat(raw(i,4)),'samples')==1
      count=count+1;
      Sample_Num(count)=cell2mat(raw(i,6));
   end
end
count=0;
RHS=zeros(Nv*max(Sample_Num)+Nv,Nv,Nf);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Right')==1 && strcmp(cell2mat(raw(i,2)),'Hand')==1
      count=count+1;
      for j=1:min(9,Nv*Sample_Num(count)+Nv)
         RHS(1:9,:,count)=cell2mat(raw(i+4:i+12,4:4+Nv-1));
```

```
      end
      if Nv*Sample_Num(count)+Nv>9
         for j=10:Nv*Sample_Num(count)+Nv

RHS(10:Nv*Sample_Num(count)+Nv,:,count)=cell2mat(raw(i+13:i+Nv*Sample_Num(count)+4+Nv-
1,3:3+Nv-1));
         end
      end
   end
end
count=0;
Weight=zeros(Nv*max(Sample_Num)+Nv,Nv,Nf);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Kriging')==1 && strcmp(cell2mat(raw(i,2)),'variable')==1 &&
strcmp(cell2mat(raw(i,3)),'V1')==1
      count=count+1;
      for j=1:Nv
         Weight(1+(j-1)*Sample_Num(count):j*Sample_Num(count),:,count)=cell2mat(raw(i+1+(j-
1)*(Sample_Num(count)+3):i+j*(Sample_Num(count)+3)-3,6:6+Nv-1));
      end
   end
end
count=0;
Lagrange=zeros(Nv,Nv,Nf);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'-')==1 && strcmp(cell2mat(raw(i,2)),'For')==1 &&
strcmp(cell2mat(raw(i,3)),'variable')==1 && strcmp(cell2mat(raw(i,4)),'V1')==1
      count=count+1;
      for j=1:Nv
         Lagrange(:,j,count)=cell2mat(raw(i+6+(j-1)*(Nv+15):i+6+Nv-1+(j-1)*(Nv+15),6));
         Weight(Nv*Sample_Num(count)+1:Nv*Sample_Num(count)+Nv,j,count)=cell2mat(raw(i+6+(j-
1)*(Nv+15):i+6+Nv-1+(j-1)*(Nv+15),6));
      end
   end
end
count=0;
estimate=zeros(Nv,Nf);
for i=1:m
   if strcmp(cell2mat(raw(i,1)),'Variable')==1 && strcmp(cell2mat(raw(i,2)),'V1')==1
      count=count+1;
      for j=1:Nv
         estimate(j,count)=cell2mat(raw(i+(j-1)*4+1,4));
      end
   end
end
varcovar=zeros(Nv,Nv,Nf);
RHS
Weight
c0=[0.987 0.14773388 -0.0224836 0.00679483 0.3976438 0.71495431 -0.033101601
0.14773388 0.13431773 0.1275228 0.07079109 0.129107903 0.1190044 -0.003049348
-0.0224836 0.1275228 1.040791 0.11354852 -0.1706903 -0.28940986 -0.0086542
```

```
0.00679483 0.07079109 0.11354852 0.1924235 0.06081324 -0.0967181 0.00952995
0.3976438 0.129107903 -0.1706903 0.06081324 0.6336 0.2217039 -0.004777286
0.71495431 0.1190044 -0.28940986 -0.0967181 0.2217039 1.264299 -0.03384148
-0.033101601 -0.003049348 -0.0086542 0.00952995 -0.004777286 -0.03384148 0.0109
];
for i=1:Nf
   varcovar(:,:,i)=c0-Weight(:,:,i)'*RHS(:,:,i);
end
R=zeros(Nv,Nv,Nf);
for i=1:Nf
   R(:,:,i)=chol(varcovar(:,:,i),'lower');
end
%Calculate quadrature points and product fo weights
[GHYs,GHweights]=GHvars(Nq,Nv);

%calculate Backtransformed data via GH integration
est=zeros(Nv,Nq^Nv,Nf);
for i=1:Nf
  est(:,:,i)=diag(estimate(:,i))*ones(size(GHYs))+sqrt(2)*R(:,:,i)*GHYs;
end
expest=zeros(Nv+1,Nq^Nv,Nf);
for i=1:Nf
   expest(1:Nv,:,i)=exp(est(:,:,i));
   expest(Nv+1,:,i)=ones(1,Nq^Nv);
end
sumexpest=zeros(Nq^Nv,Nf);
for i=1:Nf
   sumexpest(:,i)=sum(expest(:,:,i));
end
g=zeros(Nv+1,Nq^Nv,Nf);
for i=1:Nf
   for j=1:Nv+1
      for k=1:Nq^Nv
         g(j,k,i)=expest(j,k,i)/sumexpest(k,i);
      end
   end
end
ALRestGH=zeros(Nv+1,Nf);
alr=zeros(Nv+1,Nq^Nv,Nf);
for i=1:Nf
   alr(:,:,i)=g(:,:,i).*GHweights/pi^(Nv/2);
   ALRestGH(:,i)=sum(alr(:,:,i),2);
end
M=zeros(Nv+1,Nv+1,Nf);
GHweights2=zeros(Nv+1,Nv+1,Nq^Nv);
for i=1:Nq^Nv
   for k=1:Nv+1
   GHweights2(:,k,i)=GHweights(:,i);
   end
end
M;
```

```
for i=1:Nf
   Mi=zeros(Nv+1,Nv+1,Nq^Nv);
   for k=1:Nq^Nv
   Mi(:,:,k)=g(:,k,i)*g(:,k,i)';
   end
   size(Mi)
   Mi=Mi.*GHweights2/pi^(Nv/2);
   M(:,:,i)=sum(Mi,3);
   M(:,:,i)=M(:,:,i)-ALRestGH(:,i)*ALRestGH(:,i)';
end
M;
Output=cell((Nv+1)*Nf,3+2*(Nv+1));
for i=1:Nf
   for j=1:3
      Output((Nv+1)*(i-1)+1,j)={Loc(i,j)};
   end
   for j=1:Nv+1
      for k=1:Nv+1
         Output((Nv+1)*(i-1)+j,3+k)={M(j,k,i)};
      end
   end

   for j=1:Nv+1
      Output((Nv+1)*(i-1)+1,4+Nv+j)={ALRestGH(j,i)};
   end
end
Final_Output=[Final_Output; Output];
%write Output to file

end
xlswrite('C:\Users\ccward\Documents\MATLAB\East_all\Final_Output.xlsx',Final_Output)
```

GHvars.m

```
function [GHY GHw]=GHvars(N,M)
% N is the number of quadrature points
% M is the number of variables
GHY=[];
GHw=[];
[Y W]=hermquad(N)
% Y=[-1.224744871       0          1.224744871];
% W=[0.295409 1.18164 0.295409];
ghy=zeros(M,1);
ghw=zeros(1,M);
count=0;
for i=1:N^M
  count=count+1;
  for j=1:M
     ghy(j,1)=Y(ceil((mod(count-1,N^(M+1-j))+1)/N^(M-j)),1);
     ghw(1,j)=W(ceil((mod(count-1,N^(M+1-j))+1)/N^(M-j)),1);
  end
```

```
  GHY=[GHY ghy];
  GHw=[GHw prod(ghw)*ones(M+1,1)];
end
GHY;
GHw;
```

Hermquad.m

```
function [X W] = hermquad(N)
% % [X W] = HERMQUAD(N)
% % Find the Gauss-Hermite abscissae and weights. % % Arguments:
% N - The number of abscissae and weights to return.
% % Return Values: % X - A column vector containing the abscissae.
% W - A column vector containing the corresponding weights.
% % Gauss-Hermite quadrature approximates definite integrals of the form
% % \int^{-\infty}_{\infty} dx W(x) f(x)
% % where % % W(x) = \exp( - x^2 )
% % with the sum % % \sum_{n=1}^{N} w_{n} f(x_{n}).
% % This function returns the set of abscissae and weights
% % {x_{n}, w_{n}}^{N}_{n=1}
% % for performing this calculation given N, the number of abscissae.
% These abscissae correspond to the zeros of the Nth Hermite
% polynomial. It can be shown that such integration is exact when f(x)
% is a polynomial of maximum order 2N-1.
% % The procedure in this calculation is taken more or less directly from
% % @BOOK{ press-etal-1992a,
% AUTHOR = { Press, William H. and % Flannery, Brian P. and % Teukolsky, Saul A. and % Vetterling,
William T. },
% ISBN = {0521431085}, % MONTH = {October}, % PUBLISHER = {{Cambridge University Press}},
% TITLE = {Numerical Recipes in C : The Art of Scientific Computing}, % YEAR = {1992} % }
% % precision
EPS = 3.0e-14;
% 1/\pi^{1/4}
PIM4 = 0.7511255444649425;
% maximum number of loops
MAXIT = 10;
% allocate the return values
X = zeros([N 1]);
W = zeros([N 1]);
for i=1:(N+1)/2
% good guesses at initial values for specific roots
if i == 1
   z = sqrt(2.0*N+1.0) - 1.85575*((2.0*N+1)^(-0.16667));
elseif i == 2
   z = z - (1.14 * N^0.426 / z);
elseif i == 3
   z = 1.86 * z - 0.86 * X(1);
elseif i == 4
   z = 1.91 * z - 0.91 * X(2);
```

```
else
   z = 2.0*z - X(i-2);
end
for iter=1:MAXIT+1
   p1 = PIM4;
   p2 = 0.0;
   for j=1:N
      p3 = p2;
      p2 = p1;
      p1 = z * sqrt(2.0/j) * p2 - sqrt((j-1.0)/j) * p3;
   end
% the derivative
pp = sqrt(2.0*N) * p2;
% newton step
z1 = z;
z = z1 - p1/pp;
if abs(z-z1) <= EPS break;
end
end
if iter == MAXIT+1
   fprintf('Too many iterations in hermquad.\n');
end
X(i) = z;
X(N+1-i) = -z;
W(i) = 2.0/(pp*pp);
W(N+1-i) = W(i);
end
```

Appendix I

Variogram models plots

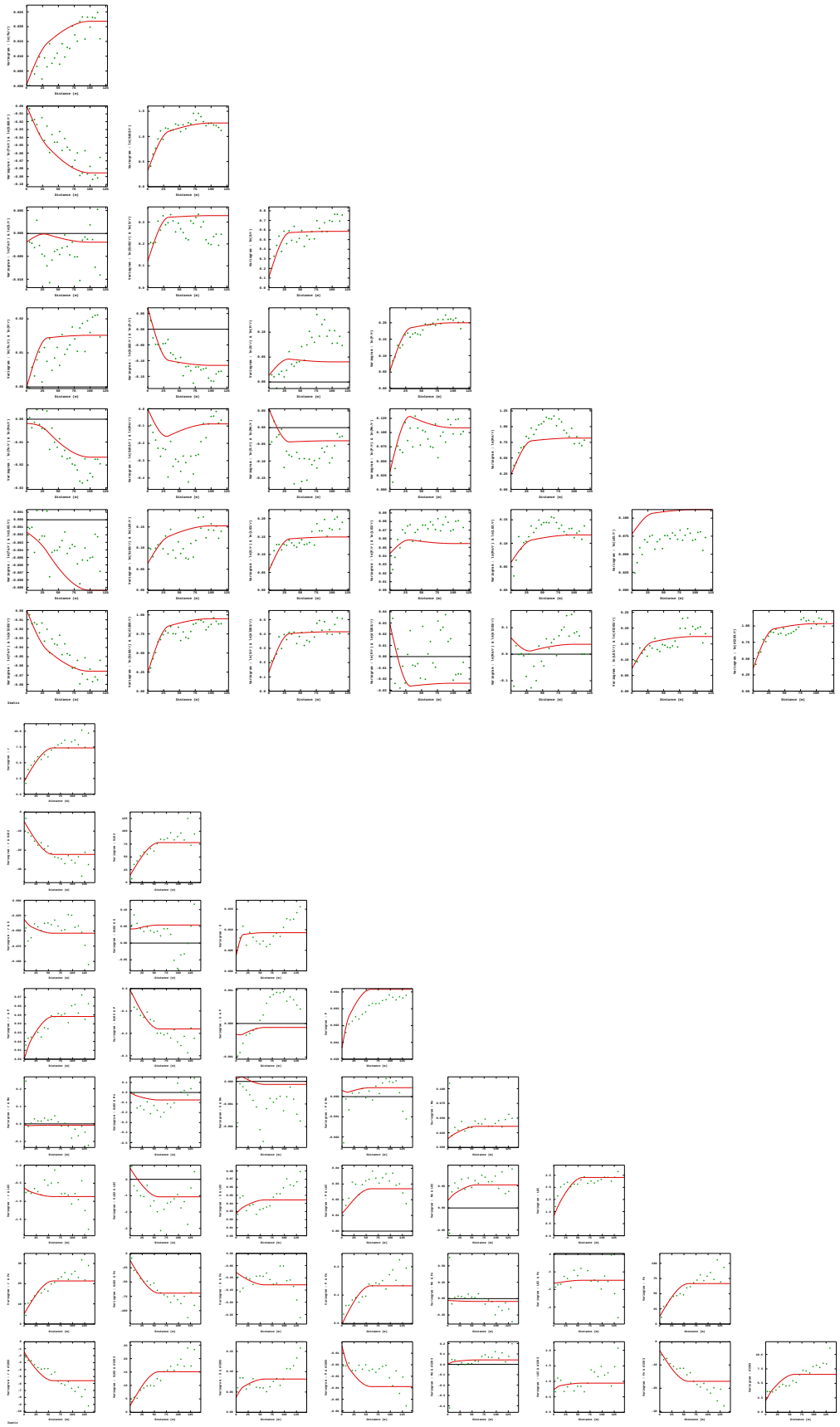


Figure I.1: Semivariogram model - East zone Spacing 2 (Top is logratio, bottom is raw space)

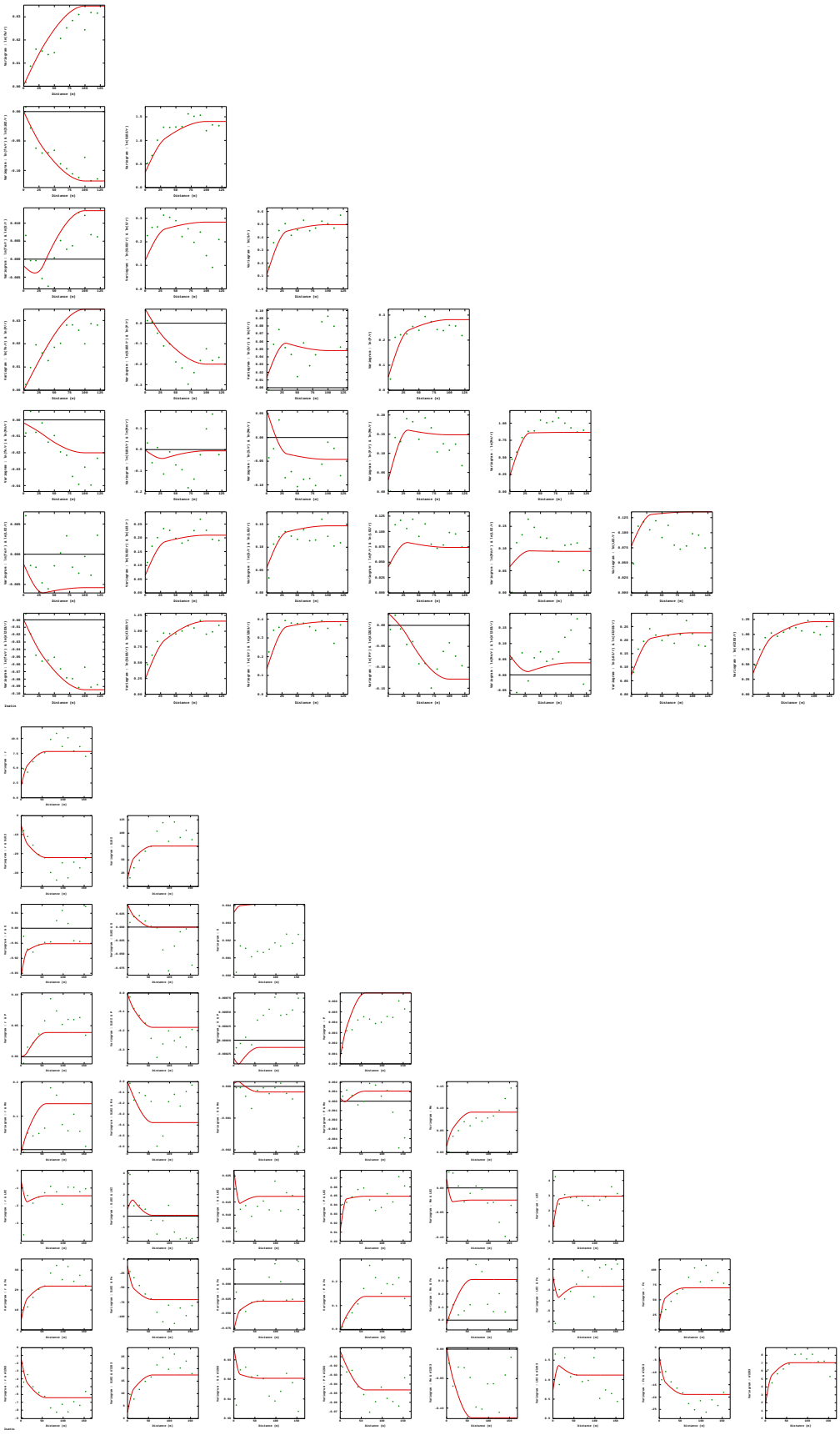


Figure I.2: Semivariogram model - East zone Spacing 3 (Top is logratio, bottom is raw space)

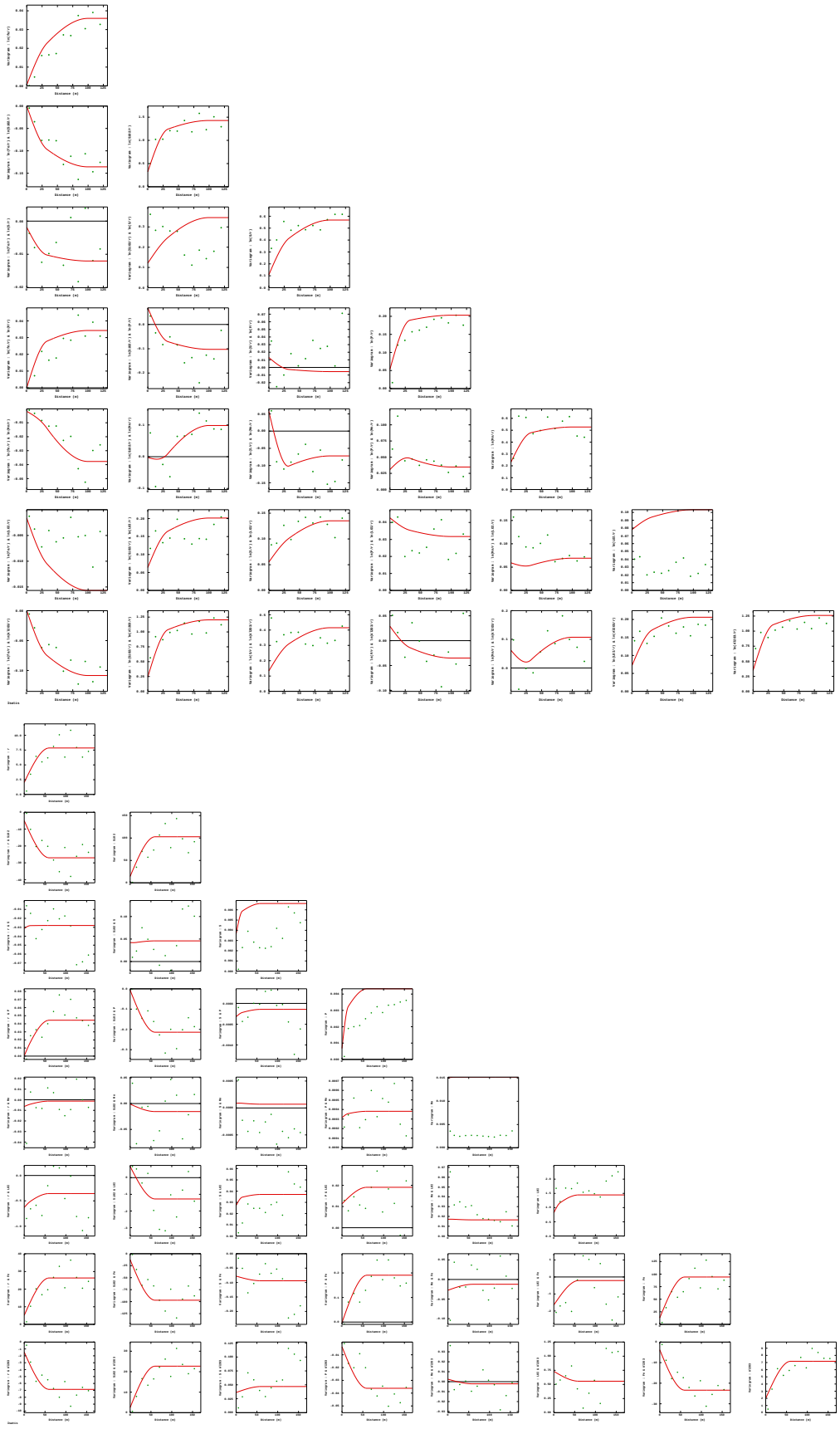


Figure I.3: Semivariogram model - East zone Spacing 4 (Top is logratio, bottom is raw space)

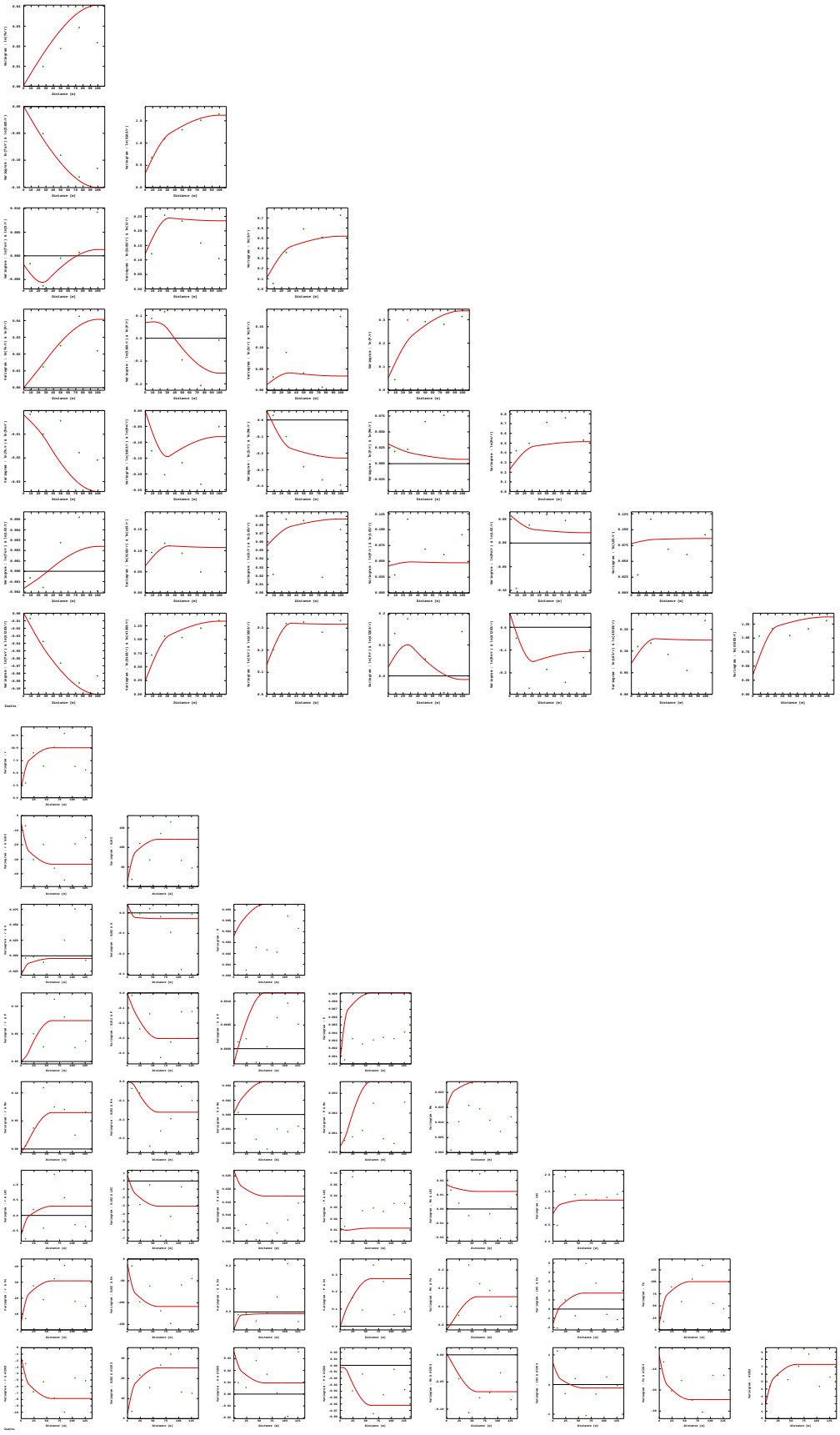


Figure I.4: Semivariogram model - East zone Spacing 5 (Top is logratio, bottom is raw space)

Appendix J

Variogram models Tabulations

J.1 Constrained space Models

---> Set name : East_1_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	1.98	-3.7932	0.7321	0.0025	-0.0037	0.0362	2.5155	-1.4632
Fe	2	-3.7932	12.95	-1.644	-0.0275	-0.0021	-0.0776	-12.5326	5.1276
LOI	3	0.7321	-1.644	0.82	0.0171	0.0113	0.0271	0.6705	-0.6307
Mn	4	0.0025	-0.0275	0.0171	0.015	0.0003	0.0001	-0.0011	-0.0064
P	5	-0.0037	-0.0021	0.0113	0.0003	0.0007	-0.0003	-0.0065	0.0003
S	6	0.0362	-0.0776	0.0271	0.0001	-0.0003	0.0036	0.0422	-0.0313
SiO2	7	2.5155	-12.5326	0.6705	-0.0011	-0.0065	0.0422	14.35	-5.0338
r	8	-1.4632	5.1276	-0.6307	-0.0064	0.0003	-0.0313	-5.0338	2.038

S2 : Spherical - Range = 5.412964880m
Directional Scales = (15.00m, 5.00m, 5.412964880m)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.1672	0.4009	-0.0161	0.0035	-0.002	0.0005	-0.7167	0.1625
Fe	2	0.4009	1.2571	-0.1049	-0.0182	-0.0089	0.0037	-2.2057	0.676
LOI	3	-0.0161	-0.1049	0.5978	0.0165	0.0112	-0.0051	-0.5048	0.0055
Mn	4	0.0035	-0.0182	0.0165	0.0119	-0.0001	-0.0001	-0.0161	0.0025
P	5	-0.002	-0.0089	0.0112	-0.0001	0.0003	-0.0001	0.0054	-0.0059
S	6	0.0005	0.0037	-0.0051	-0.0001	-0.0001	0.0001	-0.0016	0.0026
SiO2	7	-0.7167	-2.2057	-0.5048	-0.0161	0.0054	-0.0016	4.789	-1.3494
r	8	0.1625	0.676	0.0055	0.0025	-0.0059	0.0026	-1.3494	0.5062

S3 : Spherical - Range = 15.416376691m
Directional Scales = (60.00m, 15.00m, 15.416376691m)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	2.9537	-9.5479	0.1792	-0.0004	-0.0157	0.0054	9.3071	-2.8814
Fe	2	-9.5479	38.4941	0.2527	0.0718	0.1199	0.0094	-40.6989	11.2989
LOI	3	0.1792	0.2527	0.4598	-0.0078	0.0031	0.0077	-0.8536	-0.0411
Mn	4	-0.0004	0.0718	-0.0078	0.0093	0.001	-0.0007	-0.1005	0.0273
P	5	-0.0157	0.1199	0.0031	0.001	0.0023	0.0008	-0.1576	0.0463
S	6	0.0054	0.0094	0.0077	-0.0007	0.0008	0.0005	-0.0285	0.0054
SiO2	7	9.3071	-40.6989	-0.8536	-0.1005	-0.1576	-0.0285	44.4334	-11.9015
r	8	-2.8814	11.2989	-0.0411	0.0273	0.0463	0.0054	-11.9015	3.4461

Figure J.1: Semivariogram model tabulation - East zone Spacing 1 true values

---> Set name : East_2_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Variable	1	1.98	-3.7932	0.7321	0.0025	-0.0037	0.0362	2.5155	-1.4632
Variable	2	-3.7932	12.95	-1.644	-0.0275	-0.0021	-0.0776	-12.5326	5.1276
Variable	3	0.7321	-1.644	0.82	0.0171	0.0113	0.0271	0.6705	-0.6307
Variable	4	0.0025	-0.0275	0.0171	0.015	0.0003	0.0001	-0.0011	-0.0064
Variable	5	-0.0037	-0.0021	0.0113	0.0003	0.0007	-0.0003	-0.0065	0.0003
Variable	6	0.0362	-0.0776	0.0271	0.0001	-0.0003	0.0036	0.0422	-0.0313
Variable	7	2.5155	-12.5326	0.6705	-0.0011	-0.0065	0.0422	14.35	-5.0338
Variable	8	-1.4632	5.1276	-0.6307	-0.0064	0.0003	-0.0313	-5.0338	2.038

S2 : Spherical - Range = 15.00m
Directional Scales = (15.00m, 5.00m, 15.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Variable	1	0.5384	-0.0374	0.0898	0.0166	-0.0105	0.0155	-0.1087	-0.1954
Variable	2	-0.0374	0.0026	-0.007	-0.0011	0.0008	-0.001	0.0076	0.0136
Variable	3	0.0898	-0.007	0.1057	0.0051	-0.0007	0.0031	-0.0301	-0.0432
Variable	4	0.0166	-0.0011	0.0051	0.0016	-0.0002	0.0003	-0.0112	-0.0046
Variable	5	-0.0105	0.0008	-0.0007	-0.0002	0.001	-0.0001	0.0009	0.0074
Variable	6	0.0155	-0.001	0.0031	0.0003	-0.0001	0.005	-0.004	-0.0061
Variable	7	-0.1087	0.0076	-0.0301	-0.0112	0.0009	-0.004	0.0906	0.0168
Variable	8	-0.1954	0.0136	-0.0432	-0.0046	0.0074	-0.0061	0.0168	0.1019

S3 : Spherical - Range = 60.00m
Directional Scales = (60.00m, 15.00m, 60.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	4.0354	-13.1923	0.1207	0.0232	-0.0247	0.0303	12.7097	-3.9343
Fe	2	-13.1923	53.7461	0.1764	-0.0136	0.1339	-0.0501	-56.868	16.2178
LOI	3	0.1207	0.1764	1.4832	0.0302	0.0164	0.014	-1.709	-0.194
Mn	4	0.0232	-0.0136	0.0302	0.0193	0.0003	-0.0005	-0.0622	0.0034
P	5	-0.0247	0.1339	0.0164	0.0003	0.0025	0.0003	-0.1746	0.0405
S	6	0.0303	-0.0501	0.014	-0.0005	0.0003	0.0007	0.0158	-0.0165
SiO2	7	12.7097	-56.868	-1.709	-0.0622	-0.1746	0.0158	63.438	-17.2195
