2011

Plurigaussian Simulation of rocktypes using data from a gold mine in Western Australia

Robin Dunn
Edith Cowan University

EDITH COWAN UNIVERSITY

SCHOOL OF ENGINEERING

Plurigaussian Simulation of rocktypes
using data from a gold mine in Western Australia

Robin Dunn

(5820063)

May 2011

Supervisor:

Associate Professor Ute Mueller

Edith Cowan University

Copyright Warning

USE OF THESIS

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

ABSTRACT

Stochastic simulation of rocktypes, or the geometry of the geology, is a major area of continuing research as earth scientists seek a better understanding of an orebody as a precursor to the assignment of continuous rock properties, allowing more economically appropriate decisions regarding mine planning. This thesis analyses the suitability of particular geostatistical rock type modelling algorithms when applied to the five rocktypes evident in drill hole data from the Big Bell gold mine near Cue, Western Australia. The background of the geostatistical theory is considered, in particular the concept of the random function model and the link between the categorical statistics determined from the drill hole data and the three models used for estimation and simulation. The commonly applied indicator kriging (IK) and sequential indicator simulation (SIS) algorithms are compared in a non-sedimentary gold deposit environment to the more computationally demanding and more complex plurigaussian simulation (PGS). Comparisons between the three models are made by examining global and regional rocktype (lithotype) proportions of the outputs of the models, both visually and empirically. The models are validated by considering the contacts which occur in reality between different lithotypes and the proportion of contacts which do not conform to this reality in each of the models. This 'inadmissible contact' ratio measures the short range validity of the estimation and simulation techniques. Finally, cores taken from the output of the models are compared to the drill hole data in terms of transition proportions between the twenty five possible transitions for the five lithotypes. Inadmissible contacts were at a minimum with PGS, and the visual and empirical natures of the PGS output were closely linked to the reality of the drill hole data. Whilst each model produced similar 3D images, PGS was a realistic balance between the clustering effect produced by IK and the fine mosaic effect from SIS. The PGS output numerically outperformed the other two models in terms of admissible contacts and connectivity, most closely matching the drill hole data. All results indicate that, whilst demanding to implement, PGS produces the most adequate model of the study region.

COPYRIGHT AND ACCESS DECLARATION

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

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

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

    (iii) Contain any defamatory material.

    (iv) 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.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Background and significance

Geostatistics is a well-established suite of mathematical and probabilistic techniques developed in the 1960s in response to the need for mathematical techniques that would enable evaluation of the recoverable reserves in ore deposits. It can be defined as a set of statistical tools that allows description, interpretation and modelling of the spatial continuity that is a fundamental feature of many naturally occurring phenomena, taking into account both the structure and randomness of such phenomena. The techniques developed have found application in areas such as petroleum, soil sciences and hydrogeology, where there is a need for interpretation of spatially correlated data and subsequent reservoir modelling. Modelling of spatial categorical data (in this thesis: rock types) is an important, initial step in such areas as mining viability or regional hydrology. In the mining context, once a three dimensional simulation model of the rock types has been produced the different continuous variables within those rock types can be calculated.

Plurigaussian simulation (PGS) is computationally demanding and there are a number of sequenced steps needed before the simulation can be produced, however it is the incorporation of spatial variability measures in PGS that will, it is surmised, produce a more accurate simulation of the study region. PGS models provide more straightforward and realistic transitions between categories than sequential indicator simulation (SIS) (Galli et al, 1994). There are inherent smoothing issues in indicator kriging (IK) that post processing algorithms have attempted to address. It is economically important to produce realistic models that honour both the original categorical transitions and proportions. PGS employs a constraint on allowable contacts between rock types that is initially defined by the drill hole data set. Neither IK nor SIS has such a constraint and an important comparison to PGS is the proportion of IK and SIS models that have so called illegal contacts between rock types in their output, indicating that the geometry found in the initial exploration has not been honored.

1.2 Early techniques in geostatistics

The linear least squares regression algorithms introduced by Danie Krige (1951) were an initial effort at estimation of attributes of interest at unknown locations. Krige's empirical work to evaluate mineral resources was formalised in the 1960s by French engineer Georges Matheron, who subsequently contributed significantly to the discipline by the introduction of the concept of a regionalised variable (ReV). Initial techniques in geostatistics were concerned with describing the spatial variability of earth science data and estimating attribute values at unsampled locations (kriging). A well-researched account of the history and origins of kriging can be found in Cressie (1990). The process of kriging gives an accurate least squares regression value and is convenient in obtaining the conditional cumulative distribution function (ccdf). It therefore is commonly used to estimate the probability of distribution, or single location uncertainty. Three common versions of kriging are Simple Kriging (SK), where the mean of the attribute in question is considered to be constant over a study region, Ordinary Kriging (OK), where local variation in the mean of the attribute in question is accounted for, and Indicator Kriging (IK) where the attribute in question is apportioned by a series of thresholds into classes. Kriging is used both for continuous data (attributes such as mineral grades) and, in this thesis, categorical data such as rock types. Drawbacks of kriging are that the single realisation of an attribute does not reflect the full spatial and statistical variability of the sample data, there being a smoothing effect inherent in the process. This smoothing effect is minimal when the locations to be estimated are nearby the observed data but increases as the distance from the location to be estimated to the observed data increases.

The development of geostatistical simulation techniques was a necessary step in overcoming the perceived shortcomings of the estimation algorithms, and began with the work of Matheron (cited in Chiles& Delfiner, 1999, p.472) and Journel (1974). Compared to kriging, conditional simulation is an advance in characterising the spatial variation of earth variables such as rock types. The simulation approach takes into account not only the spatial variation of observed data at sample locations but also the variation in estimates at unsampled locations (Deutsch, 2006). Vann, Bertoli and

Jackson (2002) provide a summary of the development of these techniques. Numerous simulation techniques have been developed over the years and as they are computationally demanding their growth can be aligned with the increased availability and power of computers, beginning in the 1980s. Many are based on the assumption that the random function(s) is multiGaussian and include methods such as covariance matrix (LU) decomposition, turning bands, sequential indicator simulation (SIS), sequential Gaussian simulation (sGs), truncated Gaussian and the more complex PluriGaussian simulation.

Categorical data, such as rock types (lithotypes), can be modelled using a variety of algorithms. Indicator kriging (IK) was introduced by Journel (1983) and for a considerable time was one of the most widely applied grade estimation techniques in the mining industry. The original appeal of IK was that it was non-parametric- it made no assumptions about the distribution underlying the sample data. The basis of the indicator approach is the binomial coding of data into either 1 or 0 depending upon its relationship to a cut-off value, resulting in a non-linear transformation of the data values. A benefit for continuous data is that values much greater than the cut-off value will receive the same indicator value as data close to the cut-off, thus limiting the effect of very high values in the kriging process. The outcome of IK is a conditional cumulative distribution function (ccdf) which is a distribution of local uncertainty or possible values conditional to data in the neighbourhood of the block or cell to be estimated. This cumulative distribution at a block or cell should be non-decreasing and valued between 0 and 1. These two requirements are sometimes not met, leading to order relations violations. Many methods have been proposed to overcome this issue, including direct correction of the indicator values (eg. Deutsch and Journel, 1998), using nested indicator variables (Dagbert and Dimitrakopoulos (1992) and probability kriging (Isaaks, 1984) which is indicator co-kriging between the indicator-transformed data and a uniform transform of the sample data. A criticism of IK is the need to calculate a variogram and develop a model for each cut-off value, which for multiple cut-offs was initially a major issue, however computer speed and memory capacity have removed this problem. The indicator transform lends itself to the estimation of categorical data such as rock types. In this case, instead of kriging indicators at a set of thresholds, categorical IK will produce the probability of a given rock type at a given location producing probability maps of given rock types. This may be combined with

indicator estimation or simulation of grade data, as described in Dowd (1996). Lloyd and Atkinson (2001) suggest that adaptation of the variogram for IK to local variation would increase the accuracy of the estimations by resolving the problem of changing means and variograms across the study region, and this approach has been formalized by the conditional simulation paradigm.

Sequential indicator simulation (SIS) is a non-parametric algorithm that, as for IK, requires the binary transformation of continuous (eg. porosity or permeability) or categorical (eg. rock type) data into a series of indicator variables (Goovaerts, 1997, Deutsch and Journel, 1998). It applies IK in a sequential fashion where a precise category is drawn by Monte Carlo simulation at each location following a random path through the three dimensional grid. Individual grid nodes are simulated, one after the other, using continually updated (and increasing-sized) conditioning data sets, the conditioning data set including the original data set and all previously simulated values within a specified neighbourhood. This ensures that closely spaced values have the correct short scale correlation. This algorithm has been employed in such varied studies as categorical soil variable simulation (Li, Zhang, 2007), simulation of facies for groundwater flow (Moysey et al, 2003), reservoir fluid flows (Seifert and Jensen (1999) and uncertainty assessment of the spatial distribution of soil organic carbon (Delbari et al (2010). It has been described in detail by Deutsch and Journel (1992) and Srivastava (1994). Deutsch (1996) uses SIS in a categorical context and compares a number of SIS algorithms, but notes two common criticisms of the technique: the models produced can appear very patch and unstructured, and SIS often leads to uncontrolled and unrealistic transitions between categories. Li and Zhang (2007) also found that SIS models (using simple or ordinary kriging) do not obey interclass relationships in simulation maps. A number of SIS variations have evolved over the years (Gomez-Hernandez and Srivastava, 1990; Goovaerts, 1994; Goovaerts, 1997), mostly relating to the use of soft secondary data arising from geological interpretation or geophysical measurements.