r	8	-3.9343	16.2178	-0.194	0.0034	0.0405	-0.0165	-17.2195	5.2082

Figure J.2: Semivariogram model tabulation - East zone Spacing 2 true values

---> Set name : East_3_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	1.98	-3.7932	0.7321	0.0025	-0.0037	0.0362	2.5155	-1.4632
Fe	2	-3.7932	12.95	-1.644	-0.0275	-0.0021	-0.0776	-12.5326	5.1276
LOI	3	0.7321	-1.644	0.82	0.0171	0.0113	0.0271	0.6705	-0.6307
Mn	4	0.0025	-0.0275	0.0171	0.015	0.0003	0.0001	-0.0011	-0.0064
P	5	-0.0037	-0.0021	0.0113	0.0003	0.0007	-0.0003	-0.0065	0.0003
S	6	0.0362	-0.0776	0.0271	0.0001	-0.0003	0.0036	0.0422	-0.0313
SiO2	7	2.5155	-12.5326	0.6705	-0.0011	-0.0065	0.0422	14.35	-5.0338
r	8	-1.4632	5.1276	-0.6307	-0.0064	0.0003	-0.0313	-5.0338	2.038

S2 : Spherical - Range = 15.00m
Directional Scales = (15.00m, 5.00m, 15.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	2.5856	-7.5506	0.7658	-0.0127	-0.0013	-0.0125	6.246	-2.0065
Fe	2	-7.5506	31.0952	-2.6566	0.0129	0.0112	0.0249	-27.8038	7.2035
LOI	3	0.7658	-2.6566	1.8519	-0.0468	0.0337	-0.0142	1.5414	-1.3367
Mn	4	-0.0127	0.0129	-0.0468	0.018	-0.0009	0.0002	-0.0087	0.0128
P	5	-0.0013	0.0112	0.0337	-0.0009	0.0009	-0.0002	-0.0247	-0.0105
S	6	-0.0125	0.0249	-0.0142	0.0002	-0.0002	0.0003	-0.0097	0.0141
SiO2	7	6.246	-27.8038	1.5414	-0.0087	-0.0247	-0.0097	25.697	-5.8668
r	8	-2.0065	7.2035	-1.3367	0.0128	-0.0105	0.0141	-5.8668	2.0944

S3 : Spherical - Range = 60.00m
Directional Scales = (60.00m, 15.00m, 60.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	2.4902	-7.5948	-0.3808	-0.1063	-0.0414	-0.0031	8.7058	-2.9441
Fe	2	-7.5948	25.6822	1.6651	0.3256	0.129	0.0236	-30.1857	9.6437
LOI	3	-0.3808	1.6651	0.2925	0.0051	0.0047	0.0042	-2.1313	0.5316
Mn	4	-0.1063	0.3256	0.0051	0.0581	0.0017	-0.0005	-0.368	0.1304
P	5	-0.0414	0.129	0.0047	0.0017	0.0053	0.0004	-0.1516	0.0492
S	6	-0.0031	0.0236	0.0042	-0.0005	0.0004	0.0001	-0.0326	0.0069
SiO2	7	8.7058	-30.1857	-2.1313	-0.368	-0.1516	-0.0326	35.7744	-11.2186
r	8	-2.9441	9.6437	0.5316	0.1304	0.0492	0.0069	-11.2186	3.6769

Figure J.3: Semivariogram model tabulation - East zone Spacing 3 true values

---> Set name : East_4_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	1.98	-3.7932	0.7321	0.0025	-0.0037	0.0362	2.5155	-1.4632
Fe	2	-3.7932	12.95	-1.644	-0.0275	-0.0021	-0.0776	-12.5326	5.1276
LOI	3	0.7321	-1.644	0.82	0.0171	0.0113	0.0271	0.6705	-0.6307
Mn	4	0.0025	-0.0275	0.0171	0.015	0.0003	0.0001	-0.0011	-0.0064
P	5	-0.0037	-0.0021	0.0113	0.0003	0.0007	-0.0003	-0.0065	0.0003
S	6	0.0362	-0.0776	0.0271	0.0001	-0.0003	0.0036	0.0422	-0.0313
SiO2	7	2.5155	-12.5326	0.6705	-0.0011	-0.0065	0.0422	14.35	-5.0338
r	8	-1.4632	5.1276	-0.6307	-0.0064	0.0003	-0.0313	-5.0338	2.038

S2 : Spherical - Range = 15.00m
Directional Scales = (15.00m, 5.00m, 15.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0	0	-0.0004	0	0	-0.0001	0	0
Fe	2	0	0.0002	-0.0057	0	0	-0.0002	-0.0001	-0.001
LOI	3	-0.0004	-0.0057	0.1885	0.0003	0	0.0061	0.0039	0.0334
Mn	4	0	0	0.0003	0.0002	0	0	0.0001	0.0006
P	5	0	0	0	0	0.0019	0.0001	0	0.0015
S	6	-0.0001	-0.0002	0.0061	0	0.0001	0.0018	-0.0017	0.0031
SiO2	7	0	-0.0001	0.0039	0.0001	0	-0.0017	0.0024	-0.0017
r	8	0	-0.001	0.0334	0.0006	0.0015	0.0031	-0.0017	0.0123

S3 : Spherical - Range = 60.00m
Directional Scales = (60.00m, 15.00m, 60.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	5.1983	-19.5941	-0.1745	-0.0047	-0.0324	0.0108	20.1853	-5.3968
Fe	2	-19.5941	81.2524	1.4409	0.0152	0.1929	-0.016	-84.5106	21.2388
LOI	3	-0.1745	1.4409	0.4378	-0.0009	0.0078	0.0038	-1.9525	0.241
Mn	4	-0.0047	0.0152	-0.0009	0	0	0	-0.0145	0.0045
P	5	-0.0324	0.1929	0.0078	0	0.0017	0.0001	-0.2074	0.0426
S	6	0.0108	-0.016	0.0038	0	0.0001	0.0012	0.0055	0.0002
SiO2	7	20.1853	-84.5106	-1.9525	-0.0145	-0.2074	0.0055	88.4598	-21.9797
r	8	-5.3968	21.2388	0.241	0.0045	0.0426	0.0002	-21.9797	5.802

Figure J.4: Semivariogram model tabulation - East zone Spacing 4 true values

---> Set name : East_5_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Variable	1	1.98	-3.7932	0.7321	0.0025	-0.0037	0.0362	2.5155	-1.4632
Variable	2	-3.7932	12.95	-1.644	-0.0275	-0.0021	-0.0776	-12.5326	5.1276
Variable	3	0.7321	-1.644	0.82	0.0171	0.0113	0.0271	0.6705	-0.6307
Variable	4	0.0025	-0.0275	0.0171	0.015	0.0003	0.0001	-0.0011	-0.0064
Variable	5	-0.0037	-0.0021	0.0113	0.0003	0.0007	-0.0003	-0.0065	0.0003
Variable	6	0.0362	-0.0776	0.0271	0.0001	-0.0003	0.0036	0.0422	-0.0313
Variable	7	2.5155	-12.5326	0.6705	-0.0011	-0.0065	0.0422	14.35	-5.0338
Variable	8	-1.4632	5.1276	-0.6307	-0.0064	0.0003	-0.0313	-5.0338	2.038

S2 : Spherical - Range = 15.00m
Directional Scales = (15.00m, 5.00m, 15.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Variable	1	2.301	-9.5157	-0.3749	0.0019	0.0213	-0.0113	10.5106	-2.9396
Variable	2	-9.5157	44.0515	1.1854	-0.0088	0.0314	0.0615	-48.4825	12.8812
Variable	3	-0.3749	1.1854	0.2188	-0.0008	-0.0027	-0.004	-1.3399	0.4247
Variable	4	0.0019	-0.0088	-0.0008	0.0038	-0.0005	0.0003	0.0493	-0.0049
Variable	5	0.0213	0.0314	-0.0027	-0.0005	0.0051	0.0001	-0.0262	-0.0139
Variable	6	-0.0113	0.0615	-0.004	0.0003	0.0001	0.0003	-0.0613	0.0149
Variable	7	10.5106	-48.4825	-1.3399	0.0493	-0.0262	-0.0613	54.0106	-14.2855
Variable	8	-2.9396	12.8812	0.4247	-0.0049	-0.0139	0.0149	-14.2855	3.9835

S3 : Spherical - Range = 60.00m
Directional Scales = (60.00m, 15.00m, 60.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	3.0453	-11.2182	-0.4723	-0.0723	-0.0797	-0.0154	12.3111	-3.4567
Fe	2	-11.2182	43.6116	2.1981	0.19	0.2465	0.008	-47.8008	12.8873
LOI	3	-0.4723	2.1981	0.1973	-0.004	0.0029	-0.0059	-2.4137	0.5054
Mn	4	-0.0723	0.19	-0.004	0.0048	0.0038	0.0019	-0.2094	0.0763
P	5	-0.0797	0.2465	0.0029	0.0038	0.0033	0.0014	-0.2692	0.0875
S	6	-0.0154	0.008	-0.0059	0.0019	0.0014	0.0026	-0.009	0.0121
SiO2	7	12.3111	-47.8008	-2.4137	-0.2094	-0.2692	-0.009	52.4523	-14.096
r	8	-3.4567	12.8873	0.5054	0.0763	0.0875	0.0121	-14.096	4.0587

Figure J.5: Semivariogram model tabulation - East zone Spacing 5 true values

---> Set name : West_1_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.9	-1.2778	0.3884	0.0173	0.0018	0.0052	0.5238	-0.5617
Fe	2	-1.2778	2.8	-1.1403	-0.4174	-0.0095	-0.0094	-0.9407	1.0111
LOI	3	0.3884	-1.1403	0.675	0.0999	0.007	0.0023	0.4167	-0.4455
Mn	4	0.0173	-0.4174	0.0999	0.36	0.0008	-0.0004	-0.0021	-0.0597
P	5	0.0018	-0.0095	0.007	0.0008	0.0005	0	0.0031	-0.0036
S	6	0.0052	-0.0094	0.0023	-0.0004	0	0.0002	0.0054	-0.0033
SiO2	7	0.5238	-0.9407	0.4167	-0.0021	0.0031	0.0054	0.42	-0.4218
r	8	-0.5617	1.0111	-0.4455	-0.0597	-0.0036	-0.0033	-0.4218	0.4845

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	1.3878	-3.4074	0.1959	0.0547	0.0051	0.0077	2.785	-1.0289
Fe	2	-3.4074	9.5069	-1.0295	-0.1752	-0.023	-0.0161	-7.7349	2.8792
LOI	3	0.1959	-1.0295	0.8864	0.0596	0.0234	0.0063	0.1854	-0.3275
Mn	4	0.0547	-0.1752	0.0596	0.0506	0.0019	0.0022	0.0593	-0.053
P	5	0.0051	-0.023	0.0234	0.0019	0.0008	0.0002	-0.001	-0.0074
S	6	0.0077	-0.0161	0.0063	0.0022	0.0002	0.0004	0.004	-0.0046
SiO2	7	2.785	-7.7349	0.1854	0.0593	-0.001	0.004	7.0328	-2.3306
r	8	-1.0289	2.8792	-0.3275	-0.053	-0.0074	-0.0046	-2.3306	0.8728

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.6194	-1.4092	0.0683	-0.0059	-0.0126	0.0009	1.2177	-0.4785
Fe	2	-1.4092	4.4685	-0.4028	-0.0532	0.0388	-0.0028	-3.8609	1.2217
LOI	3	0.0683	-0.4028	0.6854	0.0396	-0.0188	0.0025	-0.1872	-0.187
Mn	4	-0.0059	-0.0532	0.0396	0.074	0.0004	-0.0002	-0.0614	0.0068
P	5	-0.0126	0.0388	-0.0188	0.0004	0.0008	-0.0001	-0.0233	0.0149
S	6	0.0009	-0.0028	0.0025	-0.0002	-0.0001	0.0001	0.0011	-0.0014
SiO2	7	1.2177	-3.8609	-0.1872	-0.0614	-0.0233	0.0011	3.9372	-1.0231
r	8	-0.4785	1.2217	-0.187	0.0068	0.0149	-0.0014	-1.0231	0.4467

Figure J.6: Semivariogram model tabulation - West zone Spacing 1 true values

---> Set name : West_2_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.9	-1.2778	0.3884	0.0173	0.0018	0.0052	0.5238	-0.5617
Fe	2	-1.2778	2.8	-1.1403	-0.4174	-0.0095	-0.0094	-0.9407	1.0111
LOI	3	0.3884	-1.1403	0.675	0.0999	0.007	0.0023	0.4167	-0.4455
Mn	4	0.0173	-0.4174	0.0999	0.36	0.0008	-0.0004	-0.0021	-0.0597
P	5	0.0018	-0.0095	0.007	0.0008	0.0005	0	0.0031	-0.0036
S	6	0.0052	-0.0094	0.0023	-0.0004	0	0.0002	0.0054	-0.0033
SiO2	7	0.5238	-0.9407	0.4167	-0.0021	0.0031	0.0054	0.42	-0.4218
r	8	-0.5617	1.0111	-0.4455	-0.0597	-0.0036	-0.0033	-0.4218	0.4845

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Variable	2	Variable	3	Variable	4	Variable	5
Al2O3	1	0.0812	-0.2344	0.0569	0.0374	0.0013	-0.004	0.2069	-0.087
Fe	2	-0.2344	1.2649	-0.7438	-0.0912	-0.0254	0.0058	-0.7696	0.4553
LOI	3	0.0569	-0.7438	0.68	0.0046	0.0216	-0.0044	0.2517	-0.2062
Mn	4	0.0374	-0.0912	0.0046	0.0182	0.0002	-0.0011	0.0964	-0.0395
P	5	0.0013	-0.0254	0.0216	0.0002	0.0018	-0.0005	0.0101	-0.0105
S	6	-0.004	0.0058	-0.0044	-0.0011	-0.0005	0.0113	-0.0024	0.0008
SiO2	7	0.2069	-0.7696	0.2517	0.0964	0.0101	-0.0024	0.6577	-0.3433
r	8	-0.087	0.4553	-0.2062	-0.0395	-0.0105	0.0008	-0.3433	0.2524

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Variable	2	Variable	3	Variable	4	Variable	5
Al2O3	1	1.8783	-5.5872	-0.3308	0.0223	-0.0071	0.013	5.3753	-1.3535
Fe	2	-5.5872	19.0693	0.963	-0.0081	0.0466	-0.0122	-18.5428	4.469
LOI	3	-0.3308	0.963	0.4372	0.0864	-0.0051	0.0172	-1.283	0.2345
Mn	4	0.0223	-0.0081	0.0864	0.0242	-0.0032	0.0046	-0.0808	-0.0119
P	5	-0.0071	0.0466	-0.0051	-0.0032	0.0154	0.0032	-0.0403	0.0221
S	6	0.013	-0.0122	0.0172	0.0046	0.0032	0.0027	-0.0122	0.0046
SiO2	7	5.3753	-18.5428	-1.283	-0.0808	-0.0403	-0.0122	19.1095	-4.3956
r	8	-1.3535	4.469	0.2345	-0.0119	0.0221	0.0046	-4.3956	1.1889

Figure J.7: Semivariogram model tabulation - West zone Spacing 2 true values

---> Set name : West_3_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.9	-1.2778	0.3884	0.0173	0.0018	0.0052	0.5238	-0.5617
Fe	2	-1.2778	2.8	-1.1403	-0.4174	-0.0095	-0.0094	-0.9407	1.0111
LOI	3	0.3884	-1.1403	0.675	0.0999	0.007	0.0023	0.4167	-0.4455
Mn	4	0.0173	-0.4174	0.0999	0.36	0.0008	-0.0004	-0.0021	-0.0597
P	5	0.0018	-0.0095	0.007	0.0008	0.0005	0	0.0031	-0.0036
S	6	0.0052	-0.0094	0.0023	-0.0004	0	0.0002	0.0054	-0.0033
SiO2	7	0.5238	-0.9407	0.4167	-0.0021	0.0031	0.0054	0.42	-0.4218
r	8	-0.5617	1.0111	-0.4455	-0.0597	-0.0036	-0.0033	-0.4218	0.4845

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	3.7345	-10.4468	1.3953	-0.171	0.009	0.0359	8.7671	-3.4595
Fe	2	-10.4468	40.9101	-3.9346	0.4441	-0.0077	-0.0708	-38.9995	12.2373
LOI	3	1.3953	-3.9346	2.0338	-0.1241	0.0269	0.0083	2.6121	-1.8449
Mn	4	-0.171	0.4441	-0.1241	0.0104	-0.0014	-0.0015	-0.3289	0.1733
P	5	0.009	-0.0077	0.0269	-0.0014	0.0011	0.0001	-0.0133	-0.0109
S	6	0.0359	-0.0708	0.0083	-0.0015	0.0001	0.0005	0.0507	-0.0226
SiO2	7	8.7671	-38.9995	2.6121	-0.3289	-0.0133	0.0507	38.9215	-11.0156
r	8	-3.4595	12.2373	-1.8449	0.1733	-0.0109	-0.0226	-11.0156	4.1089

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	3.1335	-8.8092	-0.9391	-0.0252	-0.075	-0.0254	10.0036	-2.9342
Fe	2	-8.8092	31.6933	3.1573	0.141	0.2085	0.0626	-37.4422	10.9104
LOI	3	-0.9391	3.1573	0.3844	0.009	0.0316	0.0071	-3.6336	1.0577
Mn	4	-0.0252	0.141	0.009	0.0232	0.0004	0.0009	-0.1034	0.0349
P	5	-0.075	0.2085	0.0316	0.0004	0.0062	0.0014	-0.229	0.0638
S	6	-0.0254	0.0626	0.0071	0.0009	0.0014	0.0009	-0.0669	0.0199
SiO2	7	10.0036	-37.4422	-3.6336	-0.1034	-0.229	-0.0669	44.7857	-13.0231
r	8	-2.9342	10.9104	1.0577	0.0349	0.0638	0.0199	-13.0231	3.7965

Figure J.8: Semivariogram model tabulation - West zone Spacing 3 true values

---> Set name : West_4_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.9	-1.2778	0.3884	0.0173	0.0018	0.0052	0.5238	-0.5617
Fe	2	-1.2778	2.8	-1.1403	-0.4174	-0.0095	-0.0094	-0.9407	1.0111
LOI	3	0.3884	-1.1403	0.675	0.0999	0.007	0.0023	0.4167	-0.4455
Mn	4	0.0173	-0.4174	0.0999	0.36	0.0008	-0.0004	-0.0021	-0.0597
P	5	0.0018	-0.0095	0.007	0.0008	0.0005	0	0.0031	-0.0036
S	6	0.0052	-0.0094	0.0023	-0.0004	0	0.0002	0.0054	-0.0033
SiO2	7	0.5238	-0.9407	0.4167	-0.0021	0.0031	0.0054	0.42	-0.4218
r	8	-0.5617	1.0111	-0.4455	-0.0597	-0.0036	-0.0033	-0.4218	0.4845

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	3.1015	-9.3158	-0.4612	-0.3834	-0.0338	0.0235	8.5267	-1.4208
Fe	2	-9.3158	32.5452	1.5931	1.2671	0.0661	-0.0003	-29.638	5.1624
LOI	3	-0.4612	1.5931	0.9925	0.2345	0.0201	0.0601	-1.6455	0.063
Mn	4	-0.3834	1.2671	0.2345	0.1148	-0.0137	0.0142	-1.2065	0.1385
P	5	-0.0338	0.0661	0.0201	-0.0137	0.0139	0.0007	-0.0566	0.0195
S	6	0.0235	-0.0003	0.0601	0.0142	0.0007	0.0341	-0.0074	-0.0196
SiO2	7	8.5267	-29.638	-1.6455	-1.2065	-0.0566	-0.0074	31.6378	-5.07
r	8	-1.4208	5.1624	0.063	0.1385	0.0195	-0.0196	-5.07	1.0769

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.0717	0.0209	0.0642	0.0555	0.0188	-0.0179	-0.0209	0.1525
Fe	2	0.0209	0.016	-0.0152	0.0261	0.0121	-0.0029	0.011	0.0919
LOI	3	0.0642	-0.0152	0.2271	-0.0205	-0.0293	-0.0174	-0.0954	0.1387
Mn	4	0.0555	0.0261	-0.0205	0.0798	0.0301	-0.0271	0.0097	0.074
P	5	0.0188	0.0121	-0.0293	0.0301	0.0451	0.0306	0.0303	-0.0459
S	6	-0.0179	-0.0029	-0.0174	-0.0271	0.0306	0.1292	0.0271	0.0771
SiO2	7	-0.0209	0.011	-0.0954	0.0097	0.0303	0.0271	0.0674	0.0628
r	8	0.1525	0.0919	0.1387	0.074	-0.0459	0.0771	0.0628	2.8723

Figure J.9: Semivariogram model tabulation - West zone Spacing 4 true values

---> Set name : West_5_raw

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.9	-1.2778	0.3884	0.0173	0.0018	0.0052	0.5238	-0.5617
Fe	2	-1.2778	2.8	-1.1403	-0.4174	-0.0095	-0.0094	-0.9407	1.0111
LOI	3	0.3884	-1.1403	0.675	0.0999	0.007	0.0023	0.4167	-0.4455
Mn	4	0.0173	-0.4174	0.0999	0.36	0.0008	-0.0004	-0.0021	-0.0597
P	5	0.0018	-0.0095	0.007	0.0008	0.0005	0	0.0031	-0.0036
S	6	0.0052	-0.0094	0.0023	-0.0004	0	0.0002	0.0054	-0.0033
SiO2	7	0.5238	-0.9407	0.4167	-0.0021	0.0031	0.0054	0.42	-0.4218
r	8	-0.5617	1.0111	-0.4455	-0.0597	-0.0036	-0.0033	-0.4218	0.4845

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	3.2036	-12.3901	-0.2018	-0.035	-0.0056	-0.0006	11.7517	-1.9367
Fe	2	-12.3901	49.437	1.0511	0.1633	0.0188	0.0084	-46.692	7.5374
LOI	3	-0.2018	1.0511	1.0711	-0.03	0.0146	0.0028	-1.9433	0.2765
Mn	4	-0.035	0.1633	-0.03	0.0023	-0.0005	0.0001	-0.1151	0.0166
P	5	-0.0056	0.0188	0.0146	-0.0005	0.0003	0	-0.0325	0.0055
S	6	-0.0006	0.0084	0.0028	0.0001	0	0.0002	-0.0089	0.0008
SiO2	7	11.7517	-46.692	-1.9433	-0.1151	-0.0325	-0.0089	45.0786	-7.2848
r	8	-1.9367	7.5374	0.2765	0.0166	0.0055	0.0008	-7.2848	1.1934

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	Al2O3	Fe	LOI	Mn	P	S	SiO2	r
Al2O3	1	0.0011	-0.004	-0.0015	-0.001	-0.0002	0	0.0036	-0.0007
Fe	2	-0.004	0.0157	0.0015	0.0043	0.0007	0.0002	-0.0128	0.0086
LOI	3	-0.0015	0.0015	0.0109	-0.0003	0.0002	-0.0005	-0.004	-0.0126
Mn	4	-0.001	0.0043	-0.0003	0.0039	-0.0002	0	0.007	0.0034
P	5	-0.0002	0.0007	0.0002	-0.0002	0.0003	0	-0.0016	0.0003
S	6	0	0.0002	-0.0005	0	0	0	-0.0003	0.0007
SiO2	7	0.0036	-0.0128	-0.004	0.007	-0.0016	-0.0003	0.0666	-0.0072
r	8	-0.0007	0.0086	-0.0126	0.0034	0.0003	0.0007	-0.0072	0.0232

Figure J.10: Semivariogram model tabulation - West zone Spacing 5 true values

J.2 Logratio Models

---> Set name : east_1_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.347	0.072	0.0617	0.029	0.134	0.2428	0.0001
ln(LOI/filler)	2	0.072	0.078	0.0592	0.0427	0.0552	0.0638	-0.0016
ln(Mn/filler)	3	0.0617	0.0592	0.2308	0.0309	0.0569	-0.0024	-0.0019
ln(P/filler)	4	0.029	0.0427	0.0309	0.0534	0.013	0.0683	0
ln(S/filler)	5	0.134	0.0552	0.0569	0.013	0.1086	0.1213	-0.0019
ln(SiO2/filler)	6	0.2428	0.0638	-0.0024	0.0683	0.1213	0.3243	-0.0001
ln(fe/filler)	7	0.0001	-0.0016	-0.0019	0	-0.0019	-0.0001	0.0005

S2 : Spherical - Range = 13.00m
Directional Scales = (33.00m, 13.00m, 13.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.49	0.078	-0.0399	-0.0285	0.2281	0.466	-0.033
ln(LOI/filler)	2	0.078	0.0542	0.0483	0.0252	0.0747	0.08	-0.0015
ln(Mn/filler)	3	-0.0399	0.0483	0.48	0.0904	-0.09	-0.121	-0.0082
ln(P/filler)	4	-0.0285	0.0252	0.0904	0.089	-0.0015	-0.1	0.006
ln(S/filler)	5	0.2281	0.0747	-0.09	-0.0015	0.34	0.1969	-0.0023
ln(SiO2/filler)	6	0.466	0.08	-0.121	-0.1	0.1969	0.59	-0.0325
ln(fe/filler)	7	-0.033	-0.0015	-0.0082	0.006	-0.0023	-0.0325	0.0051

S3 : Spherical - Range = 25.00m
Directional Scales = (100.00m, 25.00m, 25.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.15	-0.0022	-0.0443	0.0063	0.0356	0.0061	-0.0002
ln(LOI/filler)	2	-0.0022	0.0021	0.02	0.0028	-0.0007	-0.0248	0.0001
ln(Mn/filler)	3	-0.0443	0.02	0.33	-0.0078	-0.1376	-0.166	0.0014
ln(P/filler)	4	0.0063	0.0028	-0.0078	0.05	0.0493	-0.065	0.0035
ln(S/filler)	5	0.0356	-0.0007	-0.1376	0.0493	0.185	-0.0965	-0.0006
ln(SiO2/filler)	6	0.0061	-0.0248	-0.166	-0.065	-0.0965	0.35	-0.0012
ln(fe/filler)	7	-0.0002	0.0001	0.0014	0.0035	-0.0006	-0.0012	0.0053

Figure J.11: Semivariogram model tabulation - East zone Spacing 1 logratios

---> Set name : east_2_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.347	0.072	0.0617	0.029	0.134	0.2428	0.0001
ln(LOI/filler)	2	0.072	0.078	0.0592	0.0427	0.0552	0.0638	-0.0016
ln(Mn/filler)	3	0.0617	0.0592	0.2308	0.0309	0.0569	-0.0024	-0.0019
ln(P/filler)	4	0.029	0.0427	0.0309	0.0534	0.013	0.0683	0
ln(S/filler)	5	0.134	0.0552	0.0569	0.013	0.1086	0.1213	-0.0019
ln(SiO2/filler)	6	0.2428	0.0638	-0.0024	0.0683	0.1213	0.3243	-0.0001
ln(fe/filler)	7	0.0001	-0.0016	-0.0019	0	-0.0019	-0.0001	0.0005

S2 : Spherical - Range = 33.00m
Directional Scales = (33.00m, 13.00m, 33.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.5371	0.0694	-0.0707	-0.0576	0.2573	0.5229	-0.023
ln(LOI/filler)	2	0.0694	0.0242	0.038	0.0197	0.0838	0.0411	0.0018
ln(Mn/filler)	3	-0.0707	0.038	0.5057	0.1149	-0.1028	-0.2173	0.0067
ln(P/filler)	4	-0.0576	0.0197	0.1149	0.1171	0.0378	-0.1535	0.0137
ln(S/filler)	5	0.2573	0.0838	-0.1028	0.0378	0.4532	0.1936	0.0032
ln(SiO2/filler)	6	0.5229	0.0411	-0.2173	-0.1535	0.1936	0.6244	-0.0214
ln(fe/filler)	7	-0.023	0.0018	0.0067	0.0137	0.0032	-0.0214	0.0074

S3 : Spherical - Range = 100.00m
Directional Scales = (100.00m, 25.00m, 100.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.152	0.032	0.0464	0.0048	0.0243	0.1841	-0.0429
ln(LOI/filler)	2	0.032	0.0096	0.0206	-0.008	0.0097	0.0477	-0.0095
ln(Mn/filler)	3	0.0464	0.0206	0.0814	-0.0373	0.0065	0.1319	-0.0214
ln(P/filler)	4	0.0048	-0.008	-0.0373	0.0293	-0.0105	-0.0298	0.0014
ln(S/filler)	5	0.0243	0.0097	0.0065	-0.0105	0.0242	0.0145	-0.0032
ln(SiO2/filler)	6	0.1841	0.0477	0.1319	-0.0298	0.0145	0.3178	-0.0637
ln(fe/filler)	7	-0.0429	-0.0095	-0.0214	0.0014	-0.0032	-0.0637	0.0139

Figure J.12: Semivariogram model tabulation - East zone Spacing 2 logratios

---> Set name : east_3_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
Variable	1	0.347	0.072	0.0617	0.029	0.134	0.2428	0.0001
Variable	2	0.072	0.078	0.0592	0.0427	0.0552	0.0638	-0.0016
Variable	3	0.0617	0.0592	0.2308	0.0309	0.0569	-0.0024	-0.0019
Variable	4	0.029	0.0427	0.0309	0.0534	0.013	0.0683	0
Variable	5	0.134	0.0552	0.0569	0.013	0.1086	0.1213	-0.0019
Variable	6	0.2428	0.0638	-0.0024	0.0683	0.1213	0.3243	-0.0001
Variable	7	0.0001	-0.0016	-0.0019	0	-0.0019	-0.0001	0.0005

S2 : Spherical - Range = 33.00m
Directional Scales = (33.00m, 13.00m, 33.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
Variable	1	0.406	0.1169	-0.0741	0.0306	0.2023	0.3619	-0.0119
Variable	2	0.1169	0.0486	0.0366	0.0462	0.0663	0.1012	-0.0055
Variable	3	-0.0741	0.0366	0.6205	0.1405	-0.081	-0.0687	0.0022
Variable	4	0.0306	0.0462	0.1405	0.1478	0.0529	-0.0361	-0.0002
Variable	5	0.2023	0.0663	-0.081	0.0529	0.2927	0.1079	-0.013
Variable	6	0.3619	0.1012	-0.0687	-0.0361	0.1079	0.3989	-0.0145
Variable	7	-0.0119	-0.0055	0.0022	-0.0002	-0.013	-0.0145	0.0019

S3 : Spherical - Range = 100.00m
Directional Scales = (100.00m, 25.00m, 100.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.4586	0.0379	0.0511	-0.188	0.0501	0.5551	-0.0827
ln(LOI/filler)	2	0.0379	0.008	-0.0022	-0.0149	0.0253	0.0446	0.0016
ln(Mn/filler)	3	0.0511	-0.0022	0.0146	-0.0231	-0.0221	0.0651	-0.0204
ln(P/filler)	4	-0.188	-0.0149	-0.0231	0.0805	-0.0178	-0.2319	0.035
ln(S/filler)	5	0.0501	0.0253	-0.0221	-0.0178	0.0971	0.0548	0.0283
ln(SiO2/filler)	6	0.5551	0.0446	0.0651	-0.2319	0.0548	0.6778	-0.1033
ln(fe/filler)	7	-0.0827	0.0016	-0.0204	0.035	0.0283	-0.1033	0.0323

Figure J.13: Semivariogram model tabulation - East zone Spacing 3 logratios

---> Set name : east_4_logratio

S1 : Nugget effect

Variance-Covariance matrix :							
Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.347	0.072	0.0617	0.029	0.134	0.2428	0.0001
ln(LOI/filler)	2	0.072	0.078	0.0592	0.0427	0.0552	0.0638	-0.0016
ln(Mn/filler)	3	0.0617	0.0592	0.2308	0.0309	0.0569	-0.0024	-0.0019
ln(P/filler)	4	0.029	0.0427	0.0309	0.0534	0.013	0.0683	0
ln(S/filler)	5	0.134	0.0552	0.0569	0.013	0.1086	0.1213	-0.0019
ln(SiO2/filler)	6	0.2428	0.0638	-0.0024	0.0683	0.1213	0.3243	-0.0001
ln(fe/filler)	7	0.0001	-0.0016	-0.0019	0	-0.0019	-0.0001	0.0005

S2 : Spherical - Range = 33.00m
Directional Scales = (33.00m, 13.00m, 33.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :							
Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.6205	0.0668	-0.1044	-0.0244	0.0819	0.6464	-0.0446
ln(LOI/filler)	2	0.0668	0.0077	-0.019	-0.004	0.0156	0.0682	-0.0042
ln(Mn/filler)	3	-0.1044	-0.019	0.2053	0.0295	-0.1846	-0.0677	0.0095
ln(P/filler)	4	-0.0244	-0.004	0.0295	0.1235	-0.0131	-0.1104	0.022
ln(S/filler)	5	0.0819	0.0156	-0.1846	-0.0131	0.1682	0.0378	-0.0067
ln(SiO2/filler)	6	0.6464	0.0682	-0.0677	-0.1104	0.0378	0.7513	-0.061
ln(fe/filler)	7	-0.0446	-0.0042	0.0095	0.022	-0.0067	-0.061	0.0102

S3 : Spherical - Range = 100.00m
Directional Scales = (100.00m, 25.00m, 100.00m)
Local Rotation = Azimuth=N110.00 (Geologist Plane)

Variance-Covariance matrix :							
Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.2928	0.0674	0.1501	-0.0397	0.1993	0.314	-0.064
ln(LOI/filler)	2	0.0674	0.0178	0.029	-0.0069	0.0645	0.0694	-0.0099
ln(Mn/filler)	3	0.1501	0.029	0.0912	-0.0256	0.0556	0.1684	-0.0453
ln(P/filler)	4	-0.0397	-0.0069	-0.0256	0.0262	-0.0054	-0.0604	0.0122
ln(S/filler)	5	0.1993	0.0645	0.0556	-0.0054	0.2911	0.1871	-0.0035
ln(SiO2/filler)	6	0.314	0.0694	0.1684	-0.0604	0.1871	0.3533	-0.0746
ln(fe/filler)	7	-0.064	-0.0099	-0.0453	0.0122	-0.0035	-0.0746	0.0253

Figure J.14: Semivariogram model tabulation - East zone Spacing 4 logratios

---> Set name : east_5_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.347	0.072	0.0617	0.029	0.134	0.2428	0.0001
ln(LOI/filler)	2	0.072	0.078	0.0592	0.0427	0.0552	0.0638	-0.0016
ln(Mn/filler)	3	0.0617	0.0592	0.2308	0.0309	0.0569	-0.0024	-0.0019
ln(P/filler)	4	0.029	0.0427	0.0309	0.0534	0.013	0.0683	0
ln(S/filler)	5	0.134	0.0552	0.0569	0.013	0.1086	0.1213	-0.0019
ln(SiO2/filler)	6	0.2428	0.0638	-0.0024	0.0683	0.1213	0.3243	-0.0001