The above models of IK and SIS are both non-parametric, in that they have no predetermined underlying model. Many simulation algorithms make the purely mathematical assumption that such an underlying model theoretically exists, and a popular set of these models use the Gaussian function as its basis. Two algorithms which use this assumption are truncated Gaussian simulation (TGS) and plurigaussian

simulation (PGS) Both TGS and PGS were initially developed in the 1980s to model oil reservoirs in sedimentary geological environments, where different rock types typically occur in a defined sequential order. A two-step approach was developed to model quantities such as porosity and permeability in reservoirs: firstly a model of the geometry of rock types in the reservoir was done and then this model was used to generate values of porosity.

The truncated Gaussian method is based on the notion of association of a region under the normal curve to a category proportion. The truncation of a standard Gaussian random function by different thresholds calculated from experimental set proportions was first applied to model the lithotype in deltaic reservoirs by Matheron et al. (1987). The covariance model of the underlying Gaussian function is obtained by an iterative process of inversion of the covariance model of the experimental indicator variables of the sets. Once this Gaussian semivariogram is known, kriging weights are available and further work is possible. If the categorical data are sequential in their contacts then truncation of one Gaussian RF is required (TGS), however PGS is an extension to the algorithm that allows for more complex patterns and different types of contacts between rock types to be modeled.

Few texts, with the exception of Armstrong et al. (2003), are solely concerned with a succinct explanation of the basis of the recent development of plurigaussian simulation techniques, although numerous theses, journal articles, course notes and conference proceedings are available. The truncated Gaussian and plurigaussian simulation algorithms have been investigated by Dowd (2003), Le Loc'h and Galli (1996) and Lantuéjoul (1994, 1997). Remacre and Zapparolli (2003) employ plurigaussian techniques to model the reservoir characteristics of an oilfield enabling the practitioners to accurately predict oil production forecasts. Deutsch (1996) suggests that PGS is a more straightforward approach than SIS when handling multiple category interactions

1.3   Objectives

The objective of this research is to explore the theory underpinning the computationally expensive plurigaussian simulation (PGS) technique and compare the output from PGS with the previously developed and more simplistic models of indicator kriging (IK) and sequential indicator simulation (SIS), with a view to assessing overall performance of the three models. Comparisons will be undertaken in three areas:

1. Visual inspection and empirical analysis of output models. Particular emphasis will be made of global and regional proportions of the five rocktypes present in the data set and by consideration of the regional spatial variability displayed in each model with comparison to the data set.
2. Analysis of the four non-bounding rocktypes within the study region.
3. Calculation of the degree to which the three models honour the specific rocktypes that appear in the data set in terms of contact admissibility and continuity.

1.4   Outline of Thesis

In this thesis Chapter 2 will provide a theoretical background for the basic geostatistical measures associated with the random function (RF), the measures associated with categorical data, the kriging paradigm, its variant indicator kriging (IK), the sequential indicator simulation (SIS) algorithm, the Gaussian function as a model for simulations, the turning band algorithm, truncated Gaussian and PGS model overview and lastly the link between the indicator semivariogrmas and the underlying Gaussian semivariograms. An extensive discussion of aspects of the workflow of the plurigaussian algorithm follows in chapter 3, including the Gibbs sampler conditioning algorithm and the turning bands simulation method. Chapter 4 explores the Big Bell gold mine data set used in the modelling processes. Chapter 5 details the steps of the plurigaussian algorithm as run on the platform of the geostatistical programme Isatis, developed by Geovariances. Chapter 6 provides pictorial and statistical analysis of the output generated from the three algorithms considered in this thesis, with particular attention on the adequacy of the models to reproduce the initial global proportions of the data set and the identification of important differences in the output of the three algorithms.

1.5 Common Notation

The notation used throughout this document is a combination of that used by Goovaerts (1997) and Armstrong et al. (2003).

A	study area
$C(0)$:	covariance value at separation distance $\lvert \mathbf{h} \rvert = 0$
$C_{ij}(\mathbf{h})$:	stationary cross covariance between the two RFs $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$
$C_Z(\mathbf{h})$:	stationary covariance of the random function
$E\{\cdot\}$:	expected value
$E\{(Z(\mathbf{u});: z \mid (n)\}$:	conditional expectation of the RV $Z(\mathbf{u})$ given the realisations of n other neighboring RVs (called data)
$F(\mathbf{u}: z)$	non-stationary cumulative distribution function of the RV $Z(\mathbf{u})$
$F(\mathbf{u}: z \mid (n))$:	non-stationary conditional cumulative distribution function of the continuous RV $Z(\mathbf{u})$ given neighboring information, such as realisations of n other RVs (called data).
$F(\mathbf{u}, \mathbf{u}': z, z')$:	non-stationary "two point" cumulative distribution function of the RV $Z(\mathbf{u})$
$G(y)$:	standard normal cumulative distribution function
$G^{-1}(y)$:	standard normal quantile function such that $G(G^{-1}(p)) = p \in [0,1]$
$\gamma(\mathbf{u}, \mathbf{u}')$:	non-stationary semivariogram between the two RVs $Z(\mathbf{u})$ and $Z(\mathbf{u}')$
$\mathbf{h}$:	separation or lag vector
$I(\mathbf{u}; s_k)$:	binary indicator RF at location $\mathbf{u}$ and for threshold s_k
$i(\mathbf{u}; s_k)$:	binary indicator value at location $\mathbf{u}$ and for threshold s_k
$\lambda_\alpha^{SK}(\mathbf{u})$:	simple kriging weight associated to z-datum at location $\mathbf{u}_\alpha$ for estimation of the attribute z at location $\mathbf{u}$. The same type of notation applies to other algorithms, such as OK.
m:	stationary mean of the RF $Z(\mathbf{u})$
$N(\mathbf{h})$:	number of pairs of data values available at lag vector $\mathbf{h}$
n:	number of data values $s(\mathbf{u}_\alpha)$ or $z(\mathbf{u}_\alpha)$ available over the study area
$n(\mathbf{u})$:	number of data values $z(\mathbf{u}_\alpha)$ used for estimation of the attribute z at location $\mathbf{u}$
$p(\mathbf{u}: s_k)$:	probability for the category s_k to prevail at location $\mathbf{u}$
$p(\mathbf{u}: s_k \mid (n))$:	conditional probability for the category s_k to prevail at location $\mathbf{u}$ given the neighboring information (n)
p_k:	global proportion of category s_k within the area A

$R(\mathbf{u})$:   residual component model in the decomposition

$Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$ where $m(\mathbf{u})$ represents the trend component model

$\rho(\mathbf{h})$:   stationary correlogram of the RF $Z(\mathbf{u})$ for lag vector $\mathbf{h}$

$S(\mathbf{u})$:   generic categorical RV at location $\mathbf{u}$ or a generic categorical RF at location $\mathbf{u}$

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

σ^2:   variance of the RV Z

$\sigma^2_{SK}(\mathbf{u})$:   simple kriging variance associated with the simple kriging estimate $Z^*_{SK}(\mathbf{u})$ at location $\mathbf{u}$. The same type of notation applies to other algorithms, such as OK

$\mathbf{u}$:   coordinate vector

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

$Var\{\cdot\}$:   variance

$Z(\mathbf{u})$:   generic continuous RV at location $\mathbf{u}$, or a generic continuous RF at location $\mathbf{u}$

$Z^*_{SK}(\mathbf{u})$:   simple kriging estimator of $Z(\mathbf{u})$

z: continuous attribute

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

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

$z^*(\mathbf{u})$:   an estimate of value $z(\mathbf{u})$

Chapter 2: Theoretical Framework

2.1 Random Variable and Random Function

Before consideration of the mathematics associated with the plurigaussian technique, a number of mathematical and geostatistical concepts must be discussed, such as random variables, the concept of a random function, the Gaussian function and the kriging paradigm.

In this section the concept of a continuous random function (RF) is discussed. Informally, a RF is a collection of random variables defined at each of the locations that comprise the study area of interest.

A random variable (RV) is a variable, denoted by $Z(\mathbf{u})$, that can take a series of outcome values according to some probability distribution. For the purposes of simulation of rock types, the random variable is a discrete (categorical) variable, however in the processes of simulation to be considered in this thesis, these categorical variables are converted into continuous (Gaussian) variables for the modelling process and then back-transformed to categorical. The following refers to continuous RVs, and when categorical attributes will be considered at a later time, specific notation will be introduced.

A random variable is characterised by its cumulative distribution function (cdf) which indicates the probability that the variable Z at location $\mathbf{u}$ is no greater than a given threshold z:

$$F(\mathbf{u}; z) = p\{Z(\mathbf{u}) \leq z\} \tag{1}$$

A random function (RF) is defined as a set $,\{Z(\mathbf{u}) \;\; \mathbf{u} \in A\}$ of (usually dependent) random variables $Z(\mathbf{u})$, one for each location $\mathbf{u}$ in the study region A. The basic paradigm of the probabilistic approach is to model a value $z(\mathbf{u})$ at an unsampled location $\mathbf{u}$ as a realisation drawn from a random variable $Z(\mathbf{u})$. The problem of assessing the

uncertainty about an attribute value at **u** thus reduces to that of modelling the probability distribution of the random variable Z at that location. The uncertainty at **u** can be reduced by accounting for (conditioning on) neighbouring sample data values $z(\mathbf{u}_\alpha), \alpha = 1,...,n$.