ln(fe/filler)	7	0.0001	-0.0016	-0.0019	0	-0.0019	-0.0001	0.0005

S2 : Spherical - Range = 33.00m
Directional Scales = (33.00m, 13.00m, 33.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.7188	0.0598	-0.25	0.1559	0.1917	0.6069	-0.0072
ln(LOI/filler)	2	0.0598	0.0055	-0.0247	0.0076	0.0152	0.0512	-0.0005
ln(Mn/filler)	3	-0.25	-0.0247	0.1946	-0.0064	-0.1775	-0.1964	0.0062
ln(P/filler)	4	0.1559	0.0076	-0.0064	0.0875	0.0344	0.1268	-0.0023
ln(S/filler)	5	0.1917	0.0152	-0.1775	0.0344	0.2199	0.1319	-0.0089
ln(SiO2/filler)	6	0.6069	0.0512	-0.1964	0.1268	0.1319	0.5185	-0.0047
ln(fe/filler)	7	-0.0072	-0.0005	0.0062	-0.0023	-0.0089	-0.0047	0.0005

S3 : Spherical - Range = 100.00m
Directional Scales = (100.00m, 25.00m, 100.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3102	-0.0064	0.0814	-0.1966	-0.0081	0.4847	-0.1011
ln(LOI/filler)	2	-0.0064	0.0025	-0.0127	-0.0026	0.0165	-0.0073	0.0045
ln(Mn/filler)	3	0.0814	-0.0127	0.089	-0.0177	-0.1089	0.1164	-0.0386
ln(P/filler)	4	-0.1966	-0.0026	-0.0177	0.1967	-0.0138	-0.3484	0.0429
ln(S/filler)	5	-0.0081	0.0165	-0.1089	-0.0138	0.1907	-0.0181	0.0121
ln(SiO2/filler)	6	0.4847	-0.0073	0.1164	-0.3484	-0.0181	0.7823	-0.1461
ln(fe/filler)	7	-0.1011	0.0045	-0.0386	0.0429	0.0121	-0.1461	0.0395

Figure J.15: Semivariogram model tabulation - East zone Spacing 5 logratios

---> Set name : west_1_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3294	0.063	0.0088	0.0653	0.1538	0.2127	-0.0006
ln(LOI/filler)	2	0.063	0.03	0.0512	0.041	0.0333	0.0537	-0.0006
ln(Mn/filler)	3	0.0088	0.0512	0.3784	0.0494	0.0477	0.0732	-0.002
ln(P/filler)	4	0.0653	0.041	0.0494	0.0623	0.0327	0.056	-0.0011
ln(S/filler)	5	0.1538	0.0333	0.0477	0.0327	0.1506	0.0942	-0.0002
ln(SiO2/filler)	6	0.2127	0.0537	0.0732	0.056	0.0942	0.1742	0.0006
ln(fe/filler)	7	-0.0006	-0.0006	-0.002	-0.0011	-0.0002	0.0006	0.0001

S2 : Spherical - Range = 12.00m
Directional Scales = (18.00m, 12.00m, 12.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.2394	0.0381	0.0099	0.0239	0.0984	0.1683	-0.0026
ln(LOI/filler)	2	0.0381	0.0318	0.0475	0.0239	0.0618	0.0378	-0.0011
ln(Mn/filler)	3	0.0099	0.0475	0.3729	0.0213	0.0608	0.0268	-0.0042
ln(P/filler)	4	0.0239	0.0239	0.0213	0.0326	0.0333	0.0205	-0.0009
ln(S/filler)	5	0.0984	0.0618	0.0608	0.0333	0.157	0.0805	0.0007
ln(SiO2/filler)	6	0.1683	0.0378	0.0268	0.0205	0.0805	0.1346	-0.0034
ln(fe/filler)	7	-0.0026	-0.0011	-0.0042	-0.0009	0.0007	-0.0034	0.0013

S3 : Spherical - Range = 27.00m
Directional Scales = (70.00m, 27.00m, 27.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3638	0.0211	0.0122	-0.0773	0.0872	0.3623	-0.0177
ln(LOI/filler)	2	0.0211	0.0176	0.0412	-0.0065	0.0441	0.0021	-0.0009
ln(Mn/filler)	3	0.0122	0.0412	0.1873	0.0332	0.0295	-0.0509	-0.0017
ln(P/filler)	4	-0.0773	-0.0065	0.0332	0.0628	-0.0712	-0.085	0.0047
ln(S/filler)	5	0.0872	0.0441	0.0295	-0.0712	0.1801	0.0554	-0.0023
ln(SiO2/filler)	6	0.3623	0.0021	-0.0509	-0.085	0.0554	0.3869	-0.0181
ln(fe/filler)	7	-0.0177	-0.0009	-0.0017	0.0047	-0.0023	-0.0181	0.0034

Figure J.16: Semivariogram model tabulation - West zone Spacing 1 logratios

---> Set name : west_2_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3294	0.063	0.0088	0.0653	0.1538	0.2127	-0.0006
ln(LOI/filler)	2	0.063	0.03	0.0512	0.041	0.0333	0.0537	-0.0006
ln(Mn/filler)	3	0.0088	0.0512	0.3784	0.0494	0.0477	0.0732	-0.002
ln(P/filler)	4	0.0653	0.041	0.0494	0.0623	0.0327	0.056	-0.0011
ln(S/filler)	5	0.1538	0.0333	0.0477	0.0327	0.1506	0.0942	-0.0002
ln(SiO2/filler)	6	0.2127	0.0537	0.0732	0.056	0.0942	0.1742	0.0006
ln(fe/filler)	7	-0.0006	-0.0006	-0.002	-0.0011	-0.0002	0.0006	0.0001

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.1814	0.0259	0.2346	-0.0154	0.0514	0.2037	-0.0053
ln(LOI/filler)	2	0.0259	0.0112	0.0286	0.0001	0.0416	0.0249	-0.0008
ln(Mn/filler)	3	0.2346	0.0286	0.4277	-0.0014	0.0334	0.3442	-0.0155
ln(P/filler)	4	-0.0154	0.0001	-0.0014	0.0054	0.0046	-0.0057	-0.001
ln(S/filler)	5	0.0514	0.0416	0.0334	0.0046	0.1863	0.0312	0.0059
ln(SiO2/filler)	6	0.2037	0.0249	0.3442	-0.0057	0.0312	0.2814	-0.0119
ln(fe/filler)	7	-0.0053	-0.0008	-0.0155	-0.001	0.0059	-0.0119	0.0042

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.4304	-0.0214	-0.1975	-0.0742	0.1037	0.326	-0.0282
ln(LOI/filler)	2	-0.0214	0.017	0.0239	0.0036	0.0351	-0.0245	0.0013
ln(Mn/filler)	3	-0.1975	0.0239	0.1069	0.0464	-0.0155	-0.1618	0.012
ln(P/filler)	4	-0.0742	0.0036	0.0464	0.1052	-0.0553	-0.0952	0.0087
ln(S/filler)	5	0.1037	0.0351	-0.0155	-0.0553	0.1461	0.0728	-0.006
ln(SiO2/filler)	6	0.326	-0.0245	-0.1618	-0.0952	0.0728	0.268	-0.024
ln(fe/filler)	7	-0.0282	0.0013	0.012	0.0087	-0.006	-0.024	0.0065

Figure J.17: Semivariogram model tabulation - West zone Spacing 2 logratios

---> Set name : west_3_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3294	0.063	0.0088	0.0653	0.1538	0.2127	-0.0006
ln(LOI/filler)	2	0.063	0.03	0.0512	0.041	0.0333	0.0537	-0.0006
ln(Mn/filler)	3	0.0088	0.0512	0.3784	0.0494	0.0477	0.0732	-0.002
ln(P/filler)	4	0.0653	0.041	0.0494	0.0623	0.0327	0.056	-0.0011
ln(S/filler)	5	0.1538	0.0333	0.0477	0.0327	0.1506	0.0942	-0.0002
ln(SiO2/filler)	6	0.2127	0.0537	0.0732	0.056	0.0942	0.1742	0.0006
ln(fe/filler)	7	-0.0006	-0.0006	-0.002	-0.0011	-0.0002	0.0006	0.0001

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3098	0.058	0.3153	0.0866	0.0366	0.1345	0.0286
ln(LOI/filler)	2	0.058	0.0543	0.1127	0.0042	0.0226	0.0616	-0.002
ln(Mn/filler)	3	0.3153	0.1127	0.5104	0.0431	0.0704	0.1485	0.0114
ln(P/filler)	4	0.0866	0.0042	0.0431	0.0351	0.0025	0.0359	0.0118
ln(S/filler)	5	0.0366	0.0226	0.0704	0.0025	0.0117	0.0253	-0.0002
ln(SiO2/filler)	6	0.1345	0.0616	0.1485	0.0359	0.0253	0.0984	0.0083
ln(fe/filler)	7	0.0286	-0.002	0.0114	0.0118	-0.0002	0.0083	0.005

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.2176	-0.0152	-0.1644	-0.0762	-0.0137	0.204	0.0012
ln(LOI/filler)	2	-0.0152	0.0027	0.0141	0.0041	0.0162	-0.015	0
ln(Mn/filler)	3	-0.1644	0.0141	0.2246	0.0276	0.0766	-0.1169	-0.0068
ln(P/filler)	4	-0.0762	0.0041	0.0276	0.0382	-0.0147	-0.0755	-0.0007
ln(S/filler)	5	-0.0137	0.0162	0.0766	-0.0147	0.1921	-0.0181	0.0071
ln(SiO2/filler)	6	0.204	-0.015	-0.1169	-0.0755	-0.0181	0.2504	-0.0184
ln(fe/filler)	7	0.0012	0	-0.0068	-0.0007	0.0071	-0.0184	0.0074

Figure J.18: Semivariogram model tabulation - West zone Spacing 3 logratios

---> Set name : west_4_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3294	0.063	0.0088	0.0653	0.1538	0.2127	-0.0006
ln(LOI/filler)	2	0.063	0.03	0.0512	0.041	0.0333	0.0537	-0.0006
ln(Mn/filler)	3	0.0088	0.0512	0.3784	0.0494	0.0477	0.0732	-0.002
ln(P/filler)	4	0.0653	0.041	0.0494	0.0623	0.0327	0.056	-0.0011
ln(S/filler)	5	0.1538	0.0333	0.0477	0.0327	0.1506	0.0942	-0.0002
ln(SiO2/filler)	6	0.2127	0.0537	0.0732	0.056	0.0942	0.1742	0.0006
ln(fe/filler)	7	-0.0006	-0.0006	-0.002	-0.0011	-0.0002	0.0006	0.0001

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.0882	-0.021	-0.0817	-0.0232	-0.1136	0.1208	-0.0199
ln(LOI/filler)	2	-0.021	0.0116	-0.0159	-0.0013	0.055	-0.0238	0.0057
ln(Mn/filler)	3	-0.0817	-0.0159	0.2696	0.058	-0.0441	-0.1422	0.0104
ln(P/filler)	4	-0.0232	-0.0013	0.058	0.0227	0.0012	-0.0459	0.0089
ln(S/filler)	5	-0.1136	0.055	-0.0441	0.0012	0.2647	-0.1349	0.0295
ln(SiO2/filler)	6	0.1208	-0.0238	-0.1422	-0.0459	-0.1349	0.1804	-0.0285
ln(fe/filler)	7	-0.0199	0.0057	0.0104	0.0089	0.0295	-0.0285	0.0095

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)
Local Rotation = Azimuth=N80.00 (Geologist Plane)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.5382	0.1033	0.0641	-0.0571	0.3599	0.4639	-0.0146
ln(LOI/filler)	2	0.1033	0.0222	0.032	-0.0155	0.0741	0.0895	-0.0048
ln(Mn/filler)	3	0.0641	0.032	0.256	-0.0377	0.0801	0.036	-0.0115
ln(P/filler)	4	-0.0571	-0.0155	-0.0377	0.0439	-0.0479	-0.0555	-0.0008
ln(S/filler)	5	0.3599	0.0741	0.0801	-0.0479	0.2517	0.3124	-0.0143
ln(SiO2/filler)	6	0.4639	0.0895	0.036	-0.0555	0.3124	0.4067	-0.0137
ln(fe/filler)	7	-0.0146	-0.0048	-0.0115	-0.0008	-0.0143	-0.0137	0.0042

Figure J.19: Semivariogram model tabulation - West zone Spacing 4 logratios

---> Set name : west_5_logratio

S1 : Nugget effect

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.3294	0.063	0.0088	0.0653	0.1538	0.2127	-0.0006
ln(LOI/filler)	2	0.063	0.03	0.0512	0.041	0.0333	0.0537	-0.0006
ln(Mn/filler)	3	0.0088	0.0512	0.3784	0.0494	0.0477	0.0732	-0.002
ln(P/filler)	4	0.0653	0.041	0.0494	0.0623	0.0327	0.056	-0.0011
ln(S/filler)	5	0.1538	0.0333	0.0477	0.0327	0.1506	0.0942	-0.0002
ln(SiO2/filler)	6	0.2127	0.0537	0.0732	0.056	0.0942	0.1742	0.0006
ln(fe/filler)	7	-0.0006	-0.0006	-0.002	-0.0011	-0.0002	0.0006	0.0001

S2 : Spherical - Range = 18.00m
Directional Scales = (18.00m, 12.00m, 18.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.0125	-0.0086	0.0053	-0.0033	-0.0364	0.0255	-0.0125
ln(LOI/filler)	2	-0.0086	0.0286	-0.0395	0.0257	0.0331	0.0049	0.0082
ln(Mn/filler)	3	0.0053	-0.0395	0.071	-0.0204	-0.055	-0.0367	-0.0057
ln(P/filler)	4	-0.0033	0.0257	-0.0204	0.0526	-0.0225	-0.002	0.0019
ln(S/filler)	5	-0.0364	0.0331	-0.055	-0.0225	0.1769	-0.0292	0.0295
ln(SiO2/filler)	6	0.0255	0.0049	-0.0367	-0.002	-0.0292	0.098	-0.0351
ln(fe/filler)	7	-0.0125	0.0082	-0.0057	0.0019	0.0295	-0.0351	0.0214

S3 : Spherical - Range = 70.00m
Directional Scales = (70.00m, 27.00m, 70.00m)

Variance-Covariance matrix :

Variable	1	ln(Al2O3/filler)	ln(LOI/filler)	ln(Mn/filler)	ln(P/filler)	ln(S/filler)	ln(SiO2/filler)	ln(fe/filler)
ln(Al2O3/filler)	1	0.9714	0.0176	-0.0066	-0.1284	-0.125	0.9347	-0.1543
ln(LOI/filler)	2	0.0176	0.0715	0.1561	0.0497	0.0461	0.017	-0.0034
ln(Mn/filler)	3	-0.0066	0.1561	0.4827	0.1125	0.1438	-0.0213	-0.0139
ln(P/filler)	4	-0.1284	0.0497	0.1125	0.055	0.0512	-0.1233	0.0202
ln(S/filler)	5	-0.125	0.0461	0.1438	0.0512	0.0588	-0.1242	0.0159
ln(SiO2/filler)	6	0.9347	0.017	-0.0213	-0.1233	-0.1242	0.901	-0.147
ln(fe/filler)	7	-0.1543	-0.0034	-0.0139	0.0202	0.0159	-0.147	0.0259

Figure J.20: Semivariogram model tabulation - West zone Spacing 5 logratios

Appendix K

Histograms - Sample and estimated data

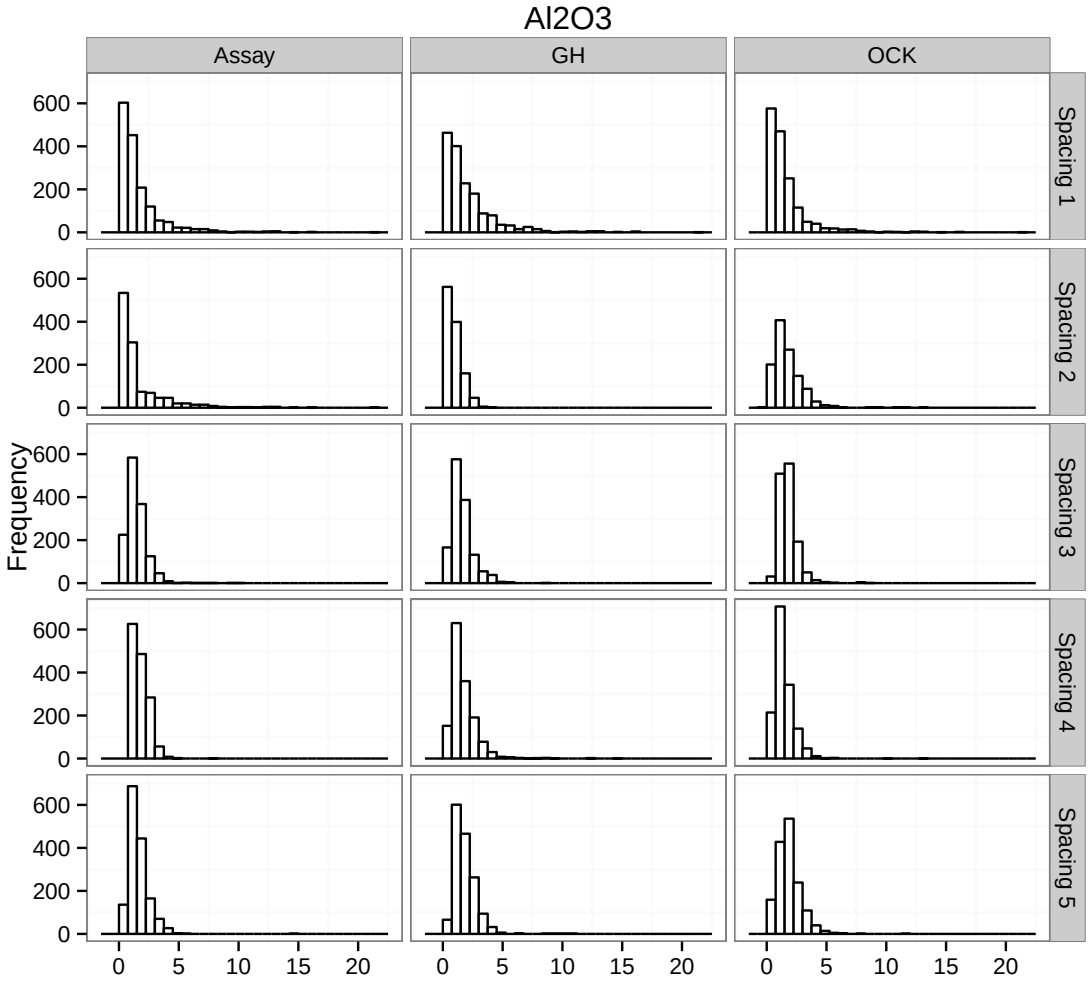


Figure K.1: Sample and Estimate histograms Al_2O_3 plots

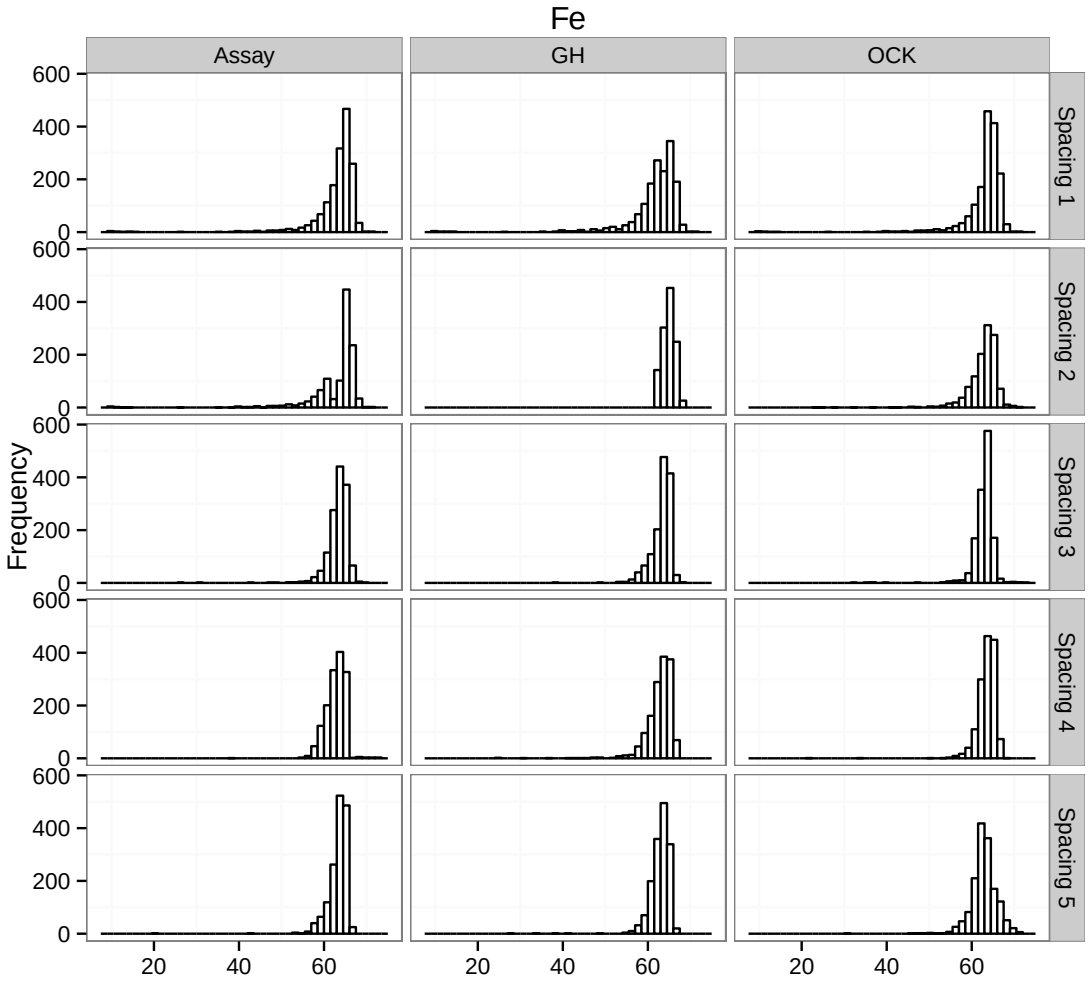


Figure K.2: Sample and Estimate histograms *Fe* plots

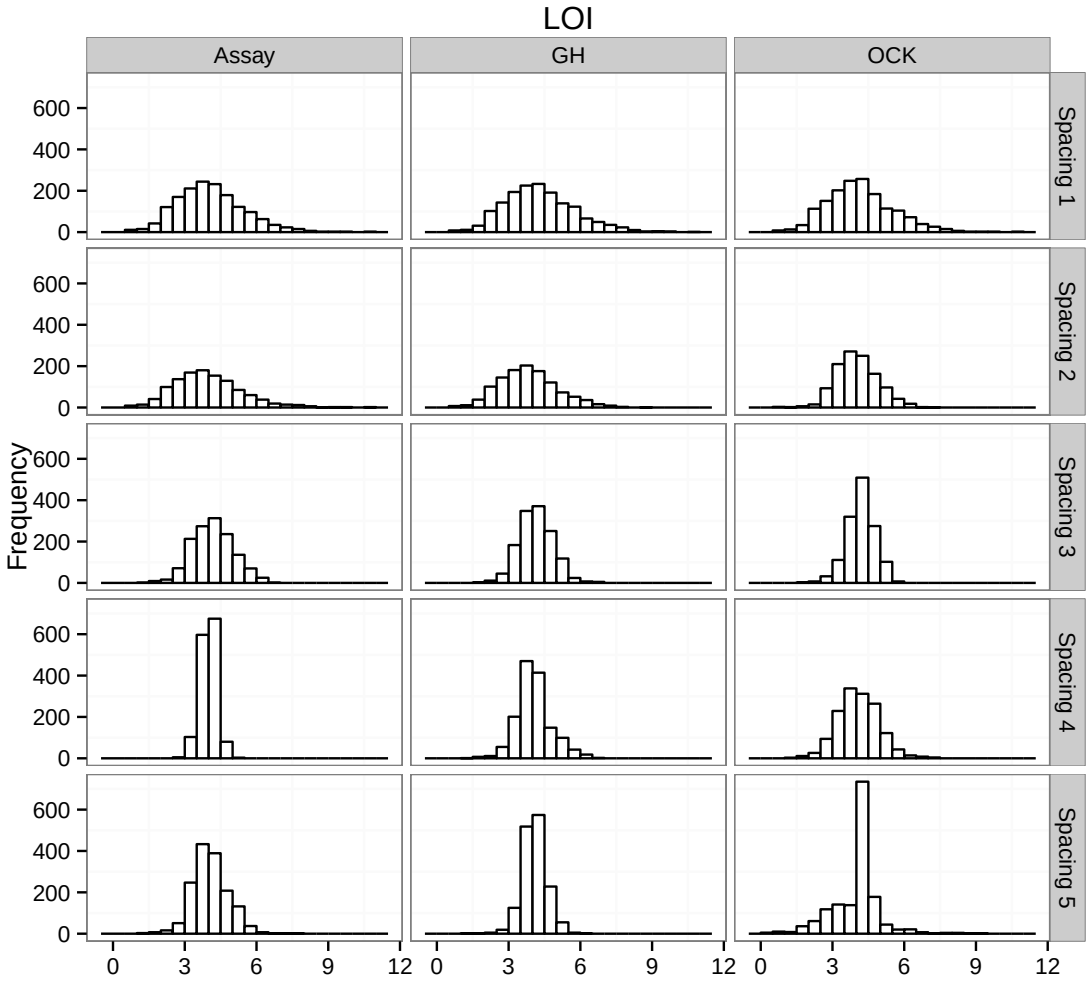


Figure K.3: Sample and Estimate histograms *LOI* plots

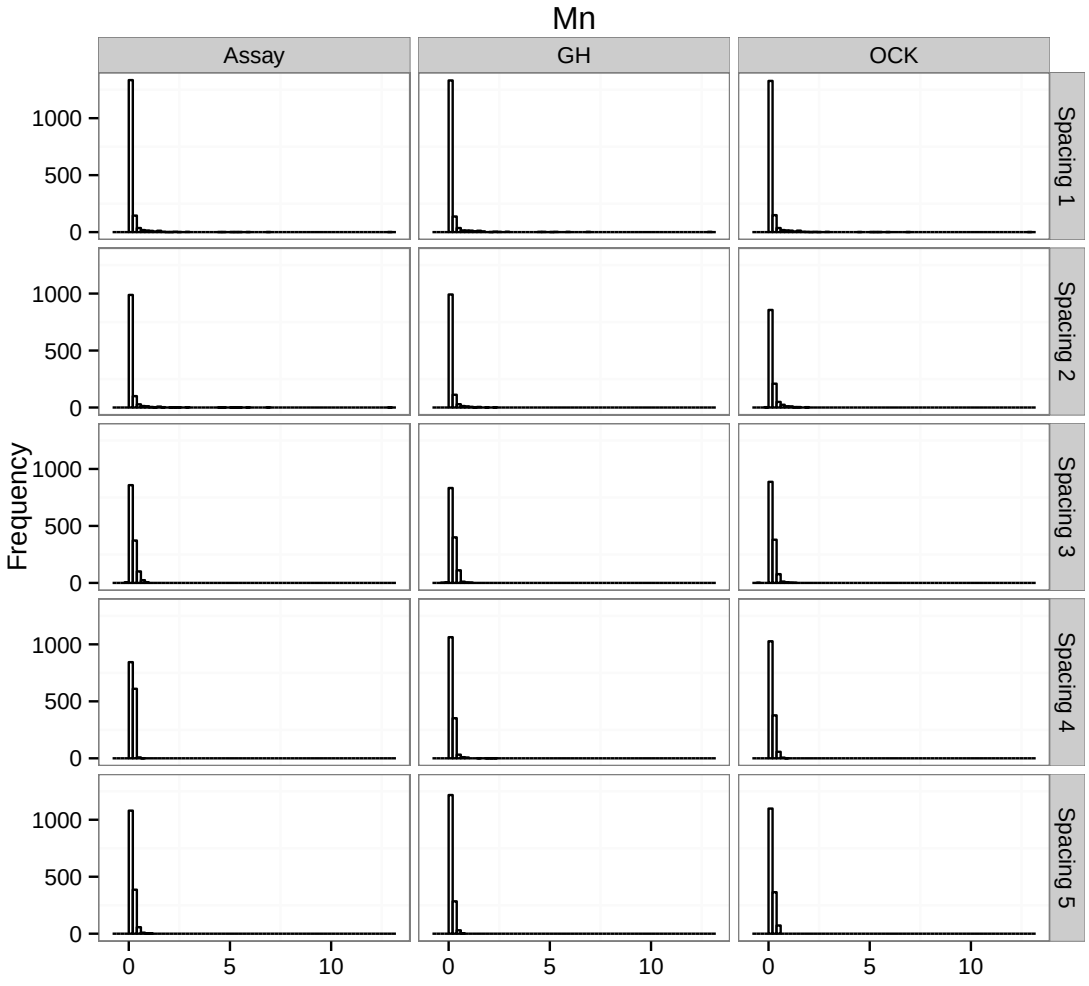


Figure K.4: Sample and Estimate histograms *Mn* plots

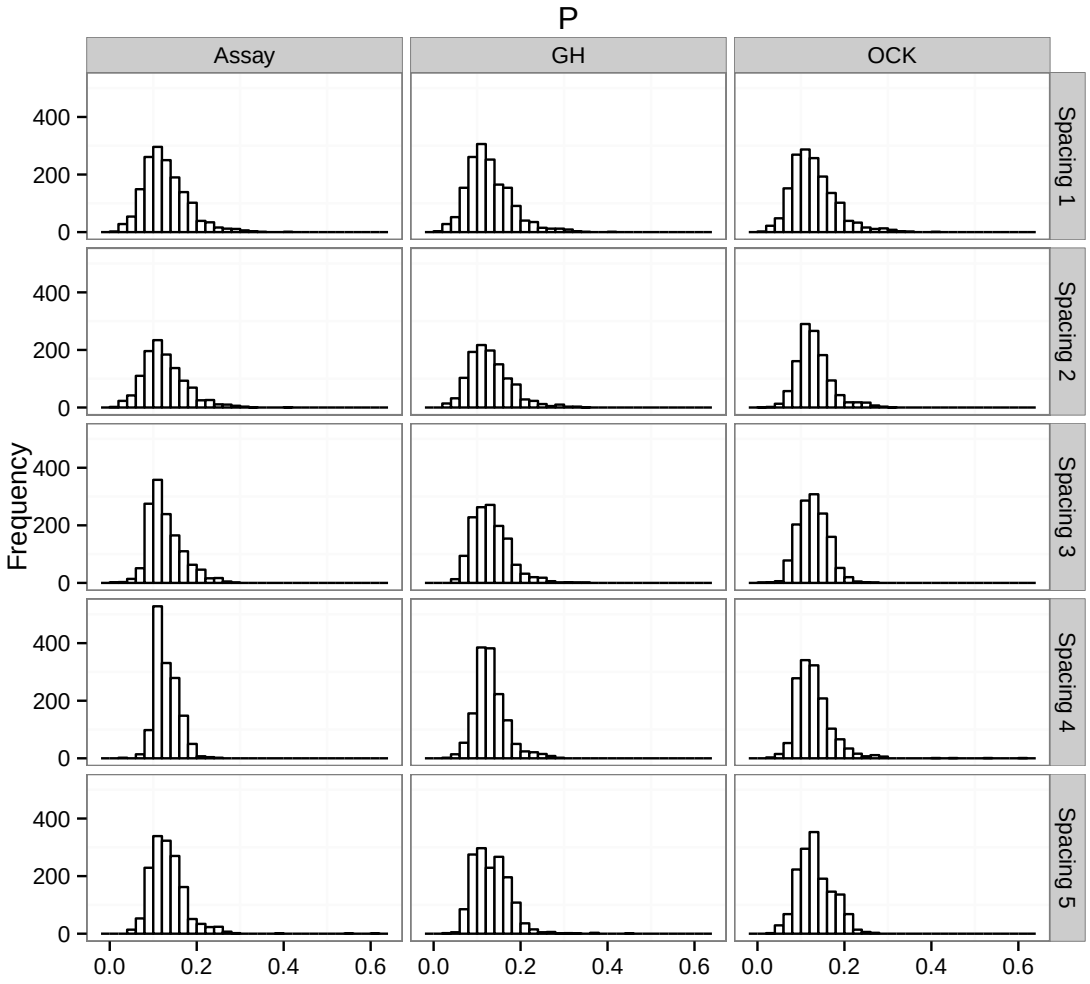


Figure K.5: Sample and Estimate histograms *P* plots

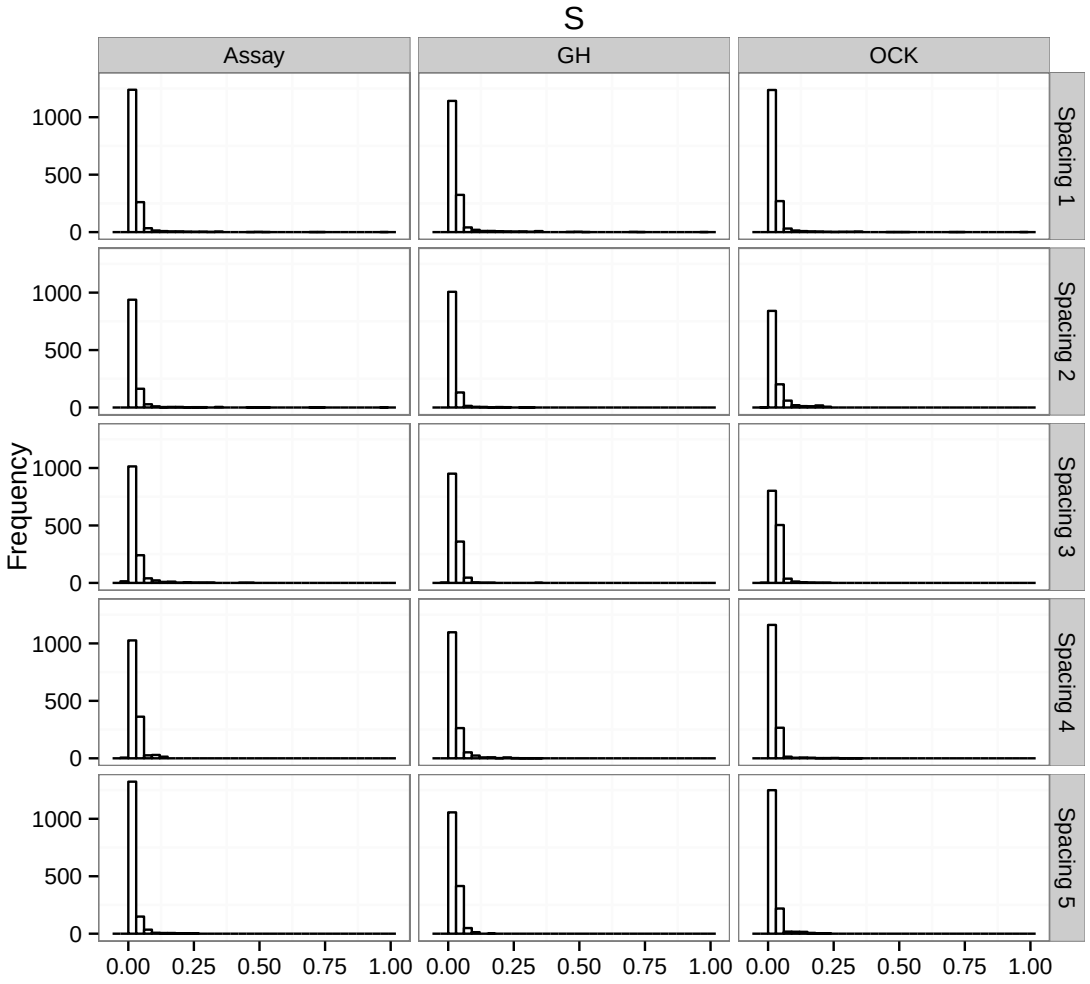


Figure K.6: Sample and Estimate histograms *S* plots

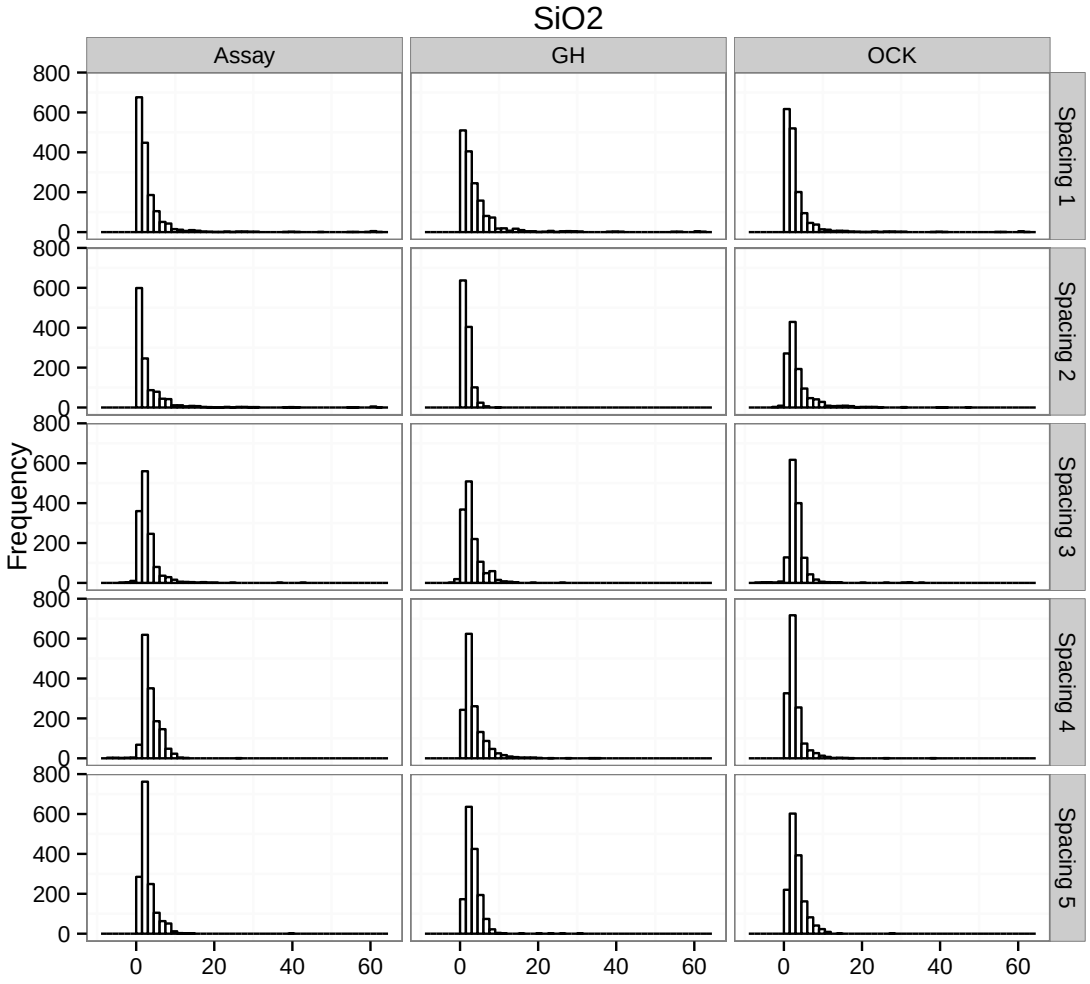


Figure K.7: SiO_2 plots

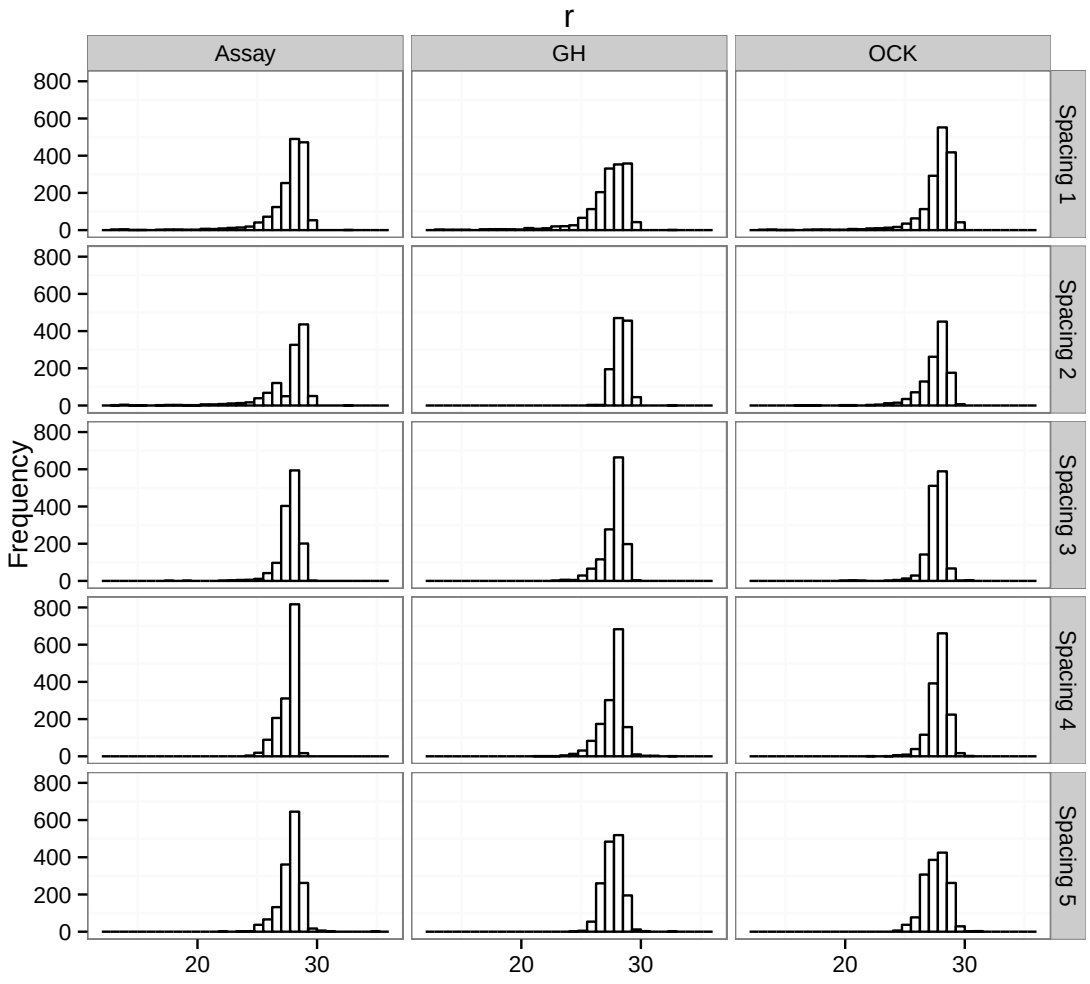


Figure K.8: *Filler* plots

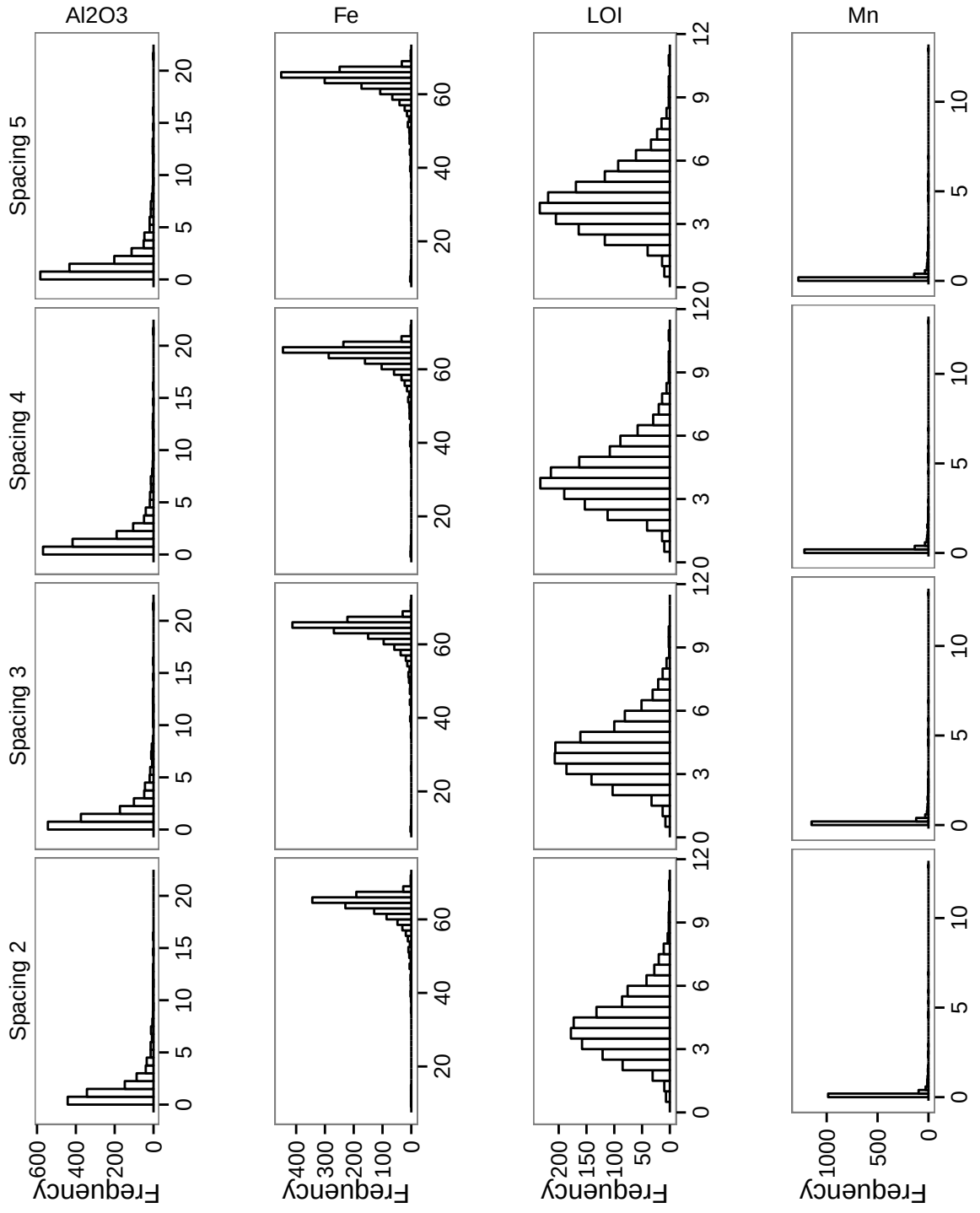


Figure K.9: Histograms of Jack-knife informing data, Al_2O_3, Fe, LOI, and Mn

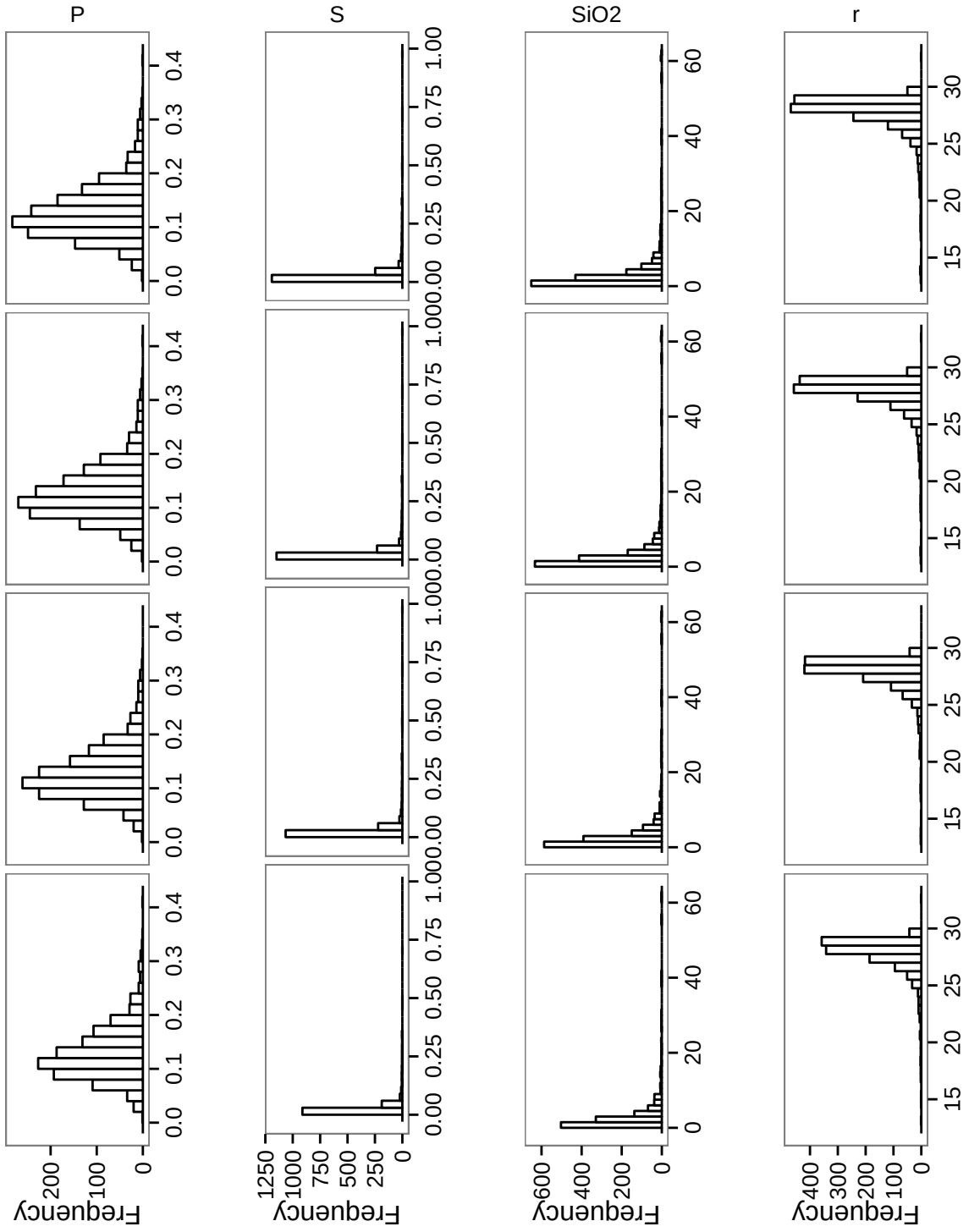


Figure K.10: Histograms of Jack-knife informing data, P, S, SiO_2, and $Filler$

Appendix L

Scatter plots - Univariate errors

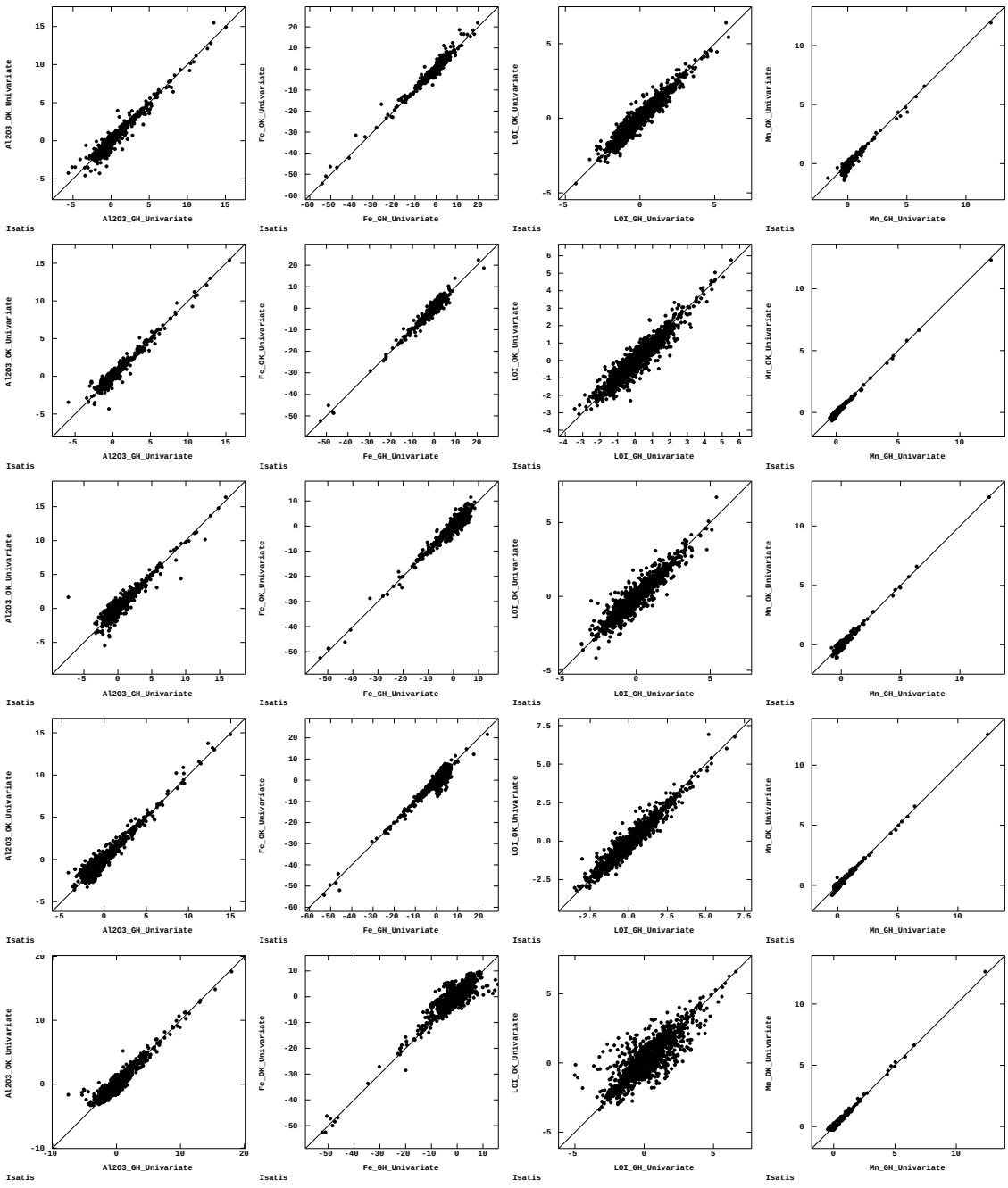


Figure L.1: Error Scatter plots - By column Al_2O_3 far left, *Fe* near left, *LOI* near right, and *Mn* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the GH distribution, the ordinate is the OK distribution.