In practice the distinction between $Z(\mathbf{u})$ and $z(\mathbf{u})$ reduces to one of context, with $z(\mathbf{u}_\alpha)$ indicating a (known) sample value and $z^*(\mathbf{u})$ an estimate of Z at location **u**. The univariate cumulative distribution function (cdf) (1) can be extended to the multivariate or n point (cdf) where:

$$F(\mathbf{u}_1,........,\mathbf{u}_n; z_1,.........z_n) = p\{Z(\mathbf{u}_1) \leq z_1,.........Z(\mathbf{u}_n) \leq z_n\} \tag{2}$$

The spatial law of the RF $Z(\mathbf{u})$ is defined by the joint uncertainty shown in (2).

Generally only the univariate and bivariate cdfs are considered when defining the moments of interest. These are:

- the one point cdf

    $$F(\mathbf{u}; z) = P\{Z(\mathbf{u}) \leq z\} \tag{3}$$

- the expected value

    $$m(\mathbf{u}) = E\{Z(\mathbf{u})\} \tag{4}$$

- the two point cdf

    $$F(\mathbf{u}, \mathbf{u}'; z, z') = P\{Z(\mathbf{u}) \leq z, Z(\mathbf{u}') \leq z'\} \tag{5}$$

- the variance

    $$\text{Var}\{Z(\mathbf{u})\} = E[Z(\mathbf{u}) - E\{Z(\mathbf{u})\}]^2 \tag{6}$$

- the two point covariance

- $$C\{\mathbf{u}, \mathbf{u}'\} = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u}')\} - E\{Z(\mathbf{u})\} \cdot E\{Z(\mathbf{u}')\} \tag{7}$$

- the two point correlogram

    $$\rho(\mathbf{u}, \mathbf{u}') = \frac{C(\mathbf{u}, \mathbf{u}')}{\sqrt{C(\mathbf{u}, \mathbf{u}) \cdot C(\mathbf{u}', \mathbf{u}')}} \tag{8}$$

- the semivariogram

    $$\gamma(\mathbf{u}, \mathbf{u}') = \frac{1}{2}\text{Var}\{Z(\mathbf{u}) - Z(\mathbf{u}')\} \tag{9}$$

The term two point refers to the value of the same attribute at two different locations as opposed to the term bivariate, which may refer to two different attributes measured at the same location.

An important statistic is the conditional cdf (ccdf) $F(\mathbf{u}; z \mid Z(\mathbf{u}') \leq z')$ of the cdf of $Z(\mathbf{u})$ where the cdf at location $\mathbf{u}$ is calculated with reference to (conditioned on) the statistics associated with the neighbouring data, specifically that $Z(\mathbf{u}') \leq z'$. Using (3) and (5);

$$F(\mathbf{u}; z \mid Z(\mathbf{u}') \leq z') = P\{Z(\mathbf{u}) \leq z \mid Z(\mathbf{u}') \leq z'\}$$

$$= \frac{P\{Z(\mathbf{u}) \leq z, Z(\mathbf{u}') \leq z'\}}{P\{Z(\mathbf{u}') \leq z'\}}$$

$$= \frac{F(\mathbf{u}, \mathbf{u}'; z, z')}{F(\mathbf{u}'; z')} \tag{10}$$

A random function is strictly stationary if its distributions are invariant under translation of the points by an arbitrary vector $\mathbf{h}$; (for all $\mathbf{u}_i, i = 1,...,n$ and $z_i, i = 1,...,n$)

$$P\{Z(\mathbf{u}_1) < z_1, ... Z(\mathbf{u}_n) < z_n\} = P\{Z(\mathbf{u}_1 + \mathbf{h}) < z_1, ... Z(\mathbf{u}_n + \mathbf{h}) < z_n\} \tag{11}$$

In physical terms this means that the phenomenon is homogenous in space. When a random function is stationary its moments are invariant under translation. In practice the analysis is limited to cdfs involving no more than two locations at a time and their corresponding moments. The two moments (4) and (8) depend on the separation vector $\mathbf{h}$:

$$m = E\{Z(\mathbf{u})\} \tag{12}$$

and

$$C(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u} + \mathbf{h})\} - E\{Z(\mathbf{u})\} \cdot E\{Z(\mathbf{u} + \mathbf{h})\} \tag{13}$$

A random function with these two properties is said to be *second order* or weakly stationary.

Stationarity is not necessarily a property of the underlying characteristic of the phenomenon under study, rather it is a decision made by analysts to facilitate inference and modelling. It may well be that non-stationarity (where the moments of the distribution vary across the study region) is a factor to be taken into account in the modelling process.

2.2 The geostatistics of categorical data

The models considered so far relate largely to a continuous random variable. The categorical data to be considered in the simulation process need to be defined, and then the link between these variables and the continuous variables $Z(\mathbf{u})$ can be made. For the purposes of explanation at a generic categorical level, sample data will be categorised into k categories S_i, $i = 1, \ldots .k$, .

.
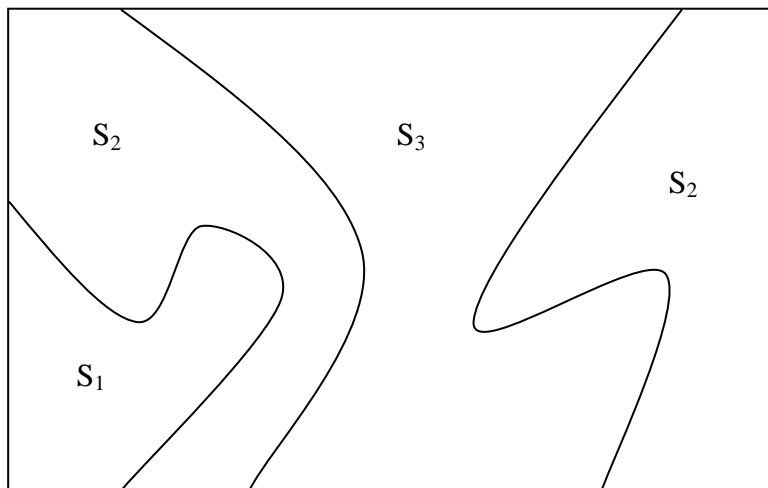
Figure 1 Categorical data contacts in a 2D region.

Figure 1 showing three categories in **A** with an important feature being which categories are neighbours for others. Category S_1 is in contact solely with S_2 whereas S_2 is in contact with both S_1 and S_3. This sequential nature of contacts between categories may be represented pictorially by partitioning a rectangle, as in Figure 2.
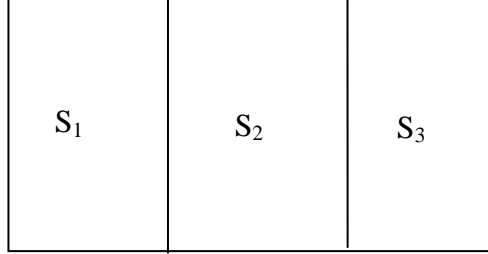
Figure 2 Representation of nature of category contacts from Figure 1.

Suppose that there are k mutually exclusive categories $s_1,...,s_k$. For each category we will define an indicator function by putting

$$i(\mathbf{u};s_k) = \begin{cases} 1 \text{ if } s(\mathbf{u}) = s_k \\ 0 \text{ otherwise} \end{cases}$$

(14)

The associated indicator RV will be denoted by

$$I(\mathbf{u};s_k) = \begin{cases} 1 \text{ if } s(\mathbf{u}) = s_k \\ 0 \text{ otherwise} \end{cases}$$

(15)

The proportion of the study region that a category occupies will be denoted by

$$p(\mathbf{u};s_k) = E\{I(\mathbf{u};s_k)\} \text{ with } 0 \le p(\mathbf{u};s_k) \le 1 \quad (16)$$

Since $\left[I(\mathbf{u};s_k) \right]^2 = I(\mathbf{u};s_k)$ the variance of the indicator function is given by

$$\text{Var}[I(\mathbf{u};s_k)] = E[\{I(\mathbf{u};s_k) - E[I(\mathbf{u};s_k)]\}^2] \quad (17)$$

Given any two locations $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$ the non-centred covariance $C_{s_k}(\mathbf{u},\mathbf{u}+\mathbf{h})$ is defined as

$$C_{s_k}(\mathbf{u},\mathbf{u}+\mathbf{h}) = E[I(\mathbf{u};s_k),I(\mathbf{u}+\mathbf{h};s_k)] = P[I(\mathbf{u};s_k)=1,I(\mathbf{u}+\mathbf{h};s_k)=1] \quad (18)$$

13

and measures the probability that the category at locations $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$ is s_k.

The centred covariance $\sigma_{S_k}(\mathbf{h})$ and centred cross-covariance $\sigma_{S_k S_j}(\mathbf{h})$ between S_k and S_j are defined as

$$\sigma_{S_k}(\mathbf{u},\mathbf{u}+\mathbf{h}) = E\{[I(\mathbf{u};s_k) - P(I(\mathbf{u};s_k))][I(\mathbf{u}+\mathbf{h};s_k) - P(I(\mathbf{u}+\mathbf{h};s_k))]\}$$
$$= C_{S_k}(\mathbf{h}) - P(I(\mathbf{u};s_k))P(I(\mathbf{u}+\mathbf{h};s_k))$$

(19)

and

$$\sigma_{S_k S_j}(\mathbf{u},\mathbf{u}+\mathbf{h}) = E\{[I(\mathbf{u};s_k) - P(I(\mathbf{u};s_k))][I(\mathbf{u}+\mathbf{h};s_j) - P(I(\mathbf{u}+\mathbf{h};s_j))]\} \quad (20)$$

The indicator semivariogram $\gamma_{S_k}(\mathbf{u},\mathbf{u}+\mathbf{h})$ of S_k (the indicator version of (9)) is defined as

$$\gamma_{S_k}(\mathbf{u},\mathbf{u}+\mathbf{h}) = \frac{1}{2}\{E[I(\mathbf{u};s_k) - I(\mathbf{u}+\mathbf{h};s_k)]^2 - (E[I(\mathbf{u};s_k) - I(\mathbf{u}+\mathbf{h};s_k)])^2\} \quad (21)$$

and measures how often one of the locations separated by a vector $\mathbf{h}$ belongs to the distinct category s_k

The cross semivariogram $\gamma_{S_i S_j}(\mathbf{u},\mathbf{u}+\mathbf{h})$ between S_i and S_j is defined as

$$\gamma_{S_i S_j}(\mathbf{u},\mathbf{u}+\mathbf{h}) = \frac{1}{2}\{E[I(\mathbf{u};s_i) - I((\mathbf{u}+\mathbf{h});s_i)][I(\mathbf{u};s_j) - I((\mathbf{u}+\mathbf{h});s_j)]\} \quad (22)$$

and measures how often two locations separated by a vector $\mathbf{h}$ belong to the different categories s_i and s_j.