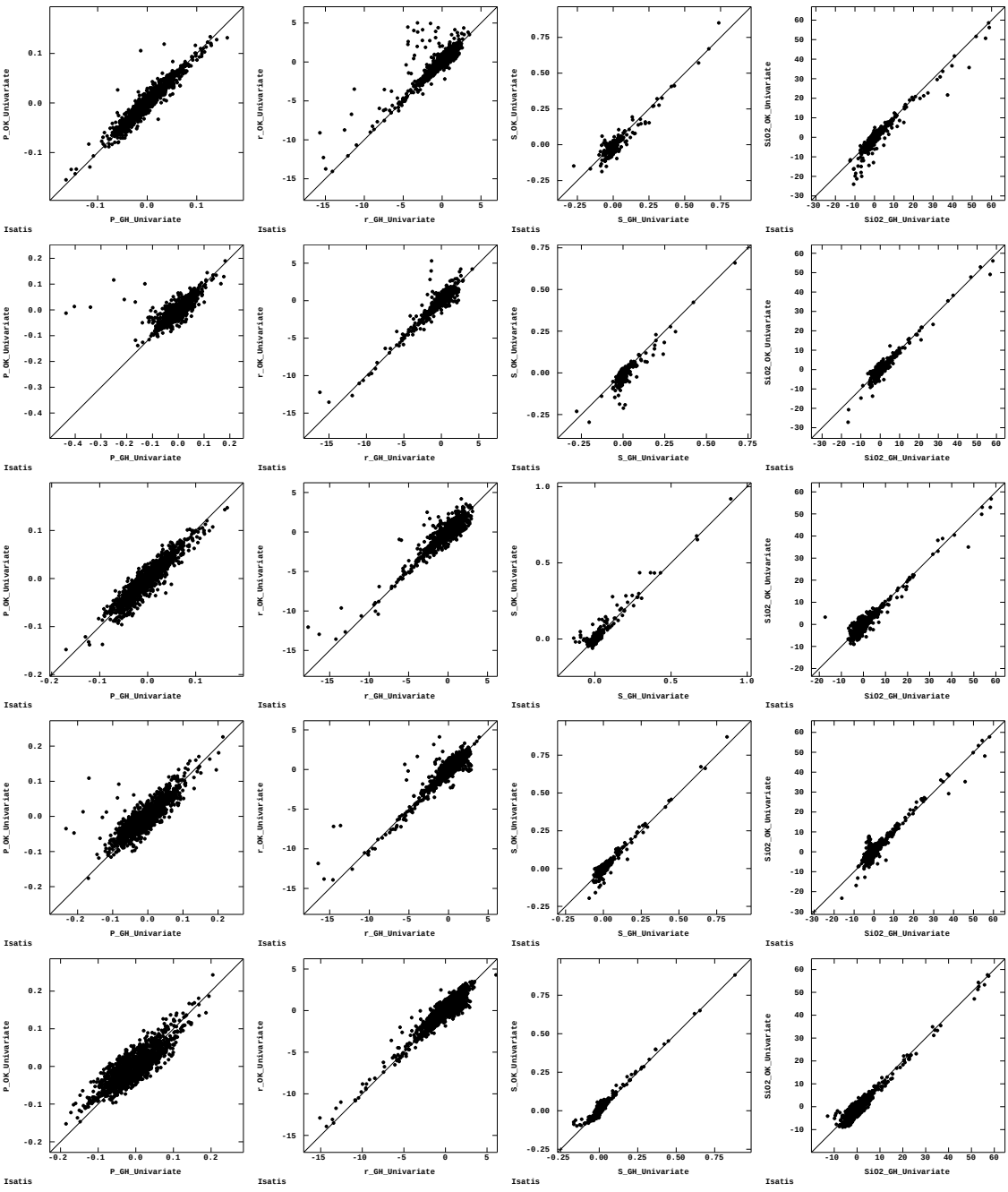


Figure L.2: Error scatter plots - By column *P* far left, *Filler* near left, *S* near right, and SiO_2 far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom. The abscissa in each case is the GH distribution, the ordinate is the OK distribution.

Appendix M

Histograms - Univariate errors

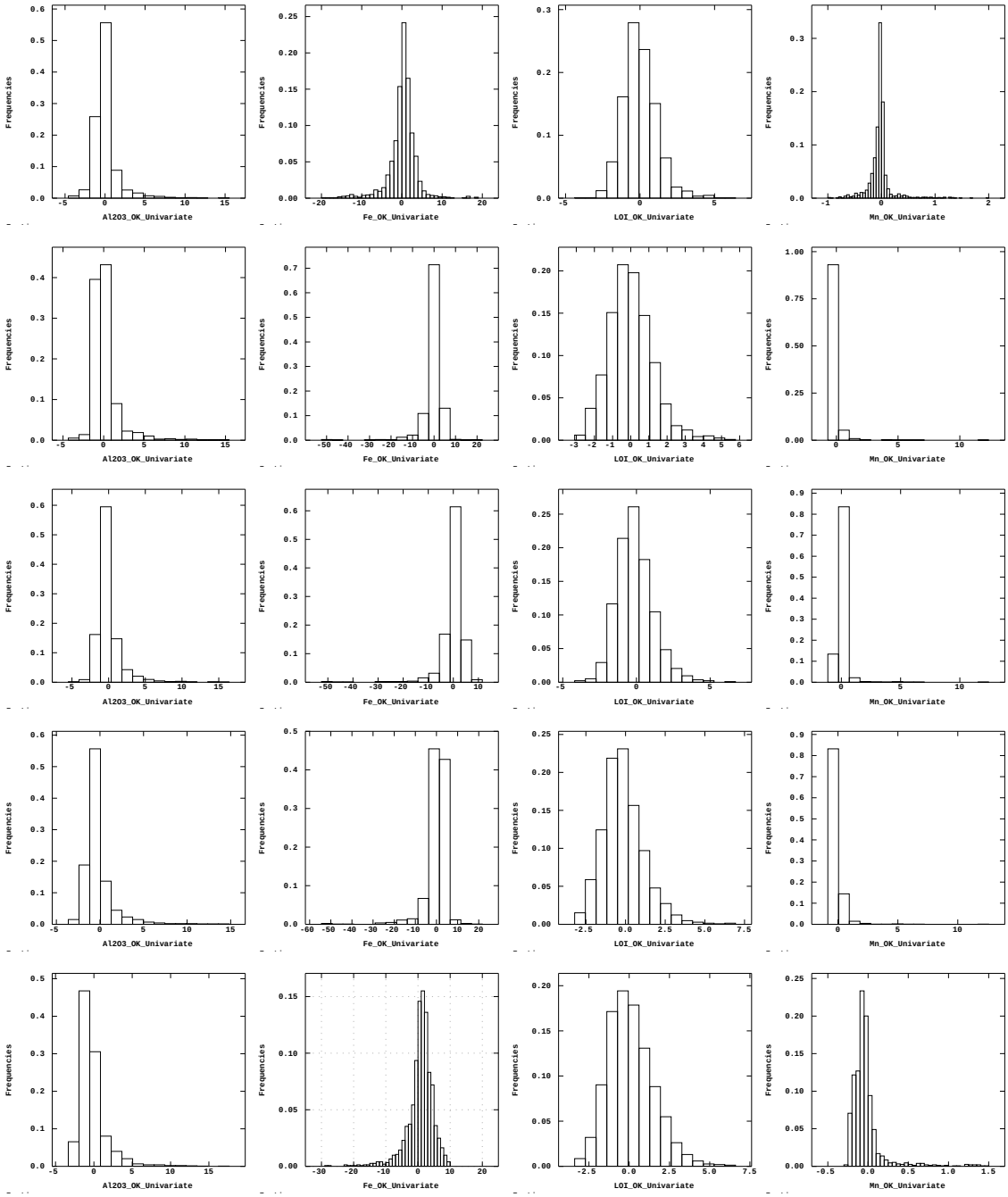


Figure M.1: Histograms of univariate errors generated from the OCK method - By column Al_2O_3 far left, *Fe* near left, *LOI* near right, and *Mn* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom.

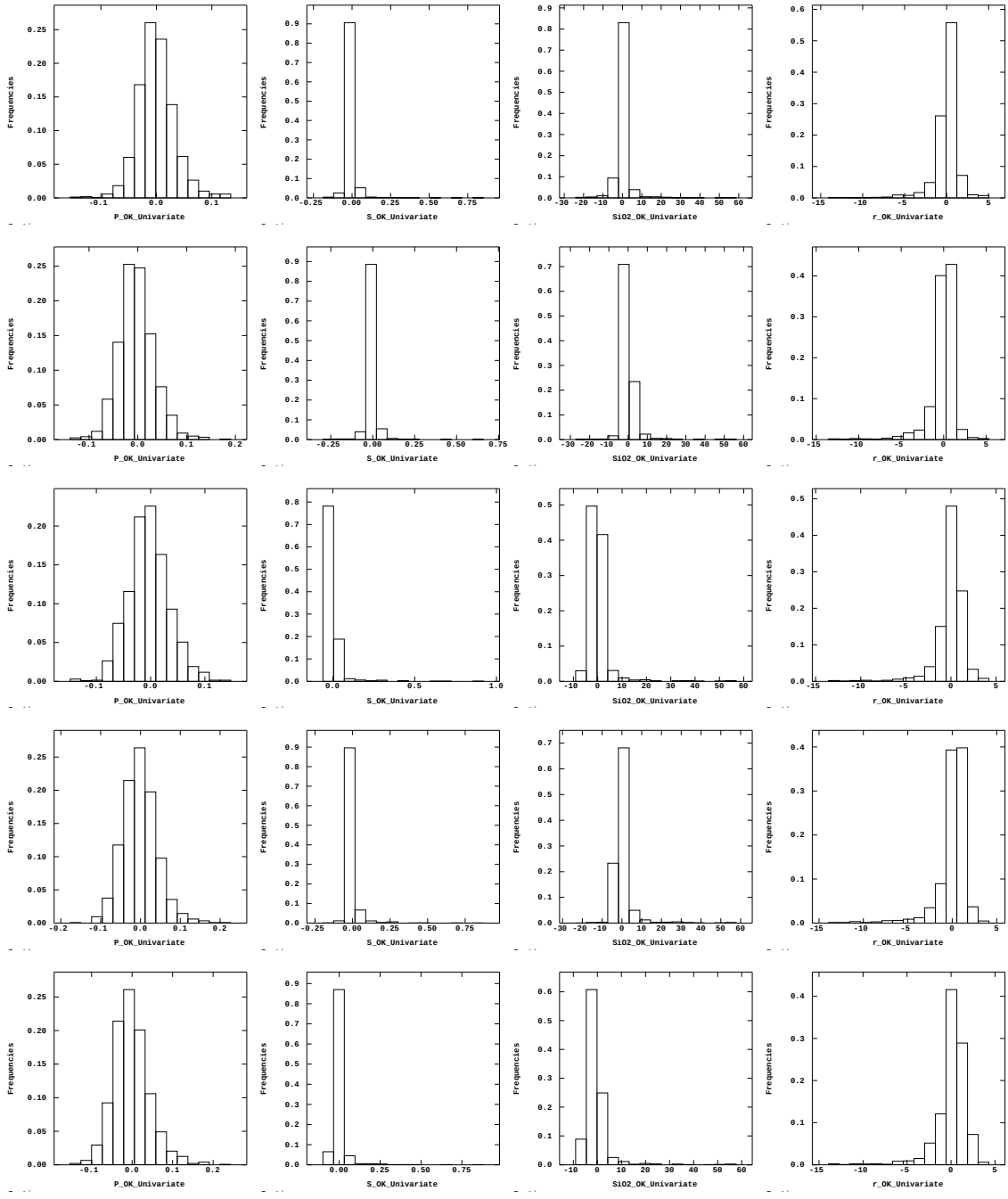


Figure M.2: Histograms of univariate errors generated from the OCK method - By column *P* far left, *Filler* near left, *S* near right, and SiO_2 far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom.

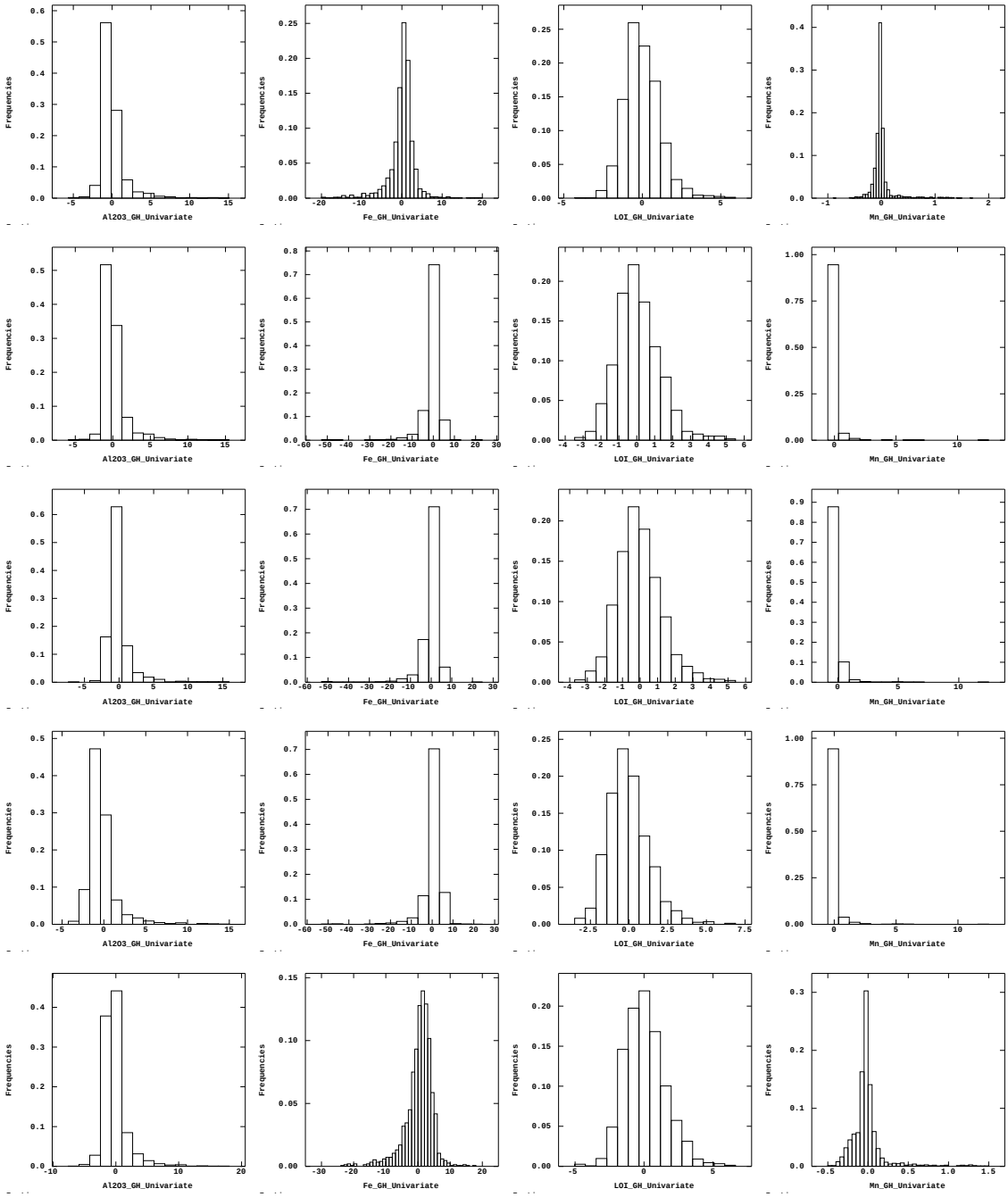


Figure M.3: Histograms of univariate errors generated from the GH method - By column Al_2O_3 far left, *Fe* near left, *LOI* near right, and *Mn* far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom.

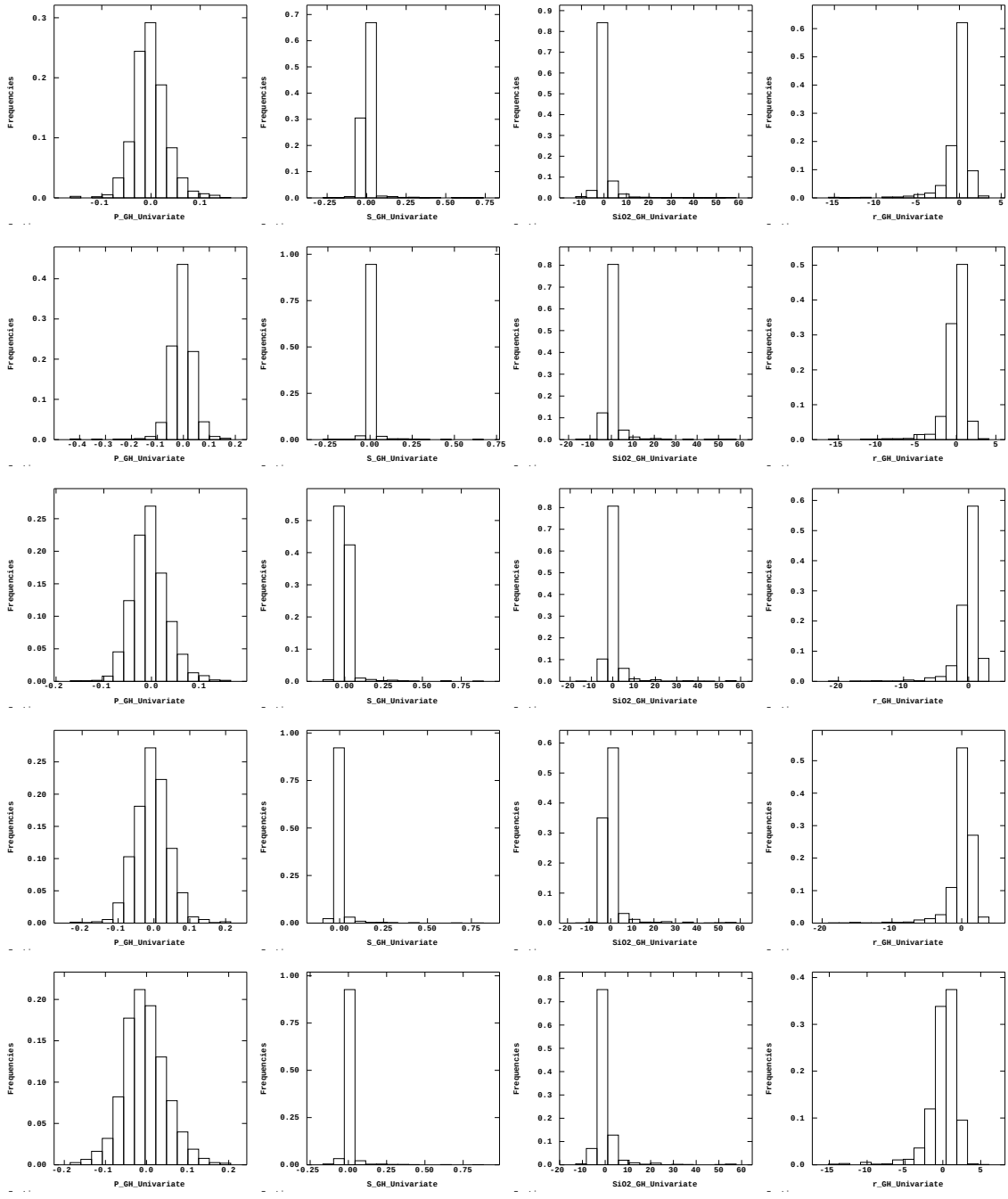


Figure M.4: Histograms of univariate errors generated from the GH method - By column *P* far left, *Filler* near left, *S* near right, and SiO_2 far right in row order of decreasing data density, Spacing 1 on the top and Spacing 5 at the bottom.

Appendix N

Histograms - Compositional errors

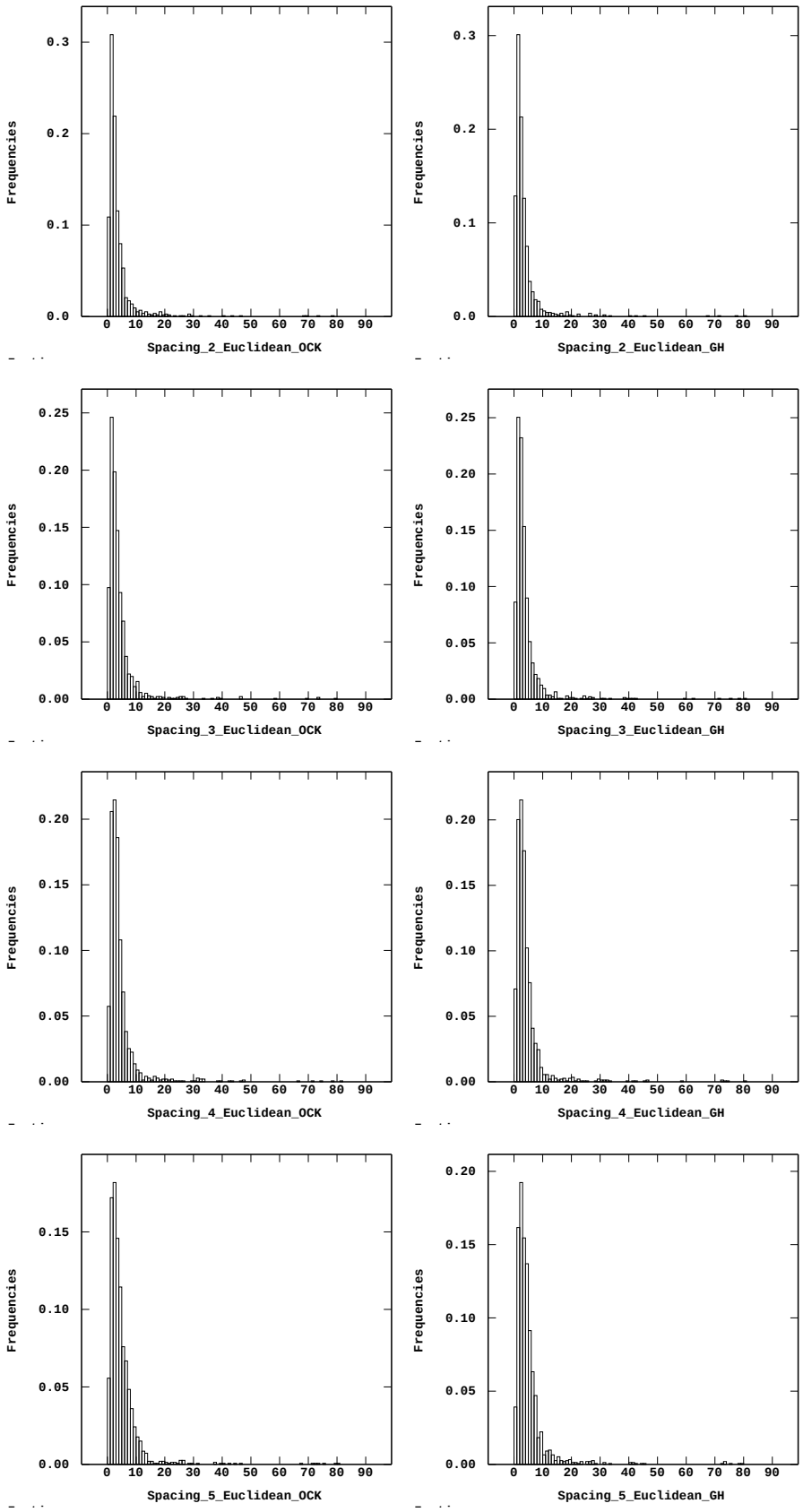


Figure N.1: Histograms of Euclidean errors - OCK estimates on the left and GH estimates on the right, in order of decreasing spatial density Spacing 2 on top, Spacing 5 on bottom.

.

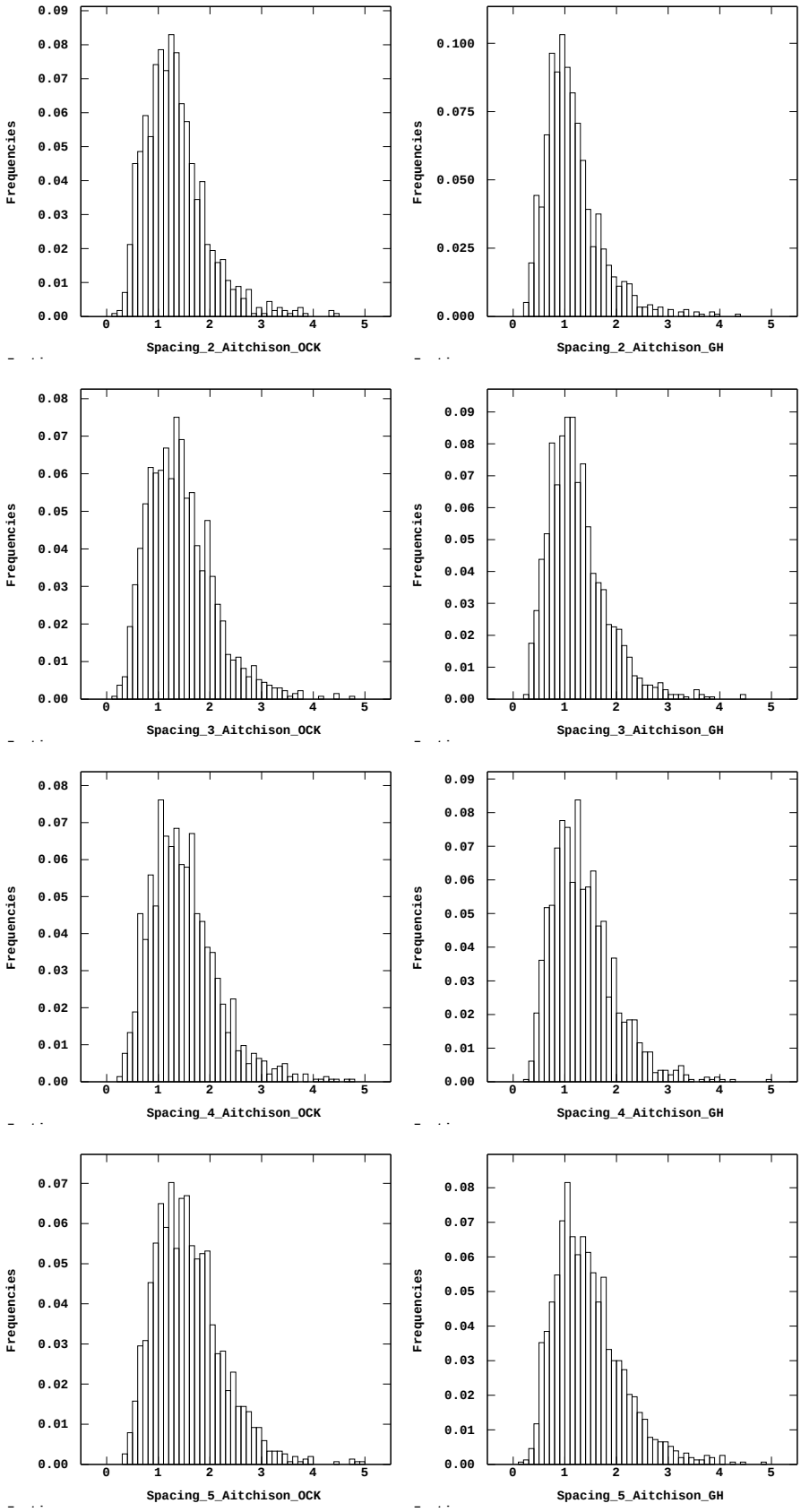


Figure N.2: Histograms of Aitchison errors - OCK estimates on the left and GH estimates on the right, in order of decreasing spatial density Spacing 2 on top, Spacing 5 on bottom.

.

Appendix O

Univariate error base maps

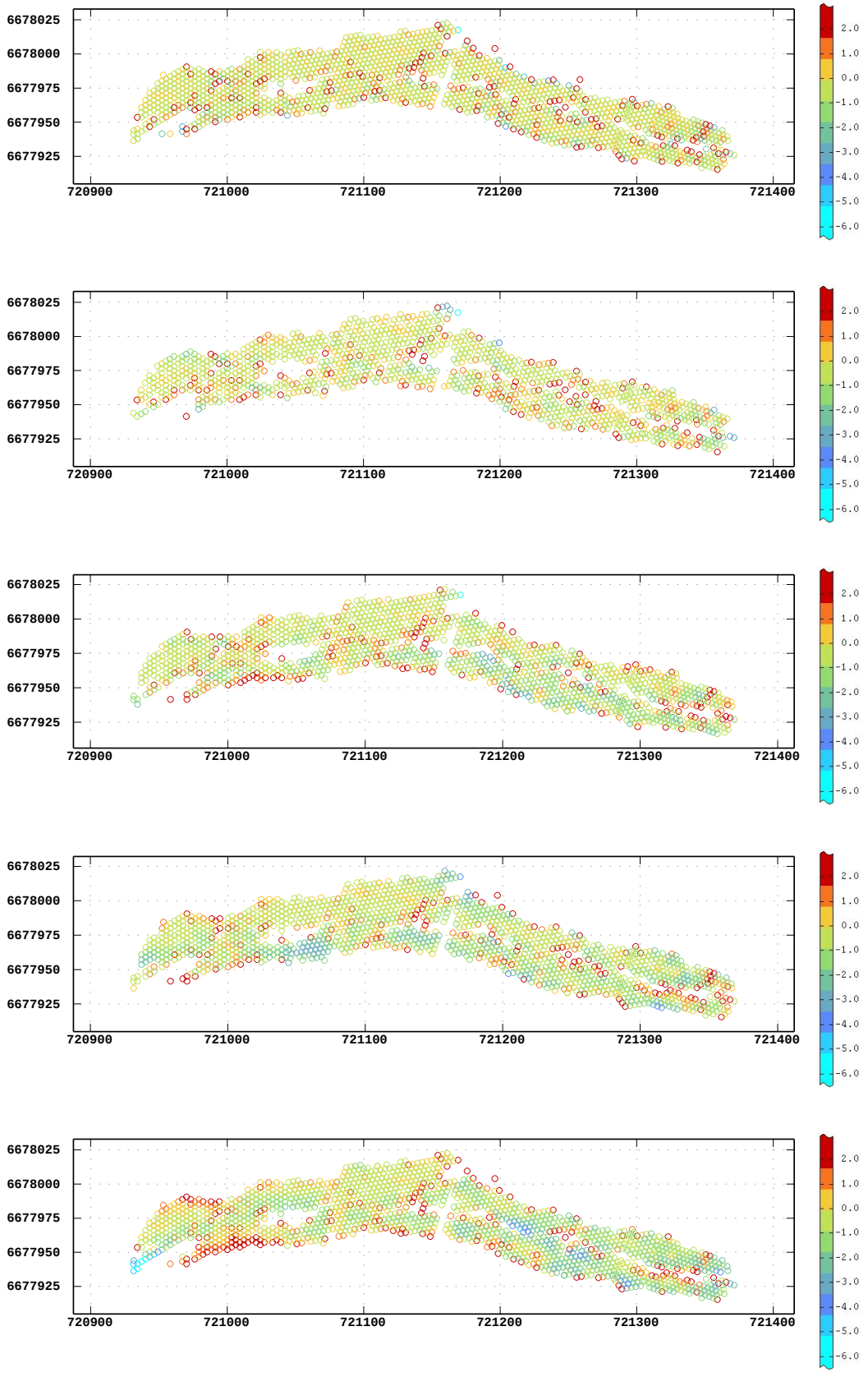


Figure O.1: Base maps of univariate Al_2O_3 errors generated by GH.

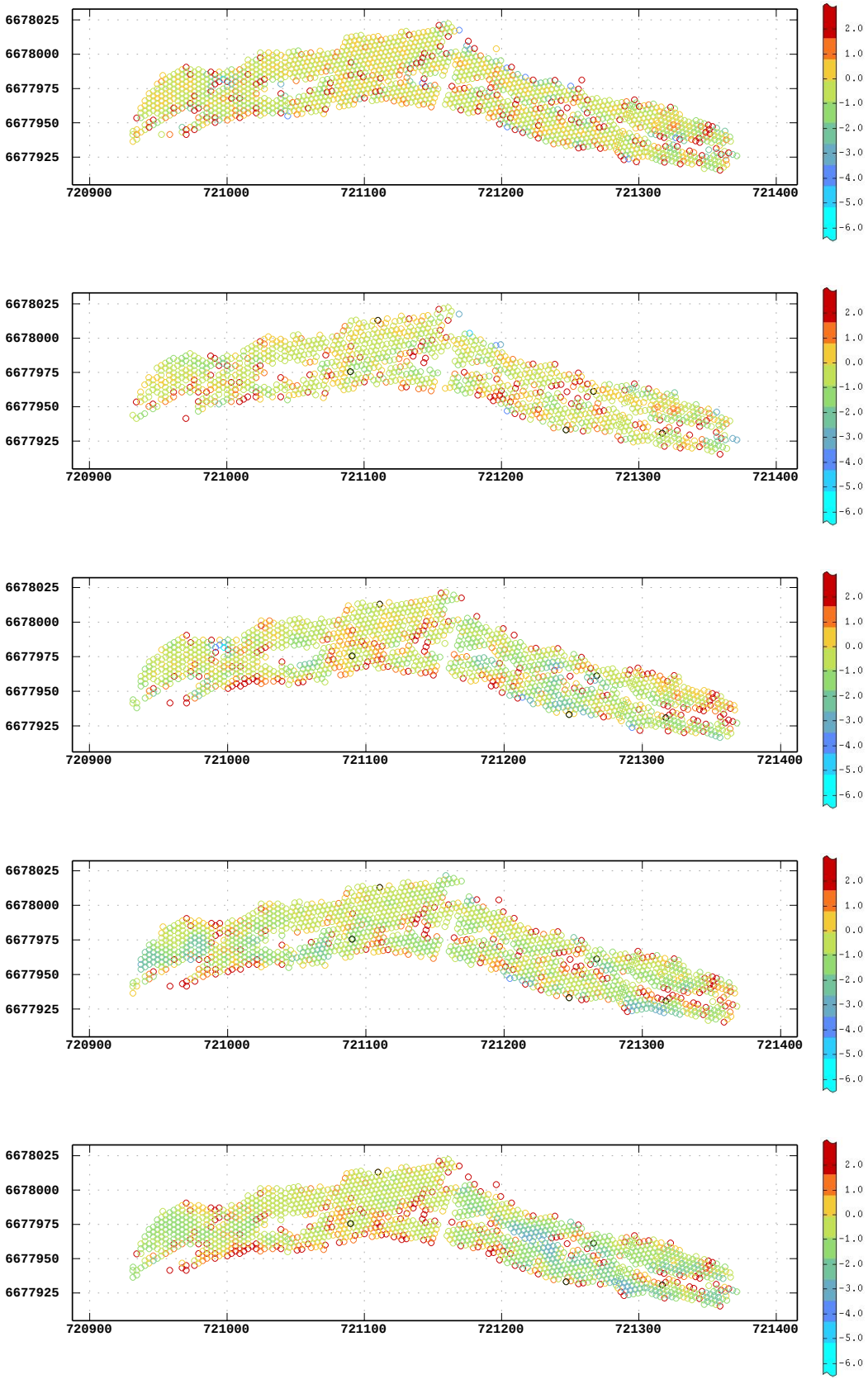


Figure O.2: Base maps of univariate Al_2O_3 errors generated by OCK.

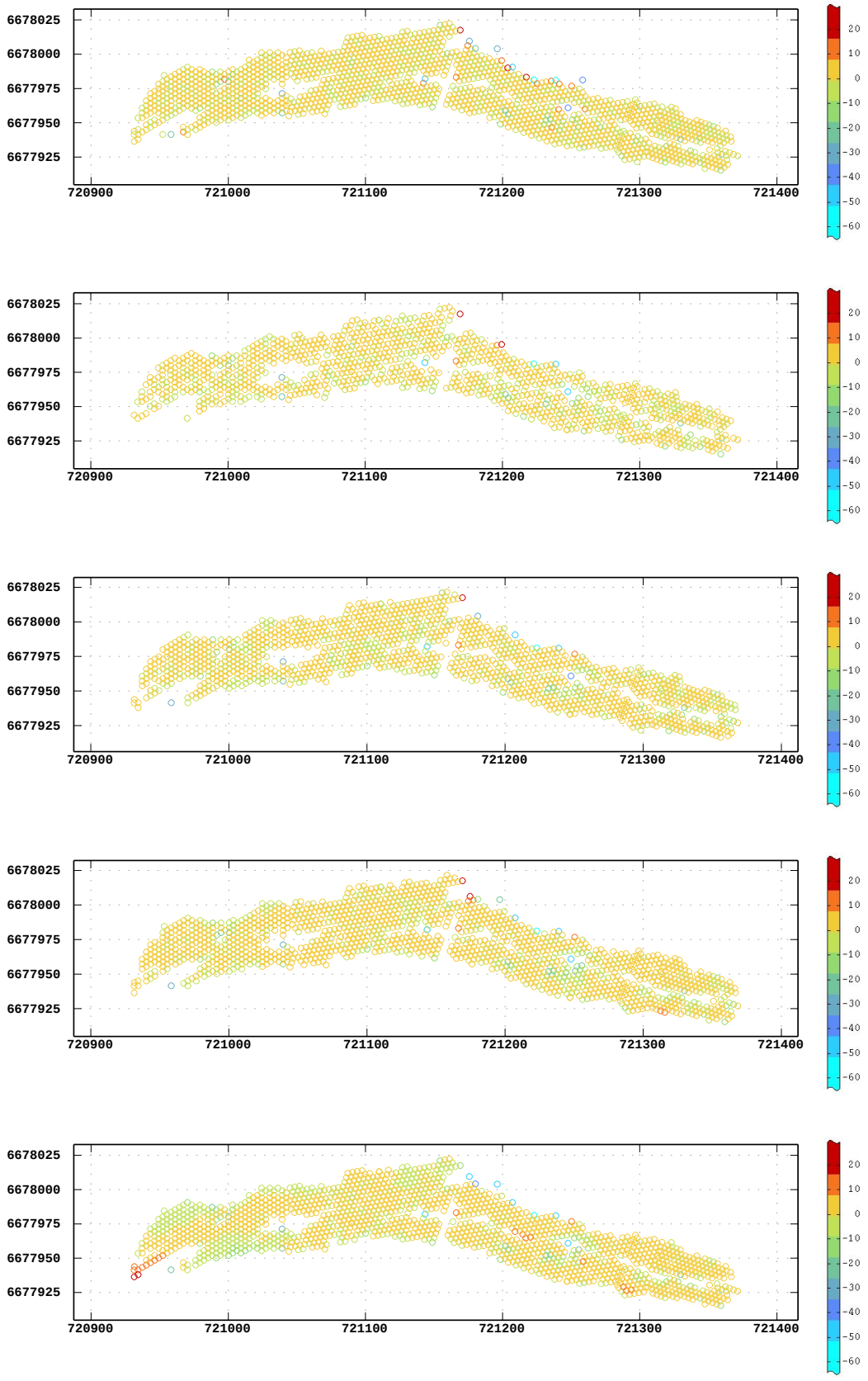


Figure O.3: Base maps of univariate *Fe* errors generated by GH.

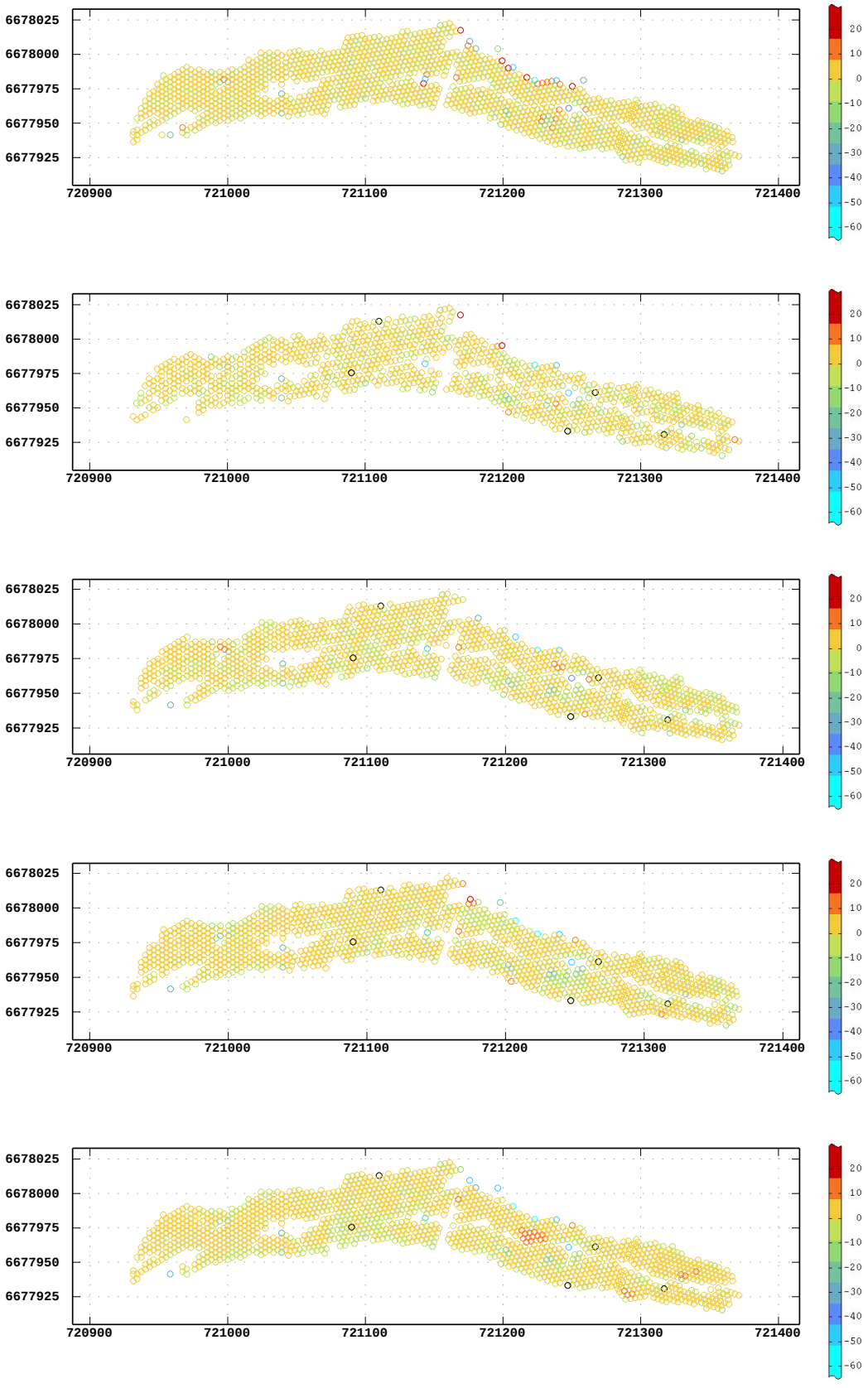


Figure O.4: Base maps of univariate *Fe* errors generated by OCK.

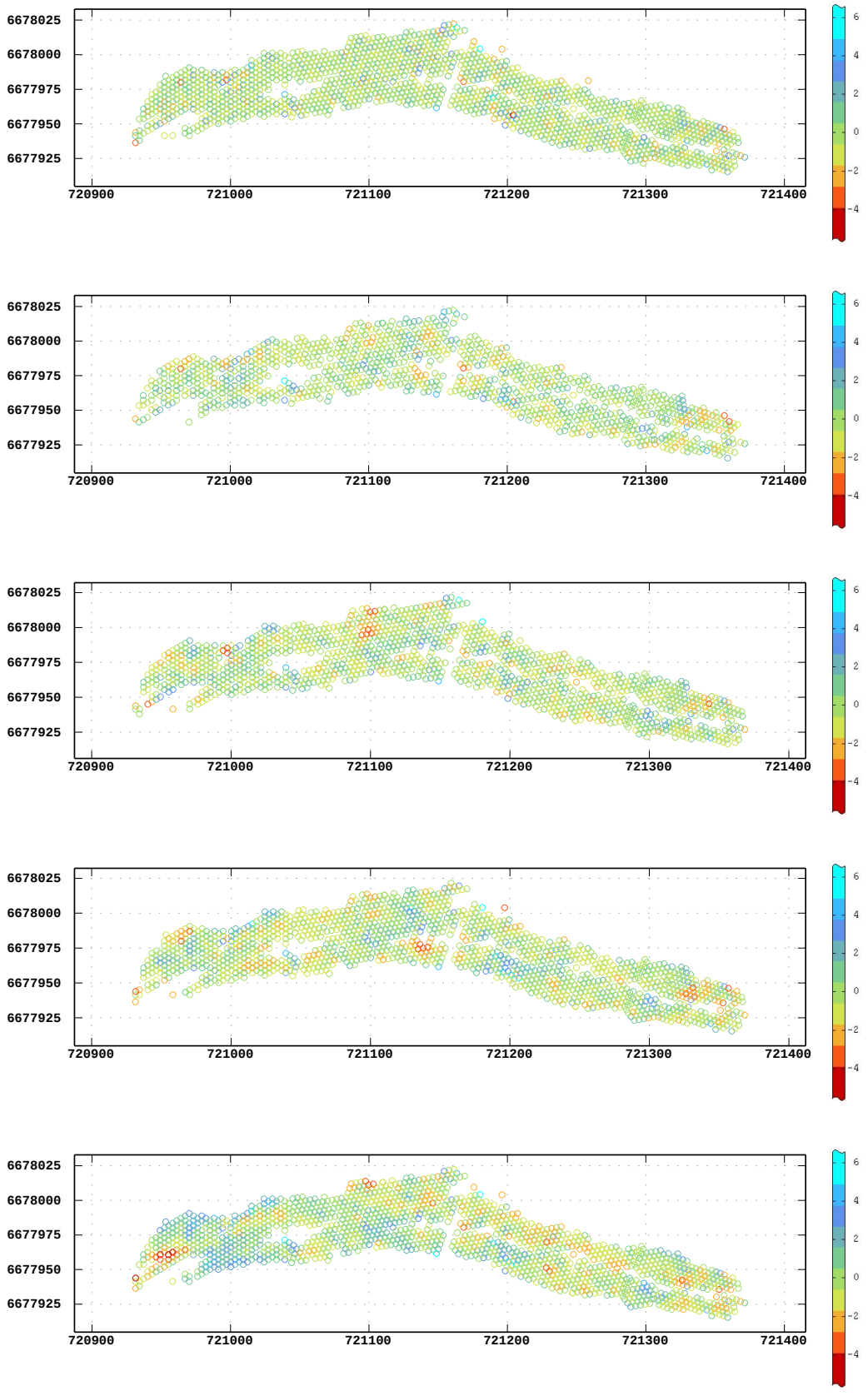


Figure O.5: Base maps of univariate *LOI* errors generated by GH.

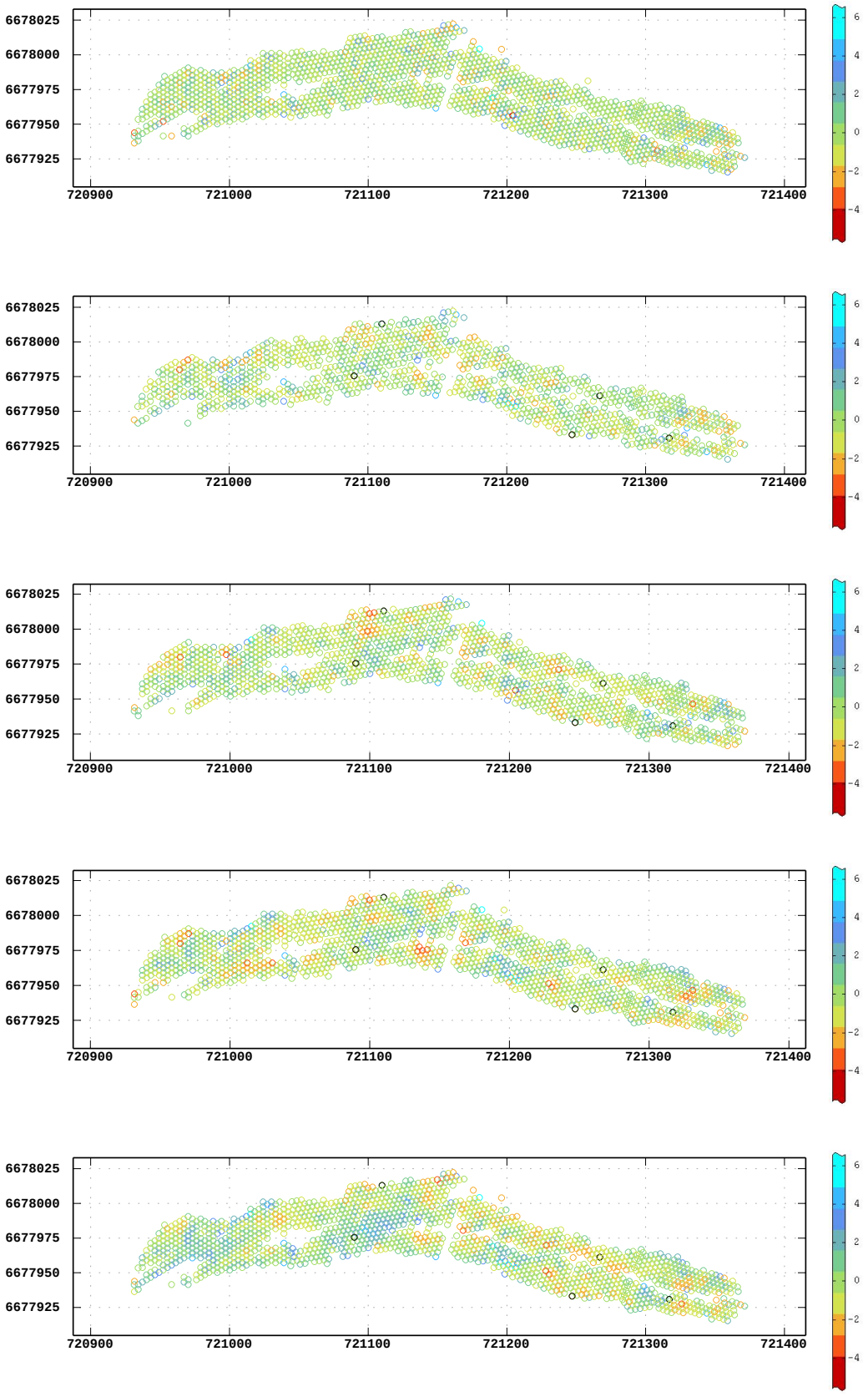


Figure O.6: Base maps of univariate *LOI* errors generated by OCK.

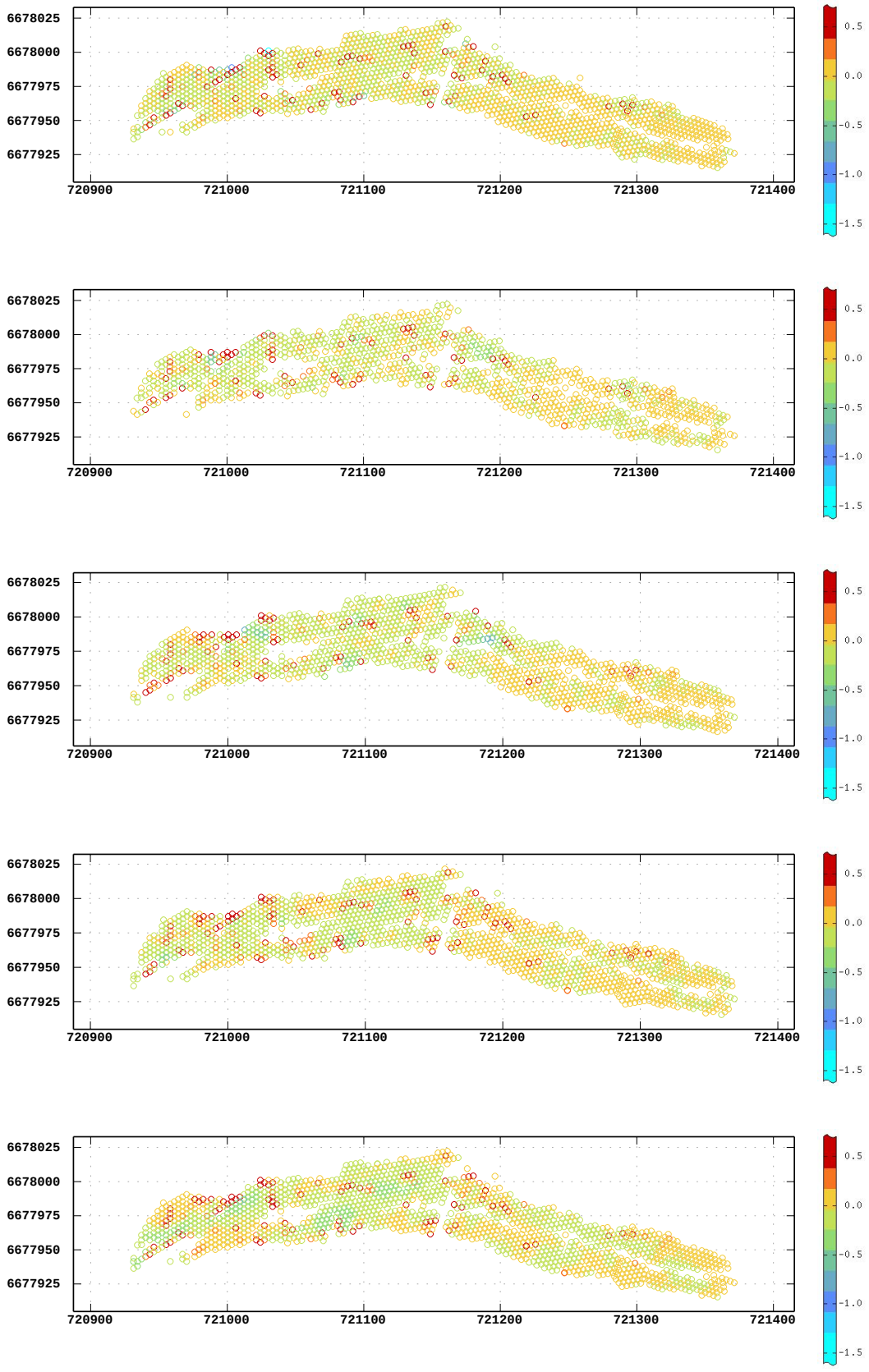


Figure O.7: Base maps of univariate *Mn* errors generated by GH.

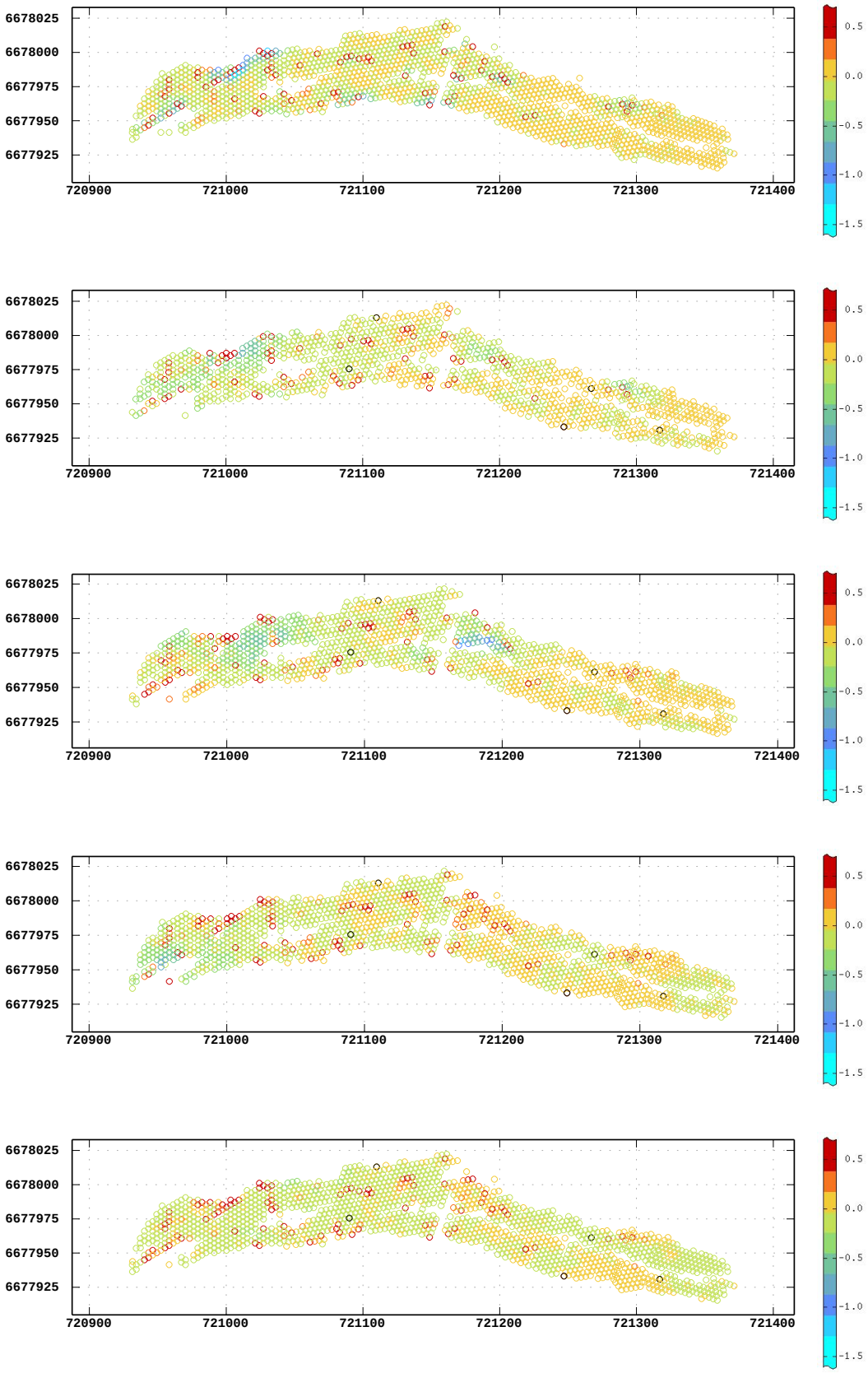


Figure O.8: Base maps of univariate *Mn* errors generated by OCK.