In reality the theoretical models shown above are unknown and need to be inferred from the experimental semivariogram and cross semivariogram using the data points $\{\mathbf{u}_\alpha, \alpha = 1,...,N\}$

The experimental semivariogram for the category S_i is:

$$\gamma_{S_i}^*(\mathbf{h}) = \frac{1}{2N} \sum_{|\mathbf{u}_\alpha - \mathbf{u}_\beta| = \mathbf{h}} [i(\mathbf{u}_\alpha;s_i) - i(\mathbf{u}_\beta;s_i)]^2$$

(23)

where the data locations $\mathbf{u}_\alpha$ and $\mathbf{u}_\beta$ are separated by the vector $\mathbf{h}$, and N denotes the number of pairs of data that are separated by each vector $\mathbf{h}$.

Similarly, the formula for the experimental cross semivariogram is given by

$$\gamma^*_{S_i S_j}(\mathbf{h}) = \frac{1}{2N} \sum_{|\mathbf{u}_\alpha - \mathbf{u}_\beta| = \mathbf{h}} [i(\mathbf{u}_\alpha; s_i) - i(\mathbf{u}_\beta; s_i)][i(\mathbf{u}_\alpha; s_j) - i(\mathbf{u}_\beta; s_j)] \qquad (24)$$

2.3 Variogram models

For estimation and simulation semivariogram values need to be computed at separation distances other than those available experimentally. To facilitate this, semivariogram models are inferred from the experiment semivariogram. In this thesis, models for indicator semivariograms as well as semivariograms of continuous variables are required. The following models defined in their isotropic form will be used.

The nugget effect model is defined as

$$g_0(h) = \begin{cases} 0 & \text{if } h=0 \\ 1 & \text{otherwise} \end{cases} \qquad (25)$$

And the spherical model (with a sill value of 1 reached at range a) is defined as

$$Sph_a(h) = \begin{cases} 0 & h=0 \\ \dfrac{3}{2}\dfrac{h}{a} - \dfrac{1}{2}\left(\dfrac{h}{a}\right)^3 & h \le a \\ 1 & h > a \end{cases} \qquad (26)$$

2.4 Kriging

Kriging is a generic term for a family of generalised least squares regression algorithms named after Danie Krige's pioneering work in the field of geostatistics. The basic linear regression estimator $Z^*(\mathbf{u})$ is defined as

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] \qquad (27)$$

15

where $\lambda_\alpha(\mathbf{u})$ is the weight assigned to datum $z(\mathbf{u}_\alpha)$ interpreted as a realisation of the RV $Z(\mathbf{u}_\alpha)$. The quantities $m(\mathbf{u})$ and $m(\mathbf{u}_\alpha)$ are the expected values of the RVs $Z(\mathbf{u})$ and $Z^*(\mathbf{u})$. $Z(\mathbf{u})$ is treated as a random function with a trend component, $m(\mathbf{u})$, and a residual component $R(\mathbf{u}) = Z(\mathbf{u}) - m(\mathbf{u})$. Kriging estimates the residual at $\mathbf{u}$ as a weighed sum of residuals at neighborhood data points. The goal is to determine the kriging weights, λ_α, which are derived from covariance function or semivariogram function, and which minimize the variance of the estimator $\sigma_E^2(\mathbf{u}) = \mathrm{Var}\{Z*(\mathbf{u}) - Z(\mathbf{u})\}$ under the unbiasedness constraint $E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0$. The random function (RF) $Z(\mathbf{u})$ is decomposed into residual and trend components, $Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$, with the residual component treated as an RF with stationary mean of 0 and a stationary covariance:

$$E\{R(\mathbf{u})\} = 0$$

$$\mathrm{Cov}\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} = E\{R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})\} = C_R(\mathbf{h})$$

The residual covariance function is generally derived from the input semivariogram model,

$$C_R(\mathbf{h}) = C_R(\mathbf{0}) - \gamma(\mathbf{h}) = Sill - \gamma(\mathbf{h})$$

$$(28)$$

In Simple Kriging (SK) the expected value m is known and constant, and the simple kriging estimate at a location $\mathbf{u}$ is given by

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

$$(29)$$

This is an unbiased estimator since $E[Z(\mathbf{u}_\alpha) - m] = 0$. The estimation error $Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u})$ is a linear combination of random variables representing residuals at data points, $\mathbf{u}_\alpha$, and the estimation point, $\mathbf{u}$:

$$Z_{SK}^*(\mathbf{u}) - Z(\mathbf{u}) = [Z_{SK}^*(\mathbf{u}) - m] - [Z(\mathbf{u}) - m]$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u})R(\mathbf{u}_\alpha) - R(\mathbf{u}) = R_{SK}^*(\mathbf{u}) - R(\mathbf{u})$$

$$(30)$$

16

Using the rules for the variance of a linear combination of random variables, the error variance is then given by

$$\sigma_E^2(\mathbf{u}) = \mathrm{Var}\{\mathrm{R}_{SK}^*(\mathbf{u})\} + \mathrm{Var}\{\mathrm{R}_{SK}(\mathbf{u})\} - 2Cov\{\mathrm{R}_{SK}^*(\mathbf{u}), \mathrm{R}_{SK}(\mathbf{u})\}$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) \lambda_\beta^{SK}(\mathbf{u}) \mathrm{C}_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mathrm{C}_{\mathrm{R}}(\mathbf{0}) - 2\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{SK}(\mathbf{u}) \mathrm{C}_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u})$$

(31)

To minimise the error variance, the derivative of the above expression is taken with respect to each of the kriging weights and then equated to zero. This leads to the system of equations:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{SK}(\mathbf{u}) \mathrm{C}_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}_\beta) = \mathrm{C}_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1,...,n(\mathbf{u})$$

(32)

For SK the mean is constant so the covariance function for $Z(\mathbf{u})$ is the same as that for the residual component, $C(\mathbf{h}) = \mathrm{C}_{\mathrm{R}}(\mathbf{h})$ so the kriging system can be written in terms of $C(\mathbf{h})$:

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

(33)

This can be written in matrix form as

$$\mathbf{K}\lambda_{SK}(\mathbf{u}) = \mathbf{k} \tag{34}$$

where $\mathbf{K}_{SK}$ is the matrix of covariances between data points, with elements $\mathbf{K}_{i,j} = C(\mathbf{u}_i - \mathbf{u}_j)$, $\mathbf{k}$ is the vector of covariances between the data points and the estimation point, with elements given by $\mathbf{k}_i = C(\mathbf{u}_i - \mathbf{u})$, and $\lambda_{SK}(\mathbf{u})$ is the vector of simple kriging weights for the surrounding data points. If the covariance model is permitted and no two data points are co-located, then the data covariance matrix is positive definite and kriging weights can be determined from the equation:

$$\lambda_{SK} = \mathbf{K}^{-1}\mathbf{k} \tag{35}$$

Once the kriging weights are established, both the kriging estimate and the kriging variance can be calculated.

In Ordinary Kriging (OK) the expected value $m(\mathbf{u})$ is considered to fluctuate over the study region and is calculated by limiting the domain of stationarity of the mean to the local neighbourhood centred on the location $\mathbf{u}$ being estimated. In this case the linear estimator (29) is determined as a linear combination of the $n(\mathbf{u})$ RVs $Z(\mathbf{u}_\alpha)$ plus the constant but unknown local mean $m(\mathbf{u})=m$.

$$Z*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\right]m(\mathbf{u}) \tag{36}$$

The unknown local mean $m(\mathbf{u})$ is filtered from the linear estimator by forcing the kriging weights to sum to 1. The ordinary kriging estimator $Z_{OK}^*(\mathbf{u})$ is thus written as a linear combination of the $n(\mathbf{u})$ RVs $Z(\mathbf{u}_\alpha)$:

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u})Z(\mathbf{u}_\alpha) \text{ with } \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) = 1 \tag{37}$$

In order to minimize the error variance subject to this unit-sum constraint on the kriging weights, a system is set up which involves the error variance plus an additional term involving a Lagrange parameter $\mu_{OK}(\mathbf{u})$:

$$L = \sigma_E^2(\mathbf{u}) + 2\mu_{OK}(\mathbf{u})[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u})] \tag{38}$$

18

so that minimization with respect to the Lagrange parameter forces the constraint to be obeyed:

$$\frac{1}{2}\frac{\partial L}{\partial \mu} = 1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{OK}(\mathbf{u}) = 0$$

(39)

In this case, the system of equations for the kriging weights is

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{OK}(\mathbf{u}) = C_R(\mathbf{u}_\alpha - \mathbf{u}) \qquad \alpha = 1,....,n(\mathbf{u}) \\ \\ \qquad\qquad \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases}$$

(40)

where $C_R(\mathbf{h})$ is the covariance function for the residual component of the variable. When dealing with Simple Kriging, $C_R(\mathbf{h}) = C(\mathbf{h})$ due to the stationarity of the mean, and although that equality does not strictly hold here, in practice this substitution is often made on the assumption that the semivariogram, from which $C(\mathbf{h})$ is derived, effectively filters the influence of large-scale trends in the mean.