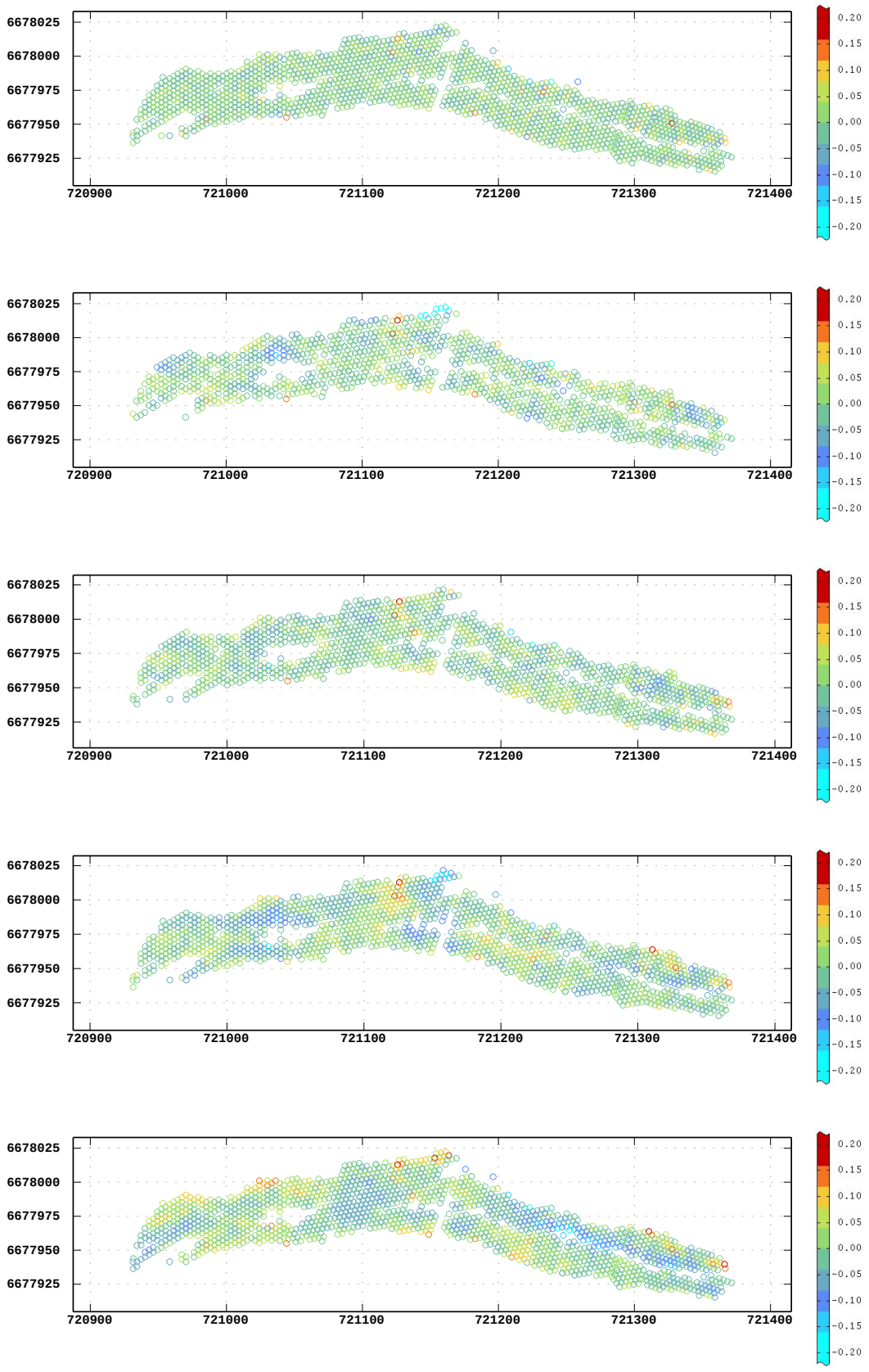


Figure O.9: Base maps of univariate *P* errors generated by GH.

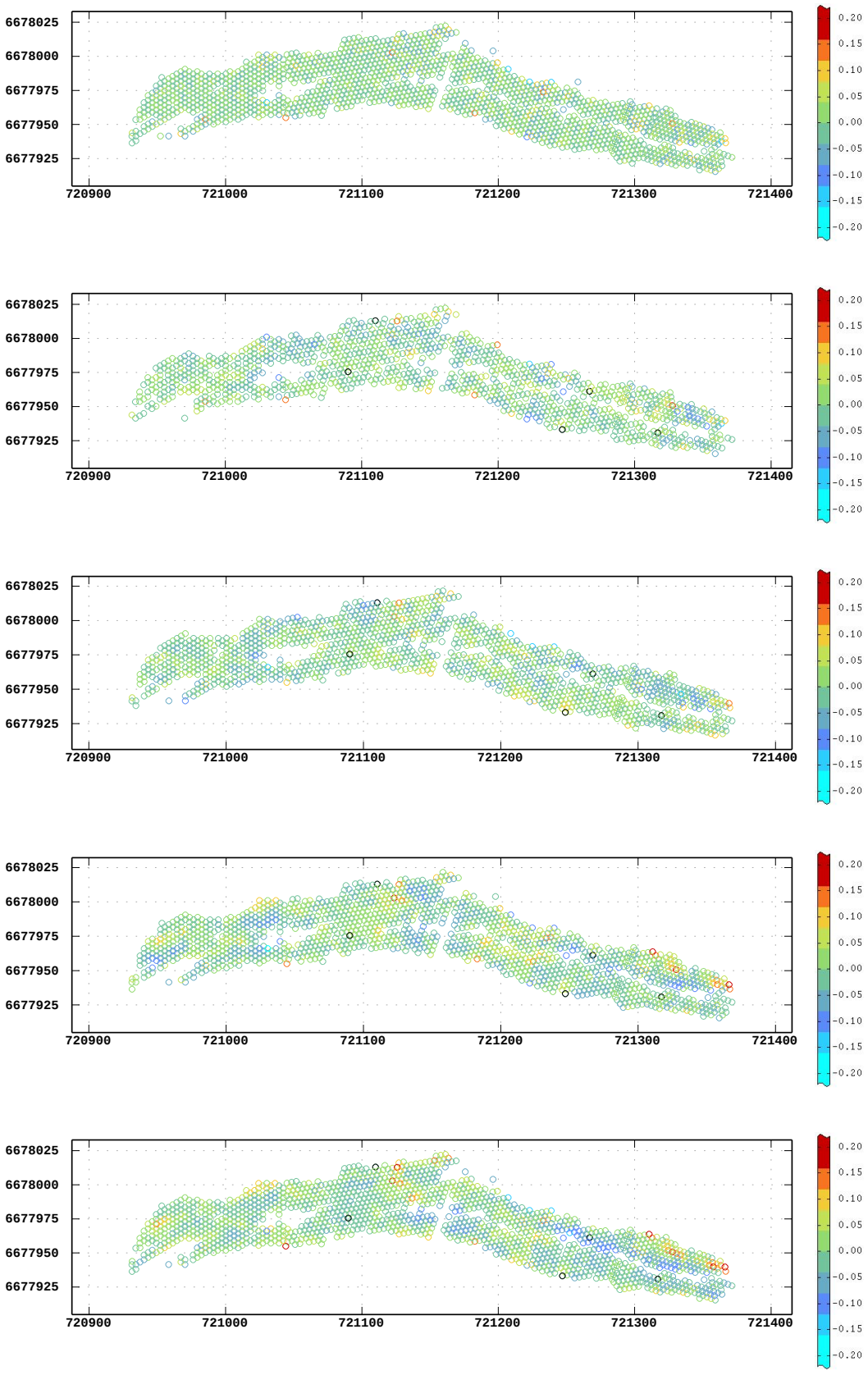


Figure O.10: Base maps of univariate *P* errors generated by OCK.

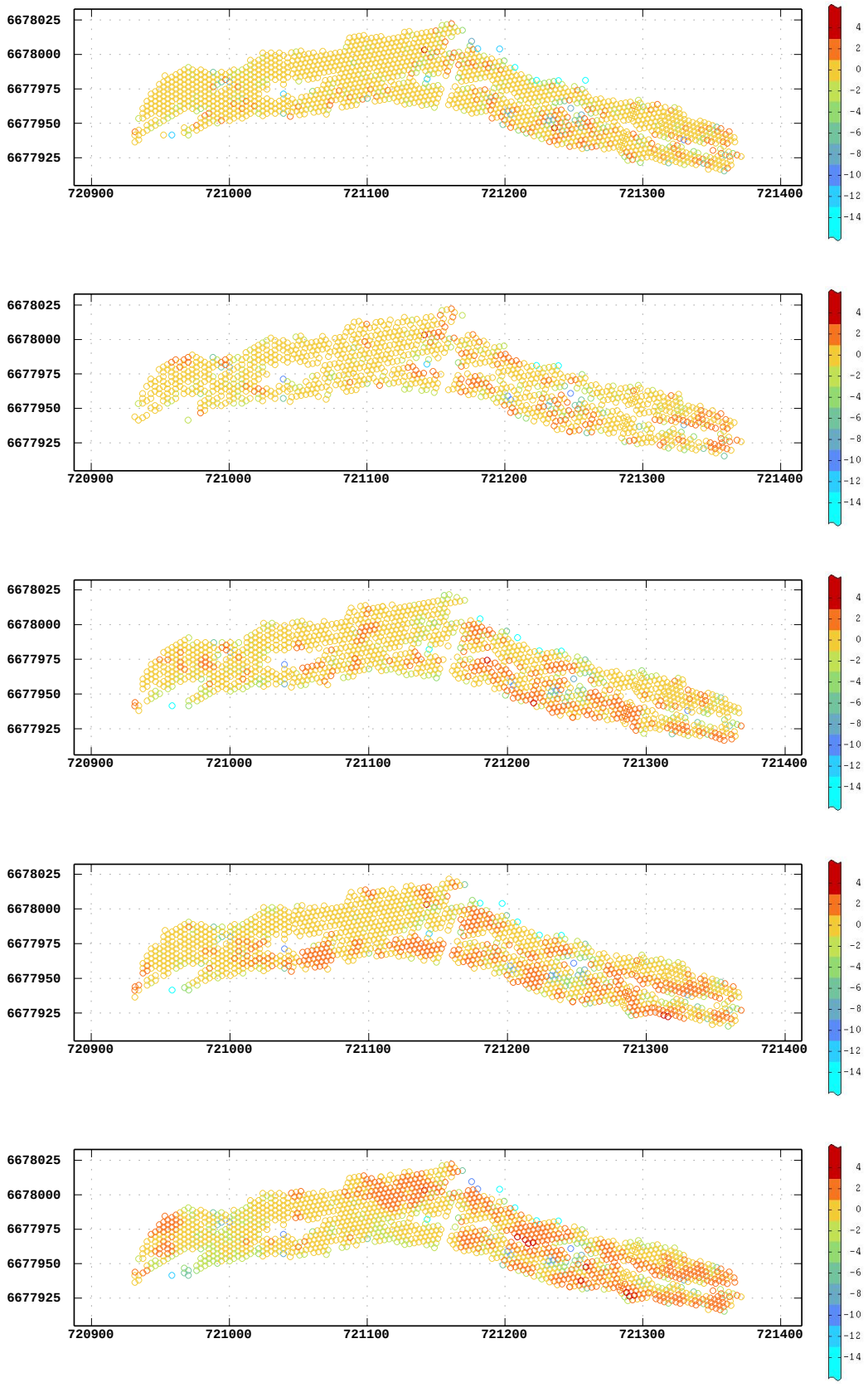


Figure O.11: Base maps of univariate *filler* errors generated by GH.

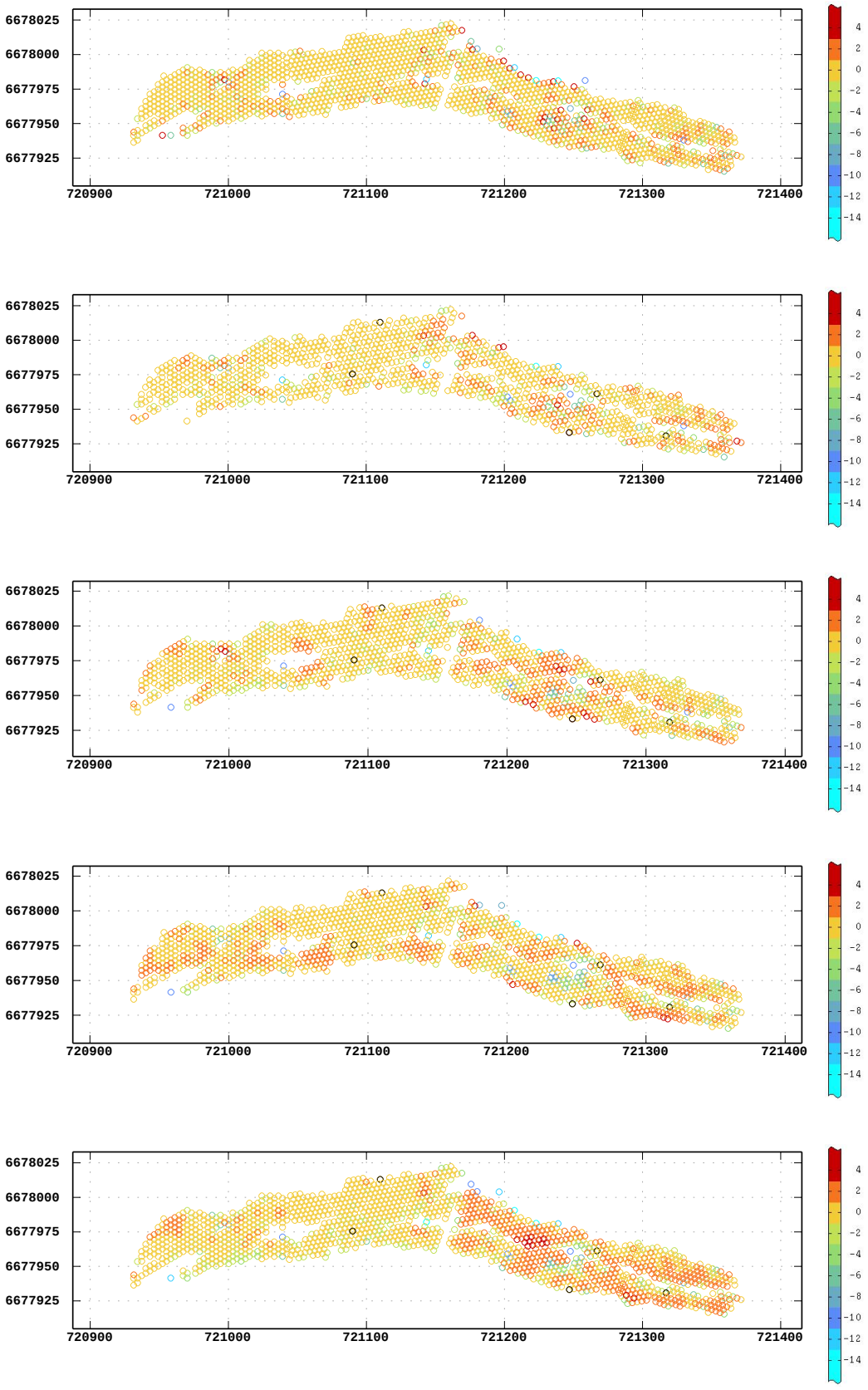


Figure O.12: Base maps of univariate *filler* errors generated by OCK.

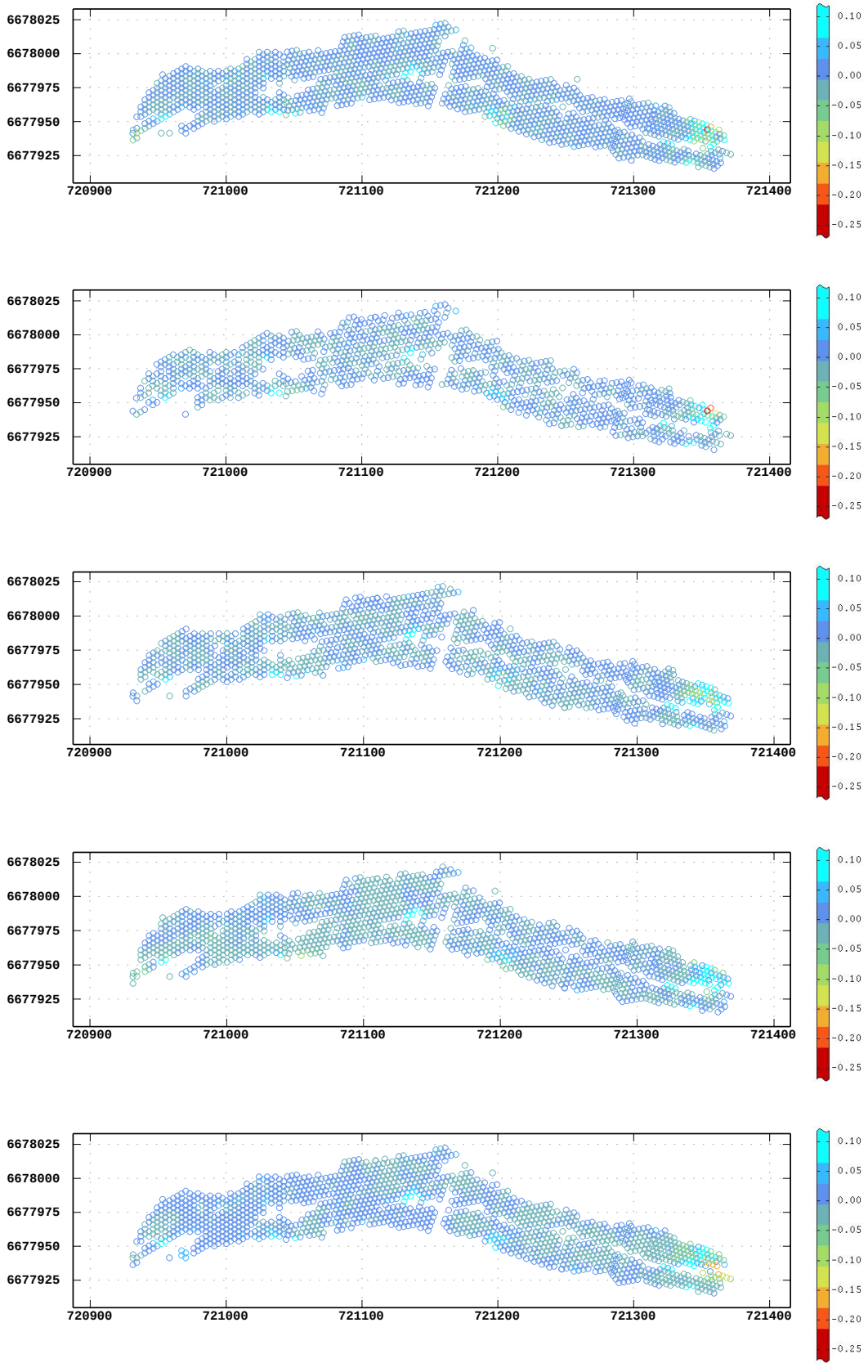


Figure O.13: Base maps of univariate *S* errors generated by GH.

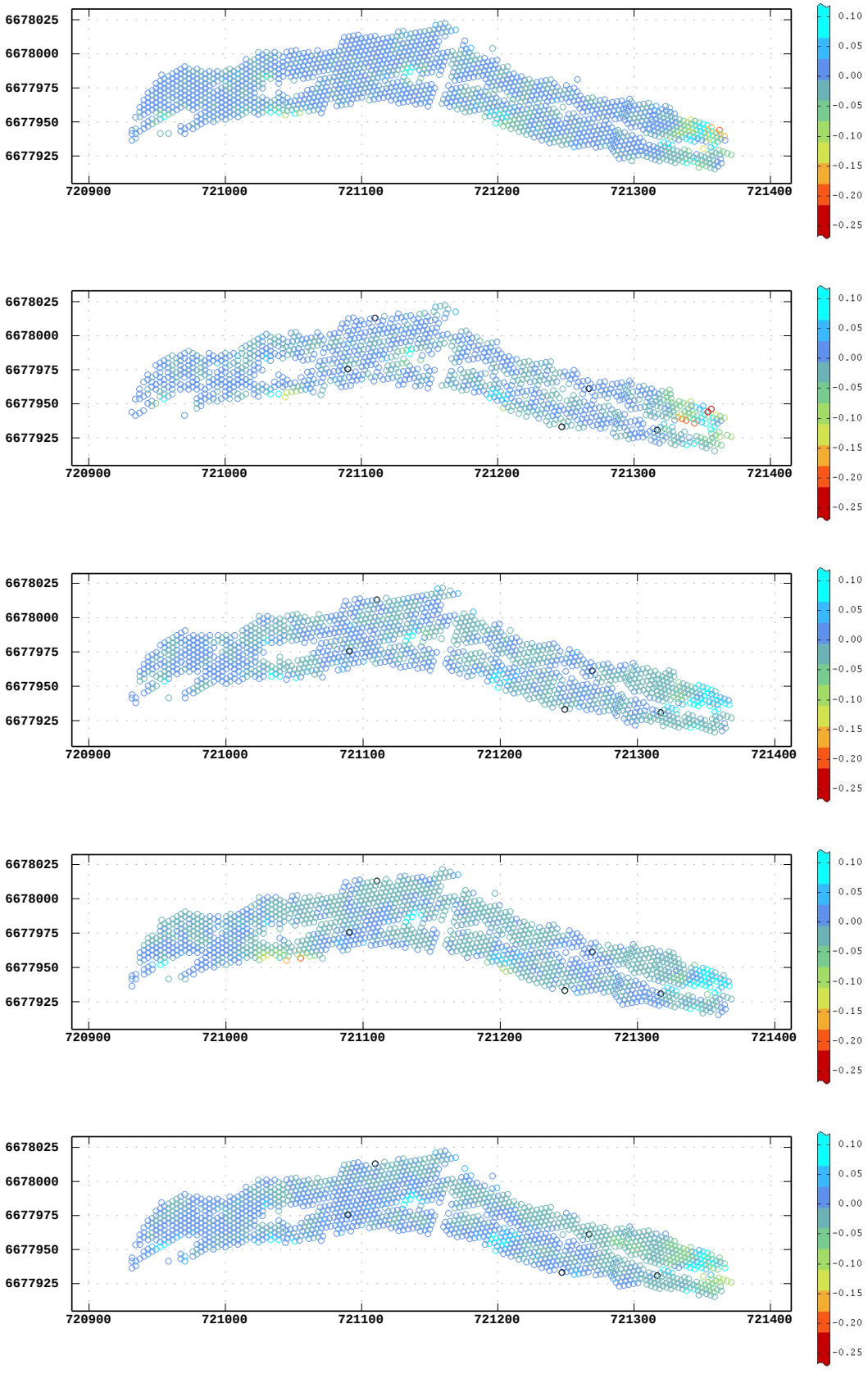


Figure O.14: Base maps of univariate *S* errors generated by OCK.

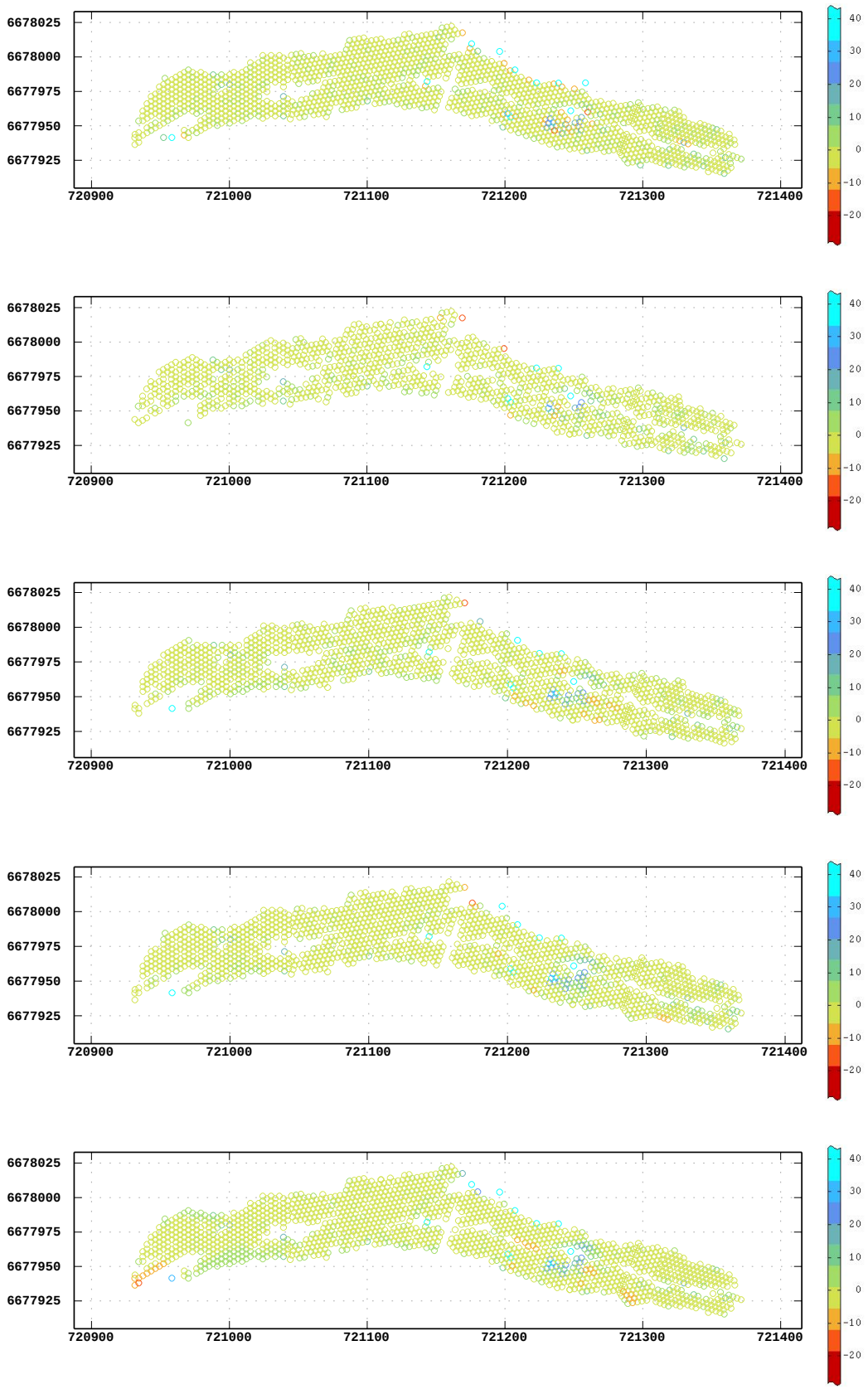


Figure O.15: Base maps of univariate SiO_2 errors generated by GH.

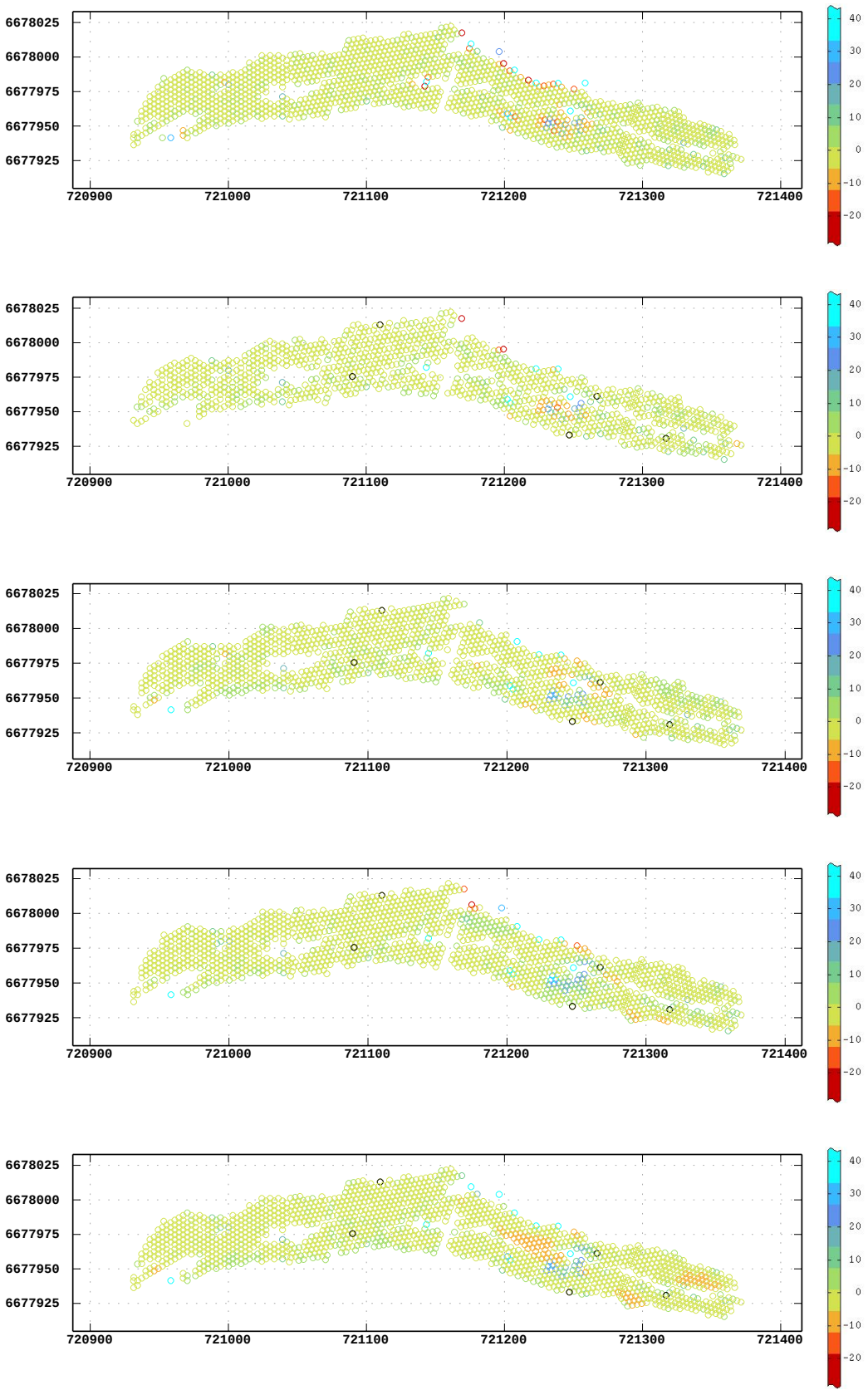


Figure O.16: Base maps of univariate SiO_2 errors generated by OCK.

Appendix P

Compositional error base maps

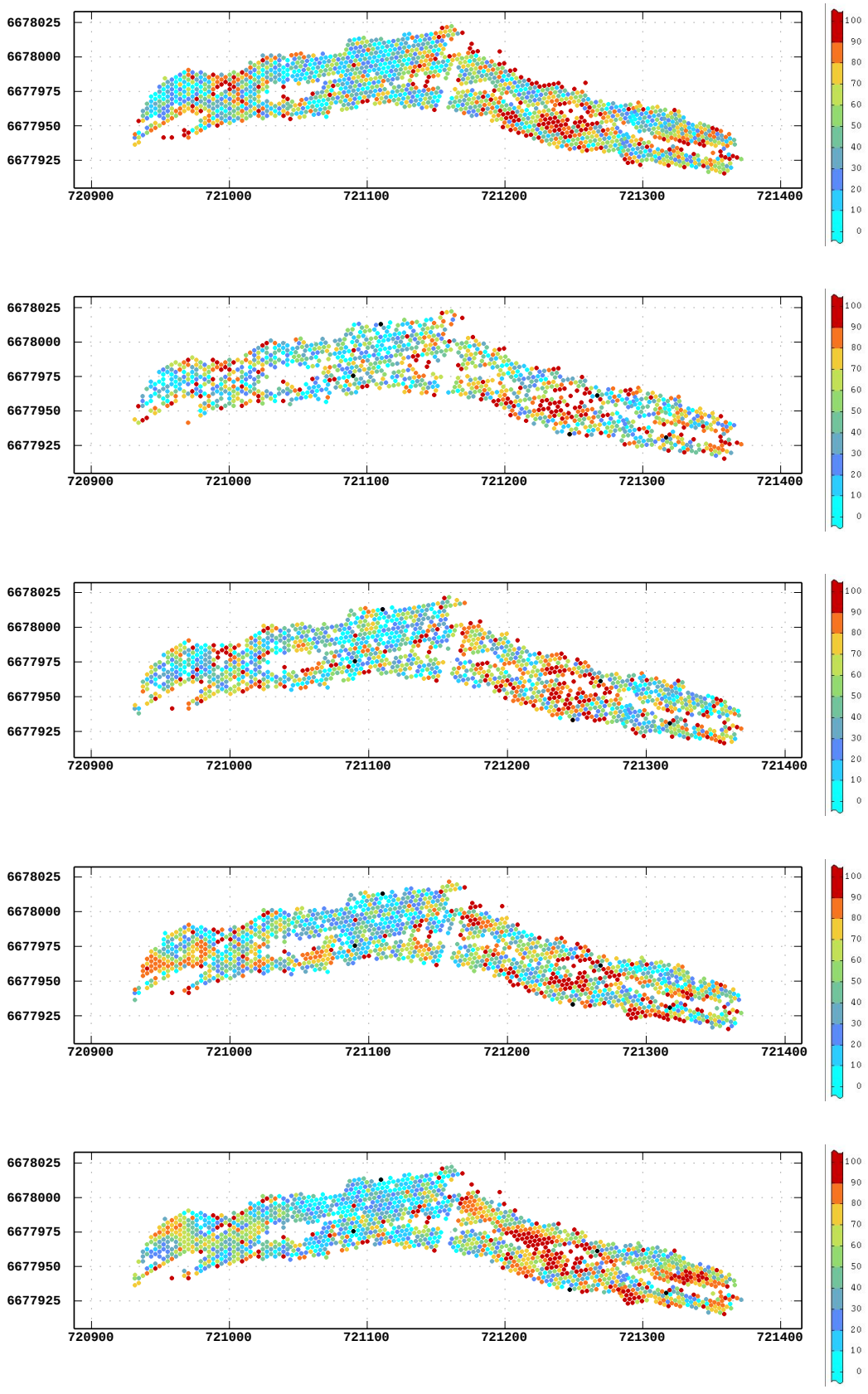


Figure P.1: Base maps of Euclidean errors generated by OCK.

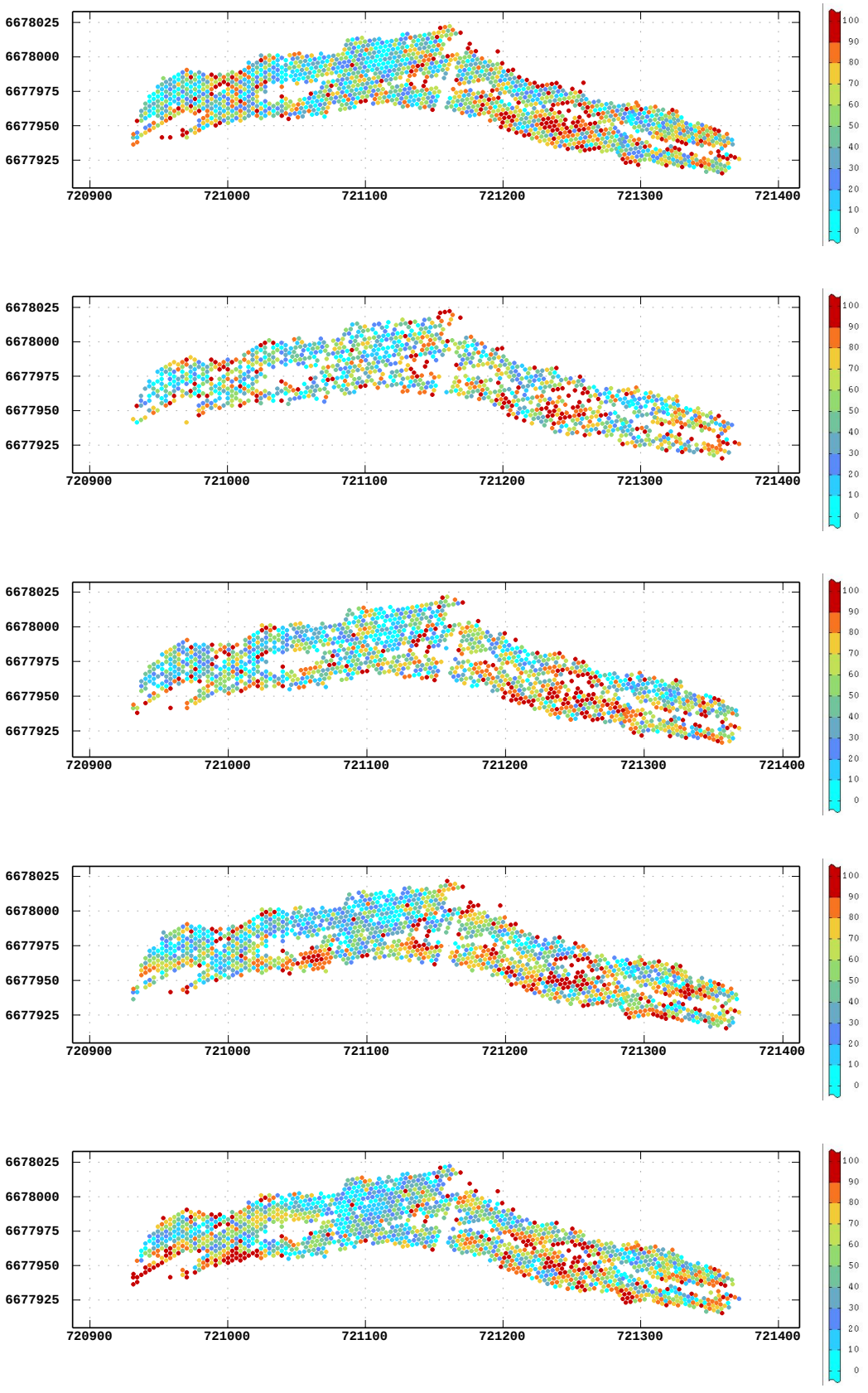


Figure P.2: Base maps of Euclidean errors generated by GH.

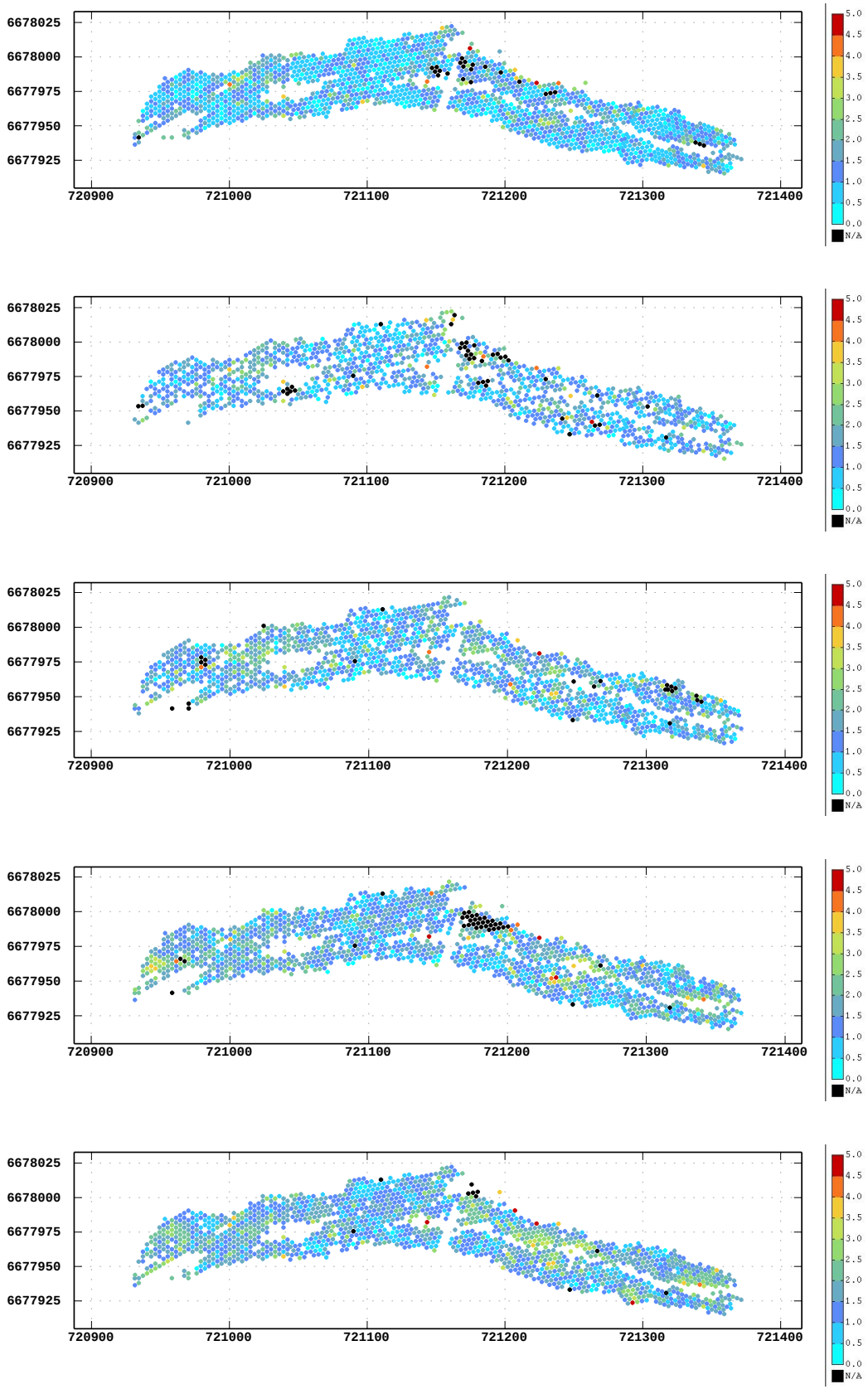


Figure P.3: Base maps of Aitchison errors generated by OCK.

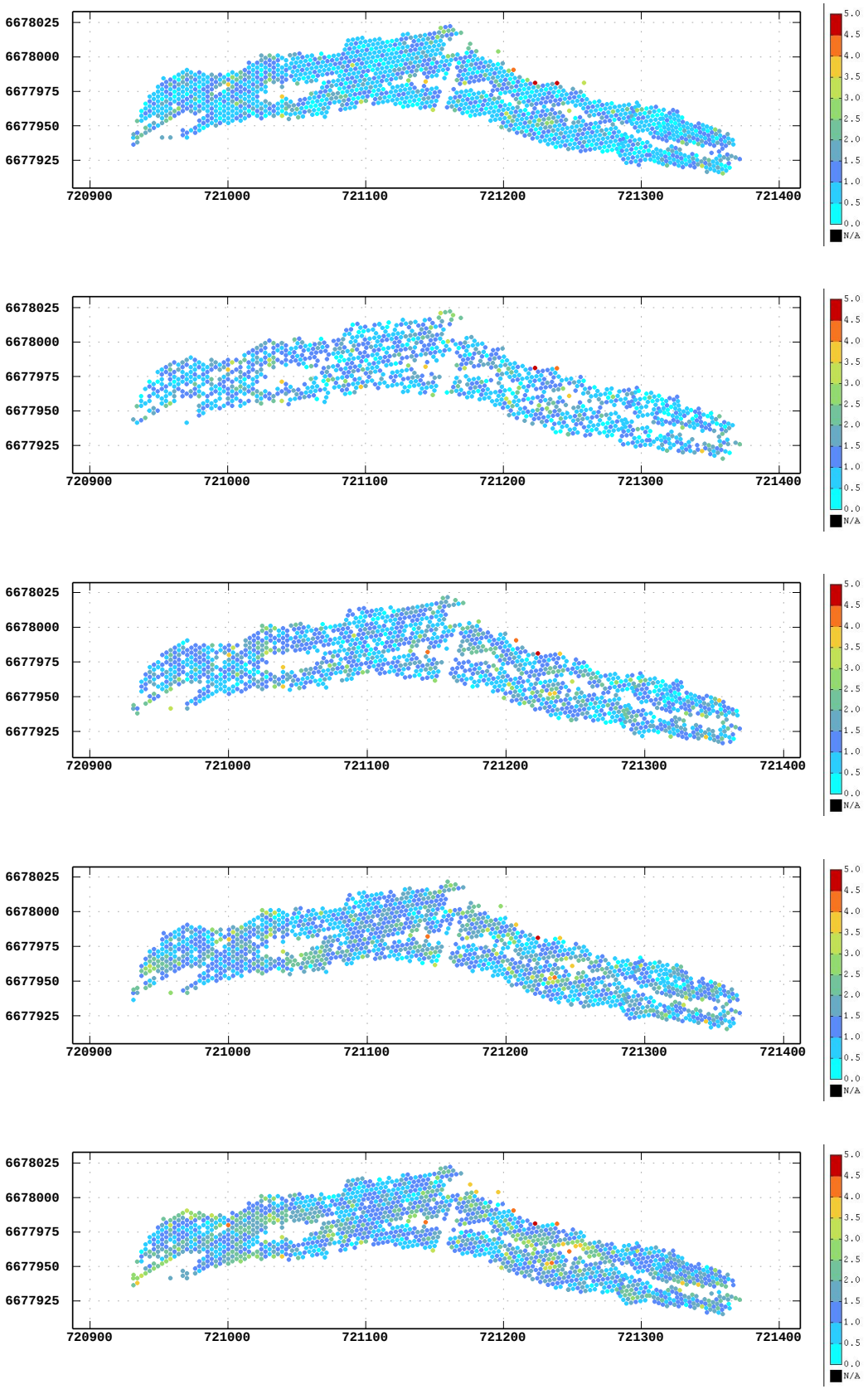


Figure P.4: Base maps of Aitchison errors generated by GH.

Appendix Q

Accuracy and Precision plots

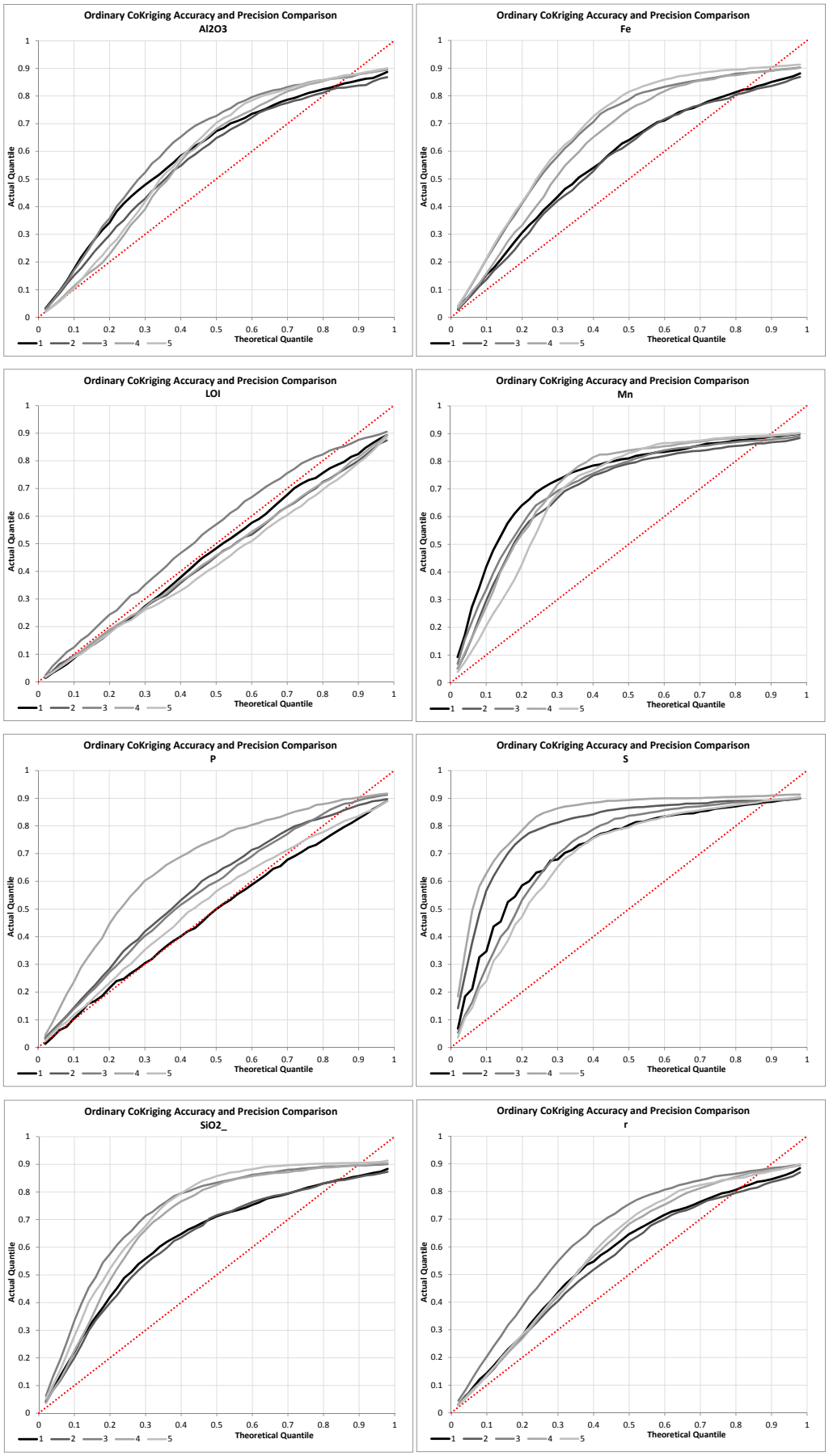


Figure Q.1: Accuracy and precision of OCK estimated distributions as a function of decreasing data density.

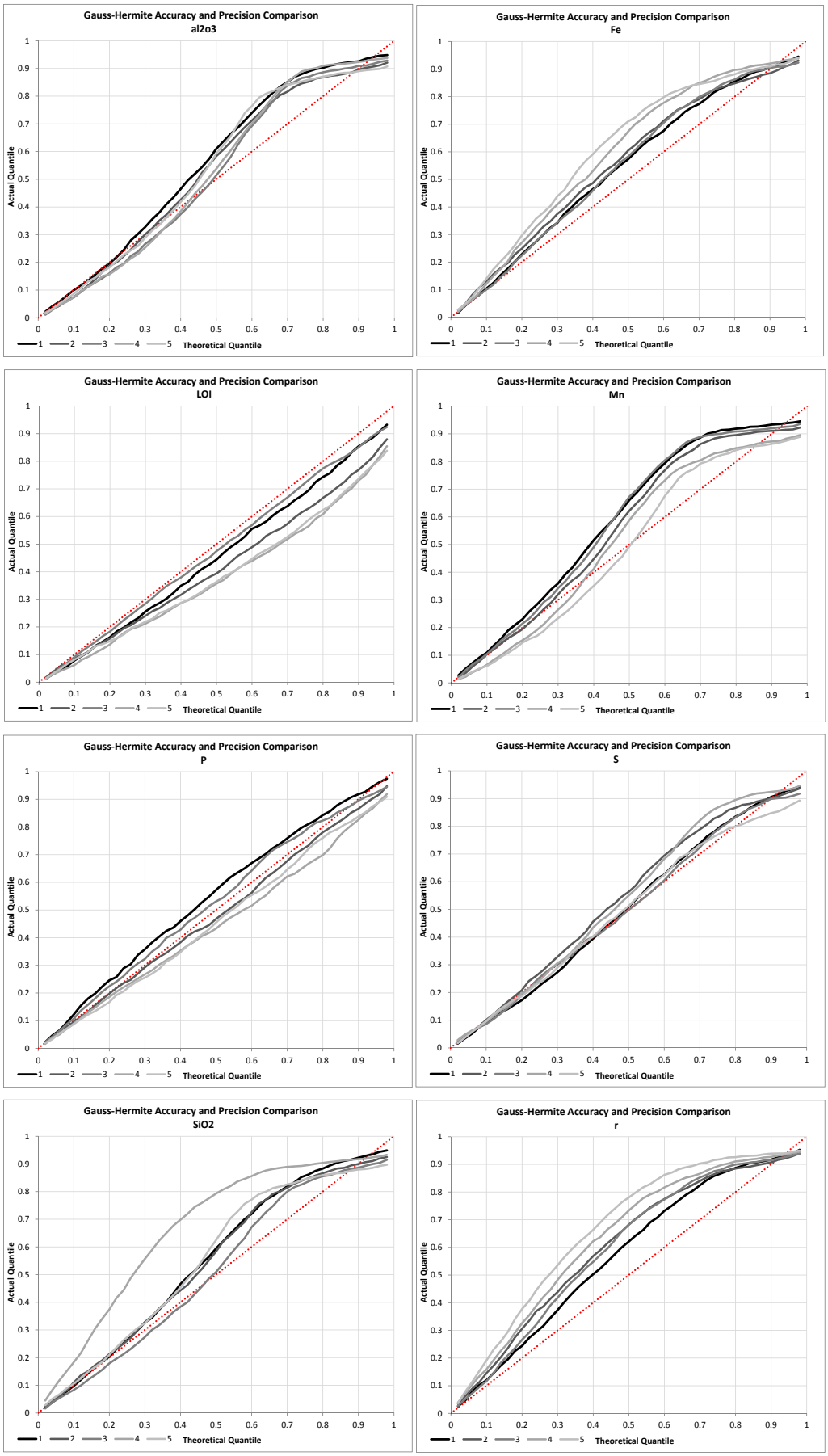


Figure Q.2: Accuracy and precision of GH estimated distributions as a function of decreasing data density.

References

[1] J. G. Manchuk, "Guide to geostatistics with compositional data," 3-133 Markin/CNRL Natural Resources Engineering Facility, Edmonton, AB, Canada T6G 2W2, 2008.

[2] K. Pearson, "Mathematical contributions to the theory of evolution.–on a form of spurious correlation which may arise when indices are used in the measurement of organs," *Proceedings of the Royal Society of London*, vol. 60, pp. 489–498, 1897.

[3] F. Chayes, "On correlation between variables of constant sum.," *Journal of Geophysical Research*, vol. 65, no. 12, pp. 4185–4193, 1960.

[4] J. Aitchison, "The statistical analysis of compositional data," *Journal of the Royal Statistical Society*, no. 2, pp. 139–172, 1982.

[5] J. Aitchison, *The statistical analysis of compositional data: (Reprint)*. Caldwell, NJ: Blackburn Press, 2003.

[6] V. Pawlowsky, "Cokriging of regionalised compositions," *Mathematical Geology*, vol. 21, no. 5, pp. 513–521, 1989.

[7] V. Pawlowsky-Glahn and R. Olea, *Geostatistical Analysis of Compositional Data*. International Association for Mathematical Geosciences: Studies in Mathematical Geosciences, Oxford University Press, USA, 2004.

[8] V. Pawlowsky-Glahn, J. J. Egozcue, and R. Tolosana-Delgado, "Lecture notes on compositional," tech. rep., University of Girona, 2007.

[9] D. Walvoort and J. de Gruijter, "Compositional kriging: A spatial interpolation method for compositional data," *Mathematical Geology*, pp. 951–966, 2001.

[10] E. H. Isaacs and R. M. Srivastava, *An introduction to Applied Geostatistics*. Oxford University Press, 1989.

[11] B. Koushavand and C. Deutsch, "Constraining the sum of multivariate estimates," Tech. rep., University of Alberta, 2008. Twelfth Annual Report of the Centre for Computational Geostatistics.

[12] E. Jarauta-Bragulat, C. Hervada-Sala, and A. Diblasi, "An experimental comparison of cokriging of regionalised compositional data using four different methods case study: Bauxites in hungary," *Journal of Hungarian Geomathematics*, vol. 1, pp. 7–13, 2003.

[13] N. Park and D. Jang, "Comparison of geostatistical kriging algorithms for intertidal surface sediment facies mapping with grain size data," *The Scientific World Journal*, vol. 2014, p. 11, 2014. Article ID 145824.

[14] A. Saavedra, C. Ord, J. Taboada, and J. Armesto, "Compositional kriging applied to the reserve estimation of a granite deposit," *Dyna*, vol. 77, no. 161, pp. 53–60, 2010.

[15] A. Menafoglio, A. Guadagnini, and P. Secchi, "Kriging prediction for functional compositional data and application to particle-size curves," in *47th Scientific Meeting of the Italian Statistical Society*, Italian Statisticl Society, 2014.

[16] X.Sun, Y. Wu, H. Wang, Y. Zhao, and G.Zhang, "Mapping soil particle size fractions using compositional kriging, cokriging and additive log-ratio cokriging in two case studies," *Mathematical Geosciences*, vol. 46, no. 4, pp. 429–443, 2014.

[17] M. Boezio, J. Costa, and J. Koppe, "Ordinary cokriging of additive log-ratios for estimating grades in iron ore deposits," in *CodaWork '11* (J. Egozcue, R. Tolosana-Delgado, and M. Ortego, eds.), 2011.

[18] M. Boezio, L., and J. Felipe, "Maf decomposition of compositional data to estimate," in *Geostatistics Oslo 2012*, 2012.

[19] C. Ward and U. Mueller, *Multivariate Estimation Using logratios: A Worked Alternative*. Quantitative Geology and Geostatistics, Springer, 2012.

[20] C. Ward and U. Mueller, *Compositions, logratios and Bias From Grade Control to Resource*. AusIMM, 2013.

[21] M. Job, "Application of logratios for geostatistical modelling of compositional data," diploma thesis, Faculty of Graduate Studies and Research, University of Alberta, 2012.

[22] J.-P. Chiles and P. Delfiner, *Geostatistics Modelling Spatial Uncertainty*. John Wiley and Sons, 1999.

[23] R. Tolosana-Delgado, U. Mueller, K. van den Boogaart, and C. Ward, "Block cokriging of a whole composition," in *International Symposium on the Applications of Computers and Operations Research in the Mineral Industry* (J. Costa, J. Koppe, and R. de Lemos Peroni, eds.), 2013.

[24] R. Tolosana-Delgado, *Geostatistics for constrained variables: positive data, compositions and probabilities. Application to environmental hazard monitoring*. PhD thesis, Universitat de Girona, 2006.

[25] R. Tolosana-Delgado, N. Otero, and V. Pawlowsky-Glahn, "Some basic concepts of compositional geometry," *Mathematical Geology*, vol. 37, no. 5, pp. 673–680, 2005.

[26] J. J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barcelo-Vidal, "Isometric logratio transformations for compositional data analysis," *Mathematical Geology*, vol. 35, 2003.

[27] J. J. Egozcue and V. Pawlowsky-Glahn, *Compositional Data Analysis*. Wiley, 2011.

[28] J. J. Egozcue and V. Pawlowsky-Glahn, "Groups of parts and their balances in compositionaldata analysis," *Mathematical Geology*, vol. 37, no. 7, pp. 795–838, 2005.

[29] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 1997.

[30] M. Armstrong, *Basic Linear Geostatistics*. Springer, 1998.

[31] D. E. Myers, "Matrix formulation of co-kriging," *Mathematical Geology*, vol. 14, no. 3, pp. 249–257, 1982.

[32] G. S. Adisoma, *The application of the jackknife in geostatistical resource estimation: Robust estimator and its measure of uncertainty*. PhD thesis, Department of Mining and Geological engineering, University of Arizona, 2014.

[33] M. J. Pyrcz and C. V. Deutsch, *Geostatistical Reservoir modelling*. Oxford University Press, first ed., 2002.

[34] C. Deutsch, "Direct assessment of local accuracy and precision," in *Wollongong '96 Proceedings of the* 5^{th} *InternationalGeostatisticsCongress*, $vol.\ 1, pp.\ 115 - -125$, *KluwerAcademic*, 1997.

[35] J. A. Martin-Fernandez, R. A. Olea-Meneses, and V. Pawlowsky-Glahn, "Criteria to compare estimation methods of regionalized compositions," *Mathematical Geology*, vol. 33, no. 8, pp. 889–909, 2001.

[36] K. Cassidy, D. Champio, B. Krapz, M. Brown, R. Blewett, P. Groenewald, and I. Tyler, "A revised geological framework for the yilgarn craton, western australia," tech. rep., Geological Survey of Western Australia, 2006.

[37] A. Riganti and S. Chen, *Geology of the Jackson 1:100,000 Sheet*. Geological Survey of Western Australia, 2002.

[38] T. Angerer and S. G. Hagemann, "The bif-hosted high-grade iron ore deposits in the archean koolyanobbing greenstone belt, western australia: Structural control on synorogenic- and weathering-related magnetite-, hematite-, and goethite-rich iron ore," *Economic Geology*, vol. 105, pp. 917–945, 2010.

[39] T. Angerer, S. G. Hagemann, and L. Danyushevsky, "High-grade iron ore at windarling, yilgarn craton: a product of syn-orogenic deformation, hypogene hydrothermal alteration and supergene modification in an archean bif-basalt lithostratigraphy," *Mineralium Deposita*, vol. 48, pp. 697–728, 2013.

[40] J. Vann, S. Jackson, and O. Bertoli, "Quantitative kriging neighbourhood analysis for the mining geologist a description of the method with worked case examples," in *5th International Mining Geology Conference*, p. 10, 2003.

[41] F. F. Pitard, "Blasthole sampling for grade control the many problems and solutions," in *AusIMM Sampling 2008 Conference*, 2008.

[42] B. Douglas, I andPerry, W. Hardtke, and L. Allen, "Angle reverse circulation versus blastholes in ore control sampling quality versus data density a case study," in *Orebody modelling and Strategic Mine planning*, pp. 193–198, 2009.

[43] J. O. R. Committee, "The JORC code," tech. rep., Australian Securities Exchange, 2012.

[44] J. Martin-Fernandez, R. Olea-Meneses, and V. Pawlowsky-Glahn, "Criteria to compare estimation methods of regionalized compositions," *Mathematical Geology*, pp. 889–909, 2001.