This OK method is used in an indicator context with the categorical variables in the following section.

2.4   *Indicator kriging of categorical variables.*

Indicator kriging (IK) uses the basic kriging paradigm but in the context of an indicator variable. In this thesis, indicators are interpreted as hard categories. The first step in indicator kriging is to model the uncertainty about a category $s_k, k = 1,...N$ of the categorical variable s at the unsampled location $\mathbf{u}$. This uncertainty is modelled by the conditional probability distribution function (cpdf) of the discrete RV $S(\mathbf{u})$:

$$p(\mathbf{u};s_k \mid (n)) = p\{S(\mathbf{u}) = s_k \mid (n)\}$$

(41)

where the conditioning information consists of n neighbourhood categorical data $s(\mathbf{u}_\alpha)$. Each conditional probability $p(\mathbf{u};s_k \mid (n))$ is also the conditional expectation of the class-indicator RV $I(\mathbf{u};s_k)$:

$$p(\mathbf{u};s_k \mid (n)) = E\{I(\mathbf{u};s_k) \mid (n)\}$$

(42)

The kriging estimator for the indicator RV is defined as

$$[I(\mathbf{u};s_k)]^* - E\{I(\mathbf{u};s_k)\} = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u};s_k)[I(\mathbf{u}_\alpha;s_k) - E\{I(\mathbf{u}_\alpha;s_k)\}]$$

(43)

where $\lambda_\alpha(\mathbf{u};s_k)$ is the weight assigned to the indicator datum $i(\mathbf{u}_\alpha;s_k)$ interpreted as a realization of the indicator RV $I(\mathbf{u}_\alpha;s_k)$.

As for the previous section, there are two indicator kriging (IK) variants; simple indicator kriging (sIK) where the indicator mean $E\{I(\mathbf{u}_\alpha;s_k)\}$ is considered constant over the study region, and the version used in this thesis, ordinary indicator kriging (oIK), where local fluctuations of the indicator mean $E\{I(\mathbf{u}_\alpha;s_k)\}$ are taken into account.

Effectively indicator kriging produces a conditional probability distribution for each category at each grid location in the study region. From this probability distribution a category needs to be derived and so there must be a post-processing classification algorithm applied to the output of IK. There are a number of different post-processing classification algorithms, such as maximum likelihood. These algorithms tend to over-represent the category (lithotype) with the greatest global proportion and under-represent or ignore the category of least global proportion (Goovaerts, 1997 p.441)

The classification method of the indicator kriging output that will be used for the purposes of comparison in this thesis is the algorithm developed by Soares (1992). In this algorithm each grid location is allocated a category under the constraint that the global proportions are reproduced. Initially the most frequently occurring category is allocated to those locations where it has greatest conditional probability until the global proportion of that category is met. This continues through the hierarchy of proportions of categories until all grid locations have been allocated a category. This algorithm allows for categories with small global proportions to be represented adequately and does not allow categories with large global proportions to dominate, which would occur in a classification algorithm such as maximum likelihood.

2.5   Sequential indicator simulation.

The SIS method is commonly used for the stochastic modelling of categorical (e.g. rock types) and also continuous attributes (e.g. porosity and permeability). SIS simulates the spatial distribution of the mutually exclusive categories (lithotypes) s_k conditional to the data set $\{s(\mathbf{u})_\alpha, \alpha = 1........, n\}$. A simplified work flow of the procedure, (Goovaerts, 1997) is as follows:

1. Define a random path visiting each node of the study area grid only once.
2. At each node $\mathbf{u}'$:
    (a) Use ordinary indicator kriging to determine the conditional probability of occurrence of each category s_k, $[p(\mathbf{u}'; s_k \mid (n))]$. The ($n$) conditioning information consists of the neighbourhood original indicator data and previously simulated indicator values.
    (b) Build a cdf-type function by adding the corresponding probabilities of occurrence of the categories
    (c) Draw a random number p uniformly distributed in [0,1]. The simulated category at location $\mathbf{u}'$ is the one corresponding to the probability interval that includes p.
    (d) Add the simulated value to the conditioning data set.
    (e) Proceed to the next node along the random path and repeat steps (a) to (d).

Repeat the entire sequential procedure with a different random path to generate another realisation.

2.6   The Gaussian function as a model for simulation

The SIS method discussed above is a simulation method where no assumption about the prior distribution at a node is made. It is often the case that a model is assumed and in many cases this is a model where the spatial law (multivariate distribution) of the RV $Z(\mathbf{u})$ is assumed to be Gaussian in nature.

A random variable Z has a Gaussian distribution if its probability density function (pdf) is:

$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$$

(44)

where μ is the mean and σ^2 the variance of the pdf.

This distribution is fully characterized by μ and σ^2 and so the problem of determining the ccdf at any point in A reduces to that of estimating these two parameters.

If $\{Z(\mathbf{u}), \mathbf{u} \in A\}$ is a standard multivariate Gaussian RF with covariance $C_Z(\mathbf{h})$ then the following properties are true:

1. All subsets of the RF Z are also multivariate Gaussian.
2. The two point distribution of any pairs of RVs $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$ is normal and fully determined by the covariance function $C_Z(\mathbf{h})$
3. All conditional distributions of any subset of the RF $Z(\mathbf{u})$ are normal. In particular the conditional distribution of a single variable given the $n(\mathbf{u})$ data $z(\mathbf{u}_\alpha)$ is normal and fully characterised by its two parameters, mean and variance:

$$G(\mathbf{u}; z \mid (n)) = G\left(\frac{z - E\{Z(\mathbf{u}) \mid (n)\}}{\sigma\{Z(\mathbf{u}) \mid (n)\}}\right)$$

(45)

where G is the standard normal cdf.

Under the multiGaussian model, the mean and variance of the ccdf at location $\mathbf{u}$ are equal to the simple kriging estimate and simple kriging variance. The ccdf (10) is then defined as

$$G(\mathbf{u} : z \mid (n))_{SK} = G\left(\frac{z - z_{SK}(\mathbf{u})}{\sigma_{SK}(\mathbf{u})}\right)$$

(46)

where $z_{SK}(\mathbf{u})$ and $\sigma^2_{SK}(\mathbf{u})$ are the simple kriging mean and variance.

2.7   Turning Band method of non-conditional simulation.

The turning band method is a non-conditional simulation algorithm that will allow the Gaussian variables to be simulated in a three dimensional space. It consists of taking the Gaussian covariance function $C_3(\mathbf{u})$ determined from the variogram model of the underlying Gaussian random function in 3 dimensional space and determining the corresponding covariance function $C_1(\mathbf{u})$ in one dimension (Armstrong et al., 2003). Next, a set of n lines with arbitrary directions in 3D space is determined. Along each line a Gaussian random function with the determined 1D covariance function $C_1(\mathbf{u})$ is simulated. The projections $\mathbf{x}_i$ at each of the n lines from any point $\mathbf{x}$ in the 3D study region are determined and the n simulation values $s(\mathbf{x}_i)$ at $\mathbf{x}$ are used to compute the grid value as

$$s(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(\mathbf{x}_i)$$

(47)

Figure 3 shows 8 lines along which the 1D Gaussian covariance function has been simulated. The grid value for the point (7,5) shown on the plane will be determined as the sum of the simulated values obtained from the projections of this point onto the simulated values from the eight 1D lines.

Once this unconditional simulation has been completed, the model is conditioned by the locations where data values occur. The error between the data value and the unconditional simulation value at those data locations is kriged using the same model used to generate the unconditional simulation.
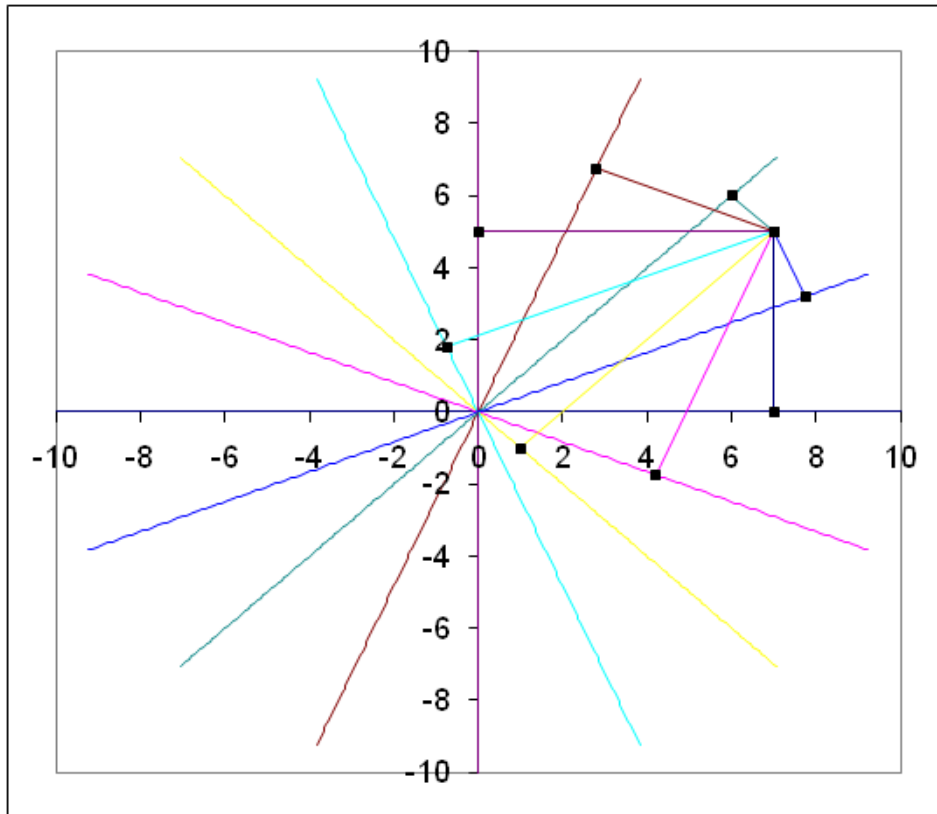
Figure 3   Projections from a point 1D covariance functions

2.8   The truncated Gaussian and Plurigaussian models

The truncated Gaussian method has been widely used when data are required to be separated into a number of thresholds (indicator coding) and the Gaussian function is assumed to be a suitable modelling tool (see section 2.6) . In this thesis the truncated Gaussian method is used with reference to the allocation of categorical (rocktype) proportions. The method is based on the notion of truncation of a standard Gaussian random function by different thresholds calculated from experimental set proportions and was first applied to model the lithotype in deltaic reservoirs by Matheron et al. (1987). The covariance model of the underlying Gaussian function is obtained by an iterative process of inversion of the covariance model of the experimental indicator variables of the sets. If the categorical data are sequential in their contacts then truncation of one Gaussian RF is required, however if two Gaussian functions are required to be truncated then this is the plurigaussian case, discussed in the sections below.

In truncated Gaussian simulation, a Gaussian random function z is simulated first and then it is transformed into the set of indicator variables by truncation. In the case of two sets, S_1 and S_2, being simulated, a numerical threshold value t_1 is produced so that:

$$P(z \leq t_1) = P(S(\mathbf{u}) = s_1) \quad \text{and} \quad P(z > t_1) = P(S(\mathbf{u}) = s_2)$$

When N sets are being simulated, (N-1) thresholds are necessary. The set S_i is associated with a subinterval $[t_{i-1}, t_i]$ of the domain of the Gaussian pdf such that

$$P(t_{i-1} \leq z \leq t_i) = P(S(\mathbf{u}) = s_i) \tag{48}$$

It is important to have the thresholds reflect the ordering of the sets

ie. $t_1 \leq t_2 \leq t_3 \leq t_4$

From (16):

$P_S(\mathbf{u}) = E[i(\mathbf{u},s)] \quad \text{and} \quad I(s) = \Leftrightarrow t_{i-1} \leq z \leq t_i$

These two statements can be incorporated so that:

$$\tag{49}$$

$$ $$

ie $\tag{50}$

where $G(t)$ is the cdf of the standard normal distribution. As the proportion of each set is known experimentally, the thresholds can be deduced from the inversion of (1.34):

$$t_1 = G^{-1}$$

$$t_2 = G^{-1}$$

$$t_i = G^{-1} \tag{51}$$

25

If the categories under consideration are not sequential in their contacts but have a more complex relationship, the plurigaussian model may be employed. If there were a fourth category, S4, added to Figure 2 which contacted both S2 and S3 then the representation shown in Figure 2 would be modified as shown in Figure 4. This is further developed in section 3.1.
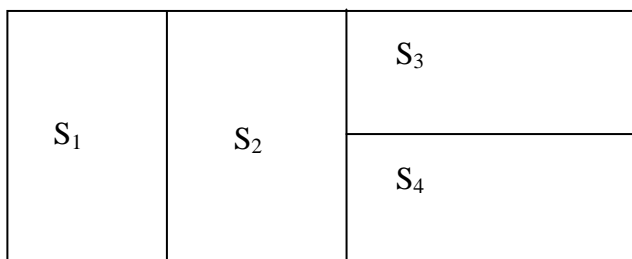
Figure 4   Representation of four categories where contacts are non-sequential.

The above model of category contacts would require the truncation of two Gaussian functions for modelling purposes, which is usual for the plurigaussian model, however the following considers the more general case where N Gaussian functions may be used.

Where N Gaussian functions may be used in the simulation a space in N dimensions is being considered for simulating sets. At a location $\mathbf{u}$ in this N-space the values of the N Gaussian functions $Z_1, Z_2, ..., Z_N$ define the co-ordinates of the simulated value. Let Q_i be the subset of the Gaussian space corresponding to the category S_i. Then

$$\text{(52)}$$

From (1.20): $P_{S_i}(\mathbf{u}) = E\{I(\mathbf{u}; s_i)\}$ and (1.36):

$$= \int_{Q_i} g_\Sigma(z_1, z_2, ..., z_N) dz_1, dz_2, ... dz_N \qquad (53)$$

where $g_\Sigma(z_1, z_2, ..., z_N)$ is the N-variate Gaussian density function with mean 0 and variance 1, and Σ is its correlation matrix.

26

2.9 The link between the indicator semivariograms and the underlying Gaussian semivariograms

The technique used in this thesis relies on a simulation of Gaussian values from the underlying Gaussian semivariogram model(s). A critical feature of this is the inference of these Gaussian semivariograms from the experimental indicator semivariograms and cross-semivariograms described in the previous section. The links between all these semivariograms are functions of the truncation process and of the conditioning to the set proportions. The semivariogram inference is based on a method described in Armstrong et al. (2003) in which the ranges of the underlying Gaussian semivariograms are adjusted iteratively. Firstly initial parameters of the underlying Gaussian models are defined and then these models are used to construct an unconditional simulation of the study area. From this simulation, indicator semivariograms of each of the sets are computed. A comparison of these computed semivariograms and the previously determined experimental semivariograms is undertaken and the parameters are adjusted iteratively until an acceptable match is obtained.

In the truncated Gaussian case, if t_{i-1} and t_i are the thresholds for the elements of set S_i then the semivariogram model is calculated using:

$$\gamma = p_{S_i} - \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t_i} g_{\rho(h)}(a,b)\, da\, db$$

$$= p_{S_i} - \frac{1}{2\pi} \int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t_i} \exp\left(-\frac{a^2 - 2\rho ab + b^2}{2(1-\rho^2)}\right) da\, db \qquad (54)$$

where p_{S_i} is the proportion of the study region where set S_i exists and $g_{\rho(h)}$ is the bigaussian density function for the same set S_i. The correlation between the two Gaussian variables a and b is $\rho(h)$. A correlation value is determined by a trial and error process in order to improve the fitting of sample indicator variograms.

The plurigaussian case is a direct extension of this. Considering two Gaussians $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$ initially, the semivariogram model is given by:

$$(55)$$

The two Gaussians give rise to four Gaussian variables, $Z_1(\mathbf{u})$, $Z_2(\mathbf{u})$, $Z_1(\mathbf{u+h})$ and $Z_2(\mathbf{u+h})$. Here Σ represents the matrix of correlations and covariances between the four variables.

A covariance for a Gaussian is a fundamental parameter and can be found from the experimental data. The correlation between the two Gaussians (ρ) is a more arbitrary value. The larger the correlation value the greater the bordering effect apparent in the outputs of the simulations, and so choosing a zero correlation minimises artificially affecting the PGS output, and neither IK nor SIS consider this border effect issue. The link between the four variables depends on the choice of coregionalisation model between $Z_1(\mathbf{u})$ and $Z_2(\mathbf{u})$. A simple model is one where the two Gaussians RFs Z_1 and Z_2 are linear combinations of independent Gaussian RFs, say $B_1(\mathbf{u})$ and $B_2(\mathbf{u})$, each with a N(0,1) distribution having covariances $\rho_{B_1}(\mathbf{h})$ and $\rho_{B_2}(\mathbf{h})$ respectively.

So:

$$Z_1(\mathbf{u}) = B_1(\mathbf{u}) \quad \text{and}$$

$$Z_2(\mathbf{u}) = \rho B_1(\mathbf{u}) + (\sqrt{1-\rho^2})B_2(\mathbf{u})$$

where $\rho_{B_1}(\mathbf{h}) = \rho_{Z_1}(\mathbf{h})$ and $\rho_{B_2}(\mathbf{h}) = \rho^2 \rho_{Z_1}(\mathbf{h}) + (1-\rho^2)\rho_{Z_2}(\mathbf{h})$

The coregionalisation correlations and covariances are summarised in Table 1:

Table 1: Correlations and covariances for RFs Z_1, Z_2

	$Z_1(\mathbf{u})$	$Z_2(\mathbf{u})$	$Z_1(\mathbf{u+h})$	$Z_2(\mathbf{u+h})$
$Z_1(\mathbf{u})$	1	ρ	$\rho_{Z_1}(\mathbf{h})$	$\rho_{B_2}(\mathbf{h})$
$Z_2(\mathbf{u})$	ρ	1	$\rho_{B_2}(\mathbf{h})$	$\rho_{Z_2}(\mathbf{h})$
$Z_1(\mathbf{u+h})$	$\rho_{Z_1}(\mathbf{h})$	$\rho_{B_2}(\mathbf{h})$	1	ρ
$Z_2(\mathbf{u+h})$	$\rho_{B_2}(\mathbf{h})$	$\rho_{Z_2}(\mathbf{h})$	ρ	1

This leads to the matrix $\Sigma = \begin{pmatrix} 1 & \rho & \rho_2(h/4) & \rho_4(h/4) \\ \rho & 1 & \rho_4(h/4) & \rho_2(h/4) \\ \rho_2(h/4) & \rho_4(h/4) & 1 & \rho \\ \rho_4(h/4) & \rho_2(h/4) & \rho & 1 \end{pmatrix}$ used in (55),

which is a starting point for the iterative process of computation of the required indicator semivariograms. In Isatis the iterative process begins with selection of (in the plurigaussian case) two possible underlying Gaussian models. These models are then tested (and altered where necessary) so as to optimise the fits of the indicator variograms to the experimental data.

Chapter 3: Aspects of Plurigaussian Simulation

The simulation of categorical (set) data outlined in the preceding sections is now applied to the simulation of rock types (lithotypes) in a geological setting. The initial application of this method was to simulate rock types in a sedimentary deposit (Armstrong, 2003)

The fundamental concept of non-stationary proportion curves is central to the model, where the rock type rule plays an essential role in producing geologically realistic models that represents the transitions between the different lithotypes.

The advantages of this model are theoretical consistency and the ability to incorporate external data such as vertical proportion curves. Truncated Gaussian simulation is an important theoretical starting point: pluriGaussian simulation is an extension of the truncated Gaussian model which allows the acquisition of two or more Gaussian variables, correlated or not, that describe different spatial behaviours of a lithotype or group of lithotypes as well as incorporating non-sequential contacts in the model.

Depending on the number of lithotypes, there is a finite number of rectangular partitions that may represent the relationships between the Gaussian random functions and the different lithotypes. The most sensible geologically speaking is often evident from the vertical proportion curves in the layered nature of the lithotypes: one particular lithotype may only be seen to be in contact with another specific lithotype. As well as this information, global proportions of each lithotype give an indicator of proportion when selecting the *rock type* rule.

3.1   *The rock type rule.*

The rock type rule splits a rectangle into sub-rectangles and assigns a lithotype to each sub-rectangle, without reference to proportions. The sub-rectangles created determine allowable contacts between lithotypes.

To illustrate the rock type rule we will consider the case where three lithotypes are to be simulated. The ordering of the lithotypes in the rock type rule must respect the

sedimentary or contact sequence in order to correctly reproduce the geology of the study area in terms of rock type. For truncated Gaussian simulation the single Gaussian function is partitioned into three regions consistent with the proportions calculated in the vertical proportion curve and the sequential nature of the lithotypes. For the case of truncated plurigaussian simulation non-sequential contacts of lithotypes are permitted and so a number of different possibilities exist, for example where both F_2 and F_3 are in contact with F_1.

These regions are diagrammatically represented by rectangles partitioned into proportions as shown in Figure 5. The case where truncated Gaussian simulation is appropriate is shown in (a), a scenario where PGS is appropriate is shown in (b).

(a)   (b)
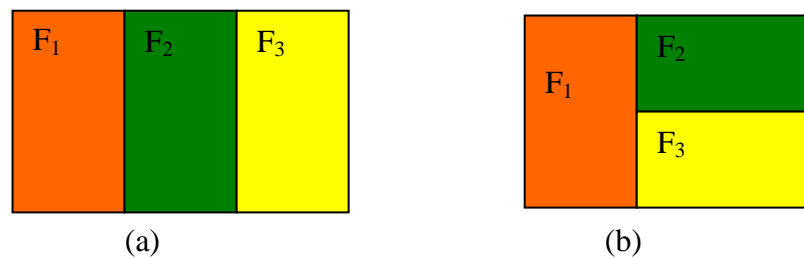
Figure 5 Possible contacts for three lithotypes

In both cases the horizontal line on the bottom of the rectangle represents the axis for the first Gaussian variable z_1 (G1) and for (b) in Figure 5 the vertical line on the left represents the second Gaussian variable z_2 (G2). The three colours represent the different lithotypes, and the equal areas of the rectangles indicate that the three lithotypes have equal proportions, however proportions need not be equal. The plurigaussian case shown in Figure 5 is further detailed in Figure 6.
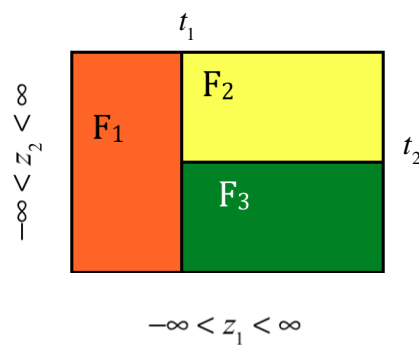
Figure 6   Plurigaussian case for three lithotypes.

In this case two thresholds, t_1 for z_1 and t_2 for z_2, are used. In Figure 6 the three lithotypes are defined within the rocktype rule by the following:

$$F_1 : \{(z_1, z_2) : z_1 < t_1, z_2 \in \mathbf{R}\}$$
$$F_2 : \{(z_1, z_2) : z_1 > t_1, z_2 > t_2\}$$
$$F_3 : \{(z_1, z_2) : z_1 > t_1, z_2 < t_2\}$$

Decreasing the value of the threshold t_1 in the first Gaussian would reduce the proportion of lithotype F_1 and increase the proportion of lithotype F_2 and F_3. Diagrams representing possible contacts for four lithotypes are shown below.
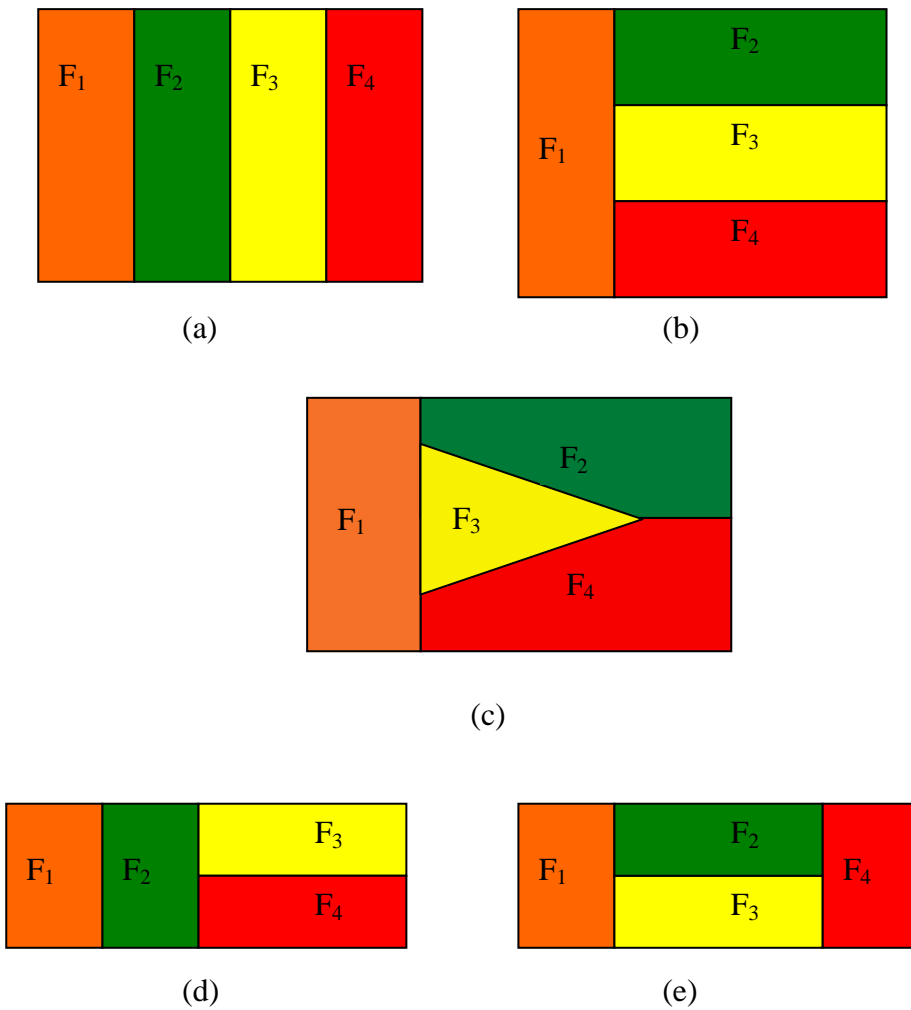
Figure 7 . Possible contacts for four lithotypes

Truncated Gaussian simulation would be applied to a situation depicted in (a) but the other lithotype contacts would require a plurigaussian treatment. Rectangular divisions of the rocktype rule are preferred as they allow thresholds to be defined in the two Gaussian functions. For diagram (c) in Figure 7 the yellow region is not defined by

edges horizontally or vertically and while possible in terms of contacts is not physically able to be modelled. If lithotype proportions vary across the study region A then rock type rules can be calculated in a grid. This would assist in realistic simulation of a horizontally non-stationary situation.

3.2 *Vertical proportion curves and reference levels.*

A basic quantity that can be deduced from a data set is the proportion of this data set that is allocated to each of the categories of interest. It is uncommon for the proportions of these categories to be either vertically or horizontally constant in A. For example, petroleum reservoirs are not stationary in the vertical direction because of cyclic changes during their deposition and non-stationarity can also occur in mining deposits (Armstrong, et al., 2003). Vertical proportion curves (VPCs) were designed to quantify these changes and it is important to analyse this non-stationarity to enable an output model to reflect the differences in these proportions in a regionally sensitive manner.

The first step is to simply count the number of occurrences of each category along lines parallel to the chosen reference level. The proportions are displayed as a graph showing the proportion of each category at each level. As it is likely that the VPCs vary horizontally depending on their location in the study region, a proportion map of VPCs is calculated over the entire study region. Initially, drill holes (wells) are grouped locally and a VPC is calculated for each set of wells. These VPCs are placed on a regular, coarse grid which covers the study region. At grid points where a VPC is not evident, an interpolation method (usually a kriging procedure) is used to generate a vertical proportion curve (with category proportions transformed where necessary so that they sum to 1). This proportion grid is used to identify local variations in category distributions.

Determination of the vertical proportion curve from the sample data is highly dependent on the choice of a reference level. This level can be defined as a guide to the system of deposition of the different categories. The sample data will be transformed into a flattened space where the reference level represents the horizontal surface at zero elevation. The simulation will be done in this artificial volume and then the simulated values transferred to real stratigraphic space.

3.3 The Gibbs sampler

This iterative algorithm (described in detail in Armstrong et al (2003) allows for local inequality constraints to be imposed on a simulation of a Gaussian random function. It is based on simulation of the underlying Gaussian function and not on the simulation of the indicator variables. Each sample data point in the study region must therefore have a Gaussian value within a predetermined area of the Gaussian function(s) that has been calculated from the lithotype proportions. Using the multivariate Gaussian approach, each lithotype data point is assigned an arbitrary starting value that is within its Gaussian thresholds. The Gibbs sampler is a process whereby these arbitrary values are then iteratively conditioned by all other regional (arbitrary) data values assigned to lithotypes until a steady state value is determined and the scores remain within their Gaussian threshold and also reflect the covariance model of the underlying Gaussian functions.

Given vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$, and corresponding initial values $z_1, \ldots, z_n$:

For $\alpha = 1$

1. Calculate the SK estimate $z_{SK}^*(\mathbf{u}_\alpha)$ and SK standard deviation $\sigma_{SK}^*(\mathbf{u}_\alpha)$ from $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \setminus \{\mathbf{u}_\alpha\}$.

2. Put $z(\mathbf{u}_\alpha) = z_{SK}^*(\mathbf{u}_\alpha) + \sigma_{SK}^*(\mathbf{u}_\alpha)r$ where $r \sim N(0,1)$

3. If $t_{\alpha-1} \leq z_\alpha \leq t_\alpha$ accept this value, otherwise draw a new r value.

Next α.

This conditioning step is repeated until a steady state is reached for each of the Gaussian values. More than 5000 iterations may be needed for the updated data to reach a steady state (Walsh, 2004).

The key issue in the successful implementation of the Gibbs sampler is the number of iterations until the values each approach their own steady state, which is the length of the burn in period. Typically the first 1000 iterations are ignored and after this a test for convergence is used to see if a steady state has been reached. A poor choice of starting values and/or proposed distribution (as defined by the model) can greatly increase the burn in time. Starting values close to the centre (mode) of the proportion in question reduce the burn in time. Iterations are mixed poorly if the values stay in a small region of the proportion, and well-mixed iterations explore the proportion randomly before approaching the steady state (Walsh, 2004).

3.4 Back-transforming the conditioned values

When the sample categorical data have been given appropriate values within the Gaussian regions, the final stage of simulation is undertaken. The values are then back-transformed into categorical data by consideration of their location within the thresholds of the Gaussian functions. This is a simple process in the case of truncated Gaussian simulation but for truncated plurigaussian simulation the values generated may fall into different regions of the Gaussians used for truncation and so the rock type rule diagram will allow back-transformation to categorical data.

As a final step the data must be relocated from the reference level to their true location so that a true three-dimensional model simulation may be generated.

Chapter 4: Exploratory Data Analysis

4.1   Background of Big Bell Gold Mine

The Big Bell gold deposit lies in the Murchison mineral field, 30km WNW of the township of Cue, which is 650 km NNE of Perth, Western Australia. Gold was discovered at Big Bell in 1904.

The Big Bell deposit is hosted by a regional volcanic (greenstone) and sedimentary sequence within the Murchison Province of the Archean Yilgarn cratonic block. At Big Bell the greenstone belt is narrow, steeply dipping, strongly attenuated and locally overturned. It forms the west limb of a north plunging structure. An open fold style is characteristic of the Big Bell regional structure. At least two generations of granitic intrusive activity are recognized in the region.

In the region of the mine, there are three major subdivisions within the volcano-sedimentary sequence: a lower mafic (rocks and minerals containing abundant magnesium and iron) sequence, a felsic (rocks and minerals rich in silica and other feldspar components such as sodium and potassium) volcanic sequence, and an upper mafic sequence.

The Big Bell mine greenstone belt is bounded by granites. The broad subdivisions described above are recognized at the Big Bell mine, where the volcano-sedimentary sequence is 1500m wide, and are structurally overturned, so that the younger upper mafic sequence now lies on the footwall of the orebody. Almost all the gold mineralisation at Big Bell is confined to the quartz-muscovite- potassium feldspar rich rocks of the ore zone.

The main orebody, mined underground and drill-indicated to a vertical depth of 1500m (Handley and Cary, 1990) forms part of a more than 1800m long and 30-70m wide zone of intensely altered amphibolite. This study will model the spatial distribution of the gold bearing altered rocks in the mined out open pit south of the main orebody, based on drill hole intersections below the base of weathering. The rock types indicated in the data are the boundary granodiorite gneiss (Lithotype 1), magnesium olivine-calcite

skarn (Lithotype 2), skarn-banded biotite (Lithotype 3), cordierite-bearing gneiss (Lithotype 4) and pyritic microcline gneiss (Lithotype 5) representing the southern tail of the main orebody.

Initial geological reports were by Woodward (1914). The staff at Big Bell Mines Limited (1953) suggested that Big Bell was a replacement deposit. More recently Chown et al. (1984) highlighted the difficulty in determining the genesis, while Philips (1985) suggested a metamorphic replacement model. Further classification and description was undertaken by Mueller et al (1991).More recently, Mueller et al (2002) modelled categorical data obtained from 42 drill holes using an ordinary indicator kriging algorithm and then a series of three different single step classification algorithms to model the rock types.

Historically, most mining at Big Bell occurred between 1937 and 1955. During that time, 5.6 megatonnes of underground ore were milled at a rate of about 30,000 tonnes/month for a recovered grade of 4 grams/tonne gold.

Redrilling of the area commenced in 1980, and was increased over a two year period when core samples from 1974 drill holes were reassayed and found to contain ore grades. In 1982 Australian Consolidated Minerals Ltd established 100% ownership of the area and in 1983 began a shallow drilling programme for assessment of possible strike extensions of the deposit. A reserve amenable to open pit mining was outlined, and an assessment of the gold recoverable from the old tailings dump was carried out.

The ACM – Placer joint venture began mine optimisation studies in 1984. A drilling programme more fully defined the deposit and tested for further reserves within the tenement. As a result of a feasibility study, the decision to reopen the mine was made in December 1987, and production commenced in February 1989. Up to 1999, when the mine closed, the mine produced 27.8 million tonnes of ore at a recovered grade of 2.38 grams/tonne from both open pit and underground mining (Mueller et al, 2002).

4.2 Data analysis

The input data for this study consist of 1501 data from a total of 42 drill holes that have been classified according to the above five lithotypes. The data were rotated in space so that the dip plane coincides with the horizontal which will facilitate modelling.

The purpose of this thesis is to simulate the five lithotypes over the study region. The motivation for consideration of lithotypes is a first step in analyzing mineral deposits in a geological region as these minerals vary greatly in grade from lithotype to lithotype. Once lithotypes are simulated mineral grades over the study region could be calculated, however this is not the basis of this thesis. Initial data analysis here simply consists of determining the positions and lithotypes associated with each of the core samples within the study region and proportions of each of the five lithotypes over the study region.

The lithotype colour scale used in this study is shown in Figure 8.

	L1 (granodiorite gneiss)
	L2 (magnesium olivine-calcite skarn)
	L3 (skarn-banded biotite)
	L4 (cordierite-bearing gneiss)
	L5 (pyritic microcline gneiss)

Figure 8 . Lithotype colour scale used in this study.

Within the drill cores, the five defined lithotypes occur in the following proportions.

Table 2 Lithotype percentages in raw data.

Lithotype	Sample count	Percentage
1	153	9.96%
2	45	2.93%
3	430	27.99%
4	700	45.57%
5	208	13.54%
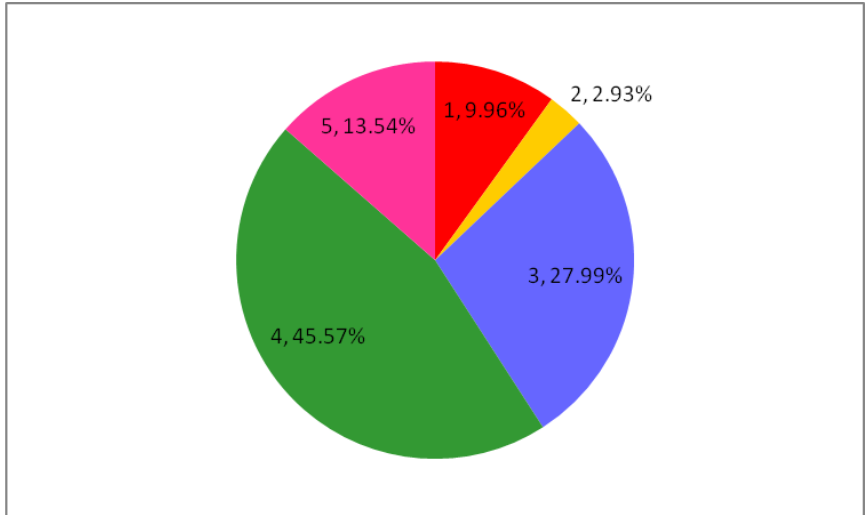
Figure 9 . Proportions of lithotypes in the study region.

When comparisons are made between the percentages of lithotypes present in the raw data and the subsequent plurigaussian simulation and ordinary kriging algorithms, lithotype 1 percentages will not be considered, as this lithotype envelops the study region, and will by necessity increase to fill the perimeter of the simulation grid.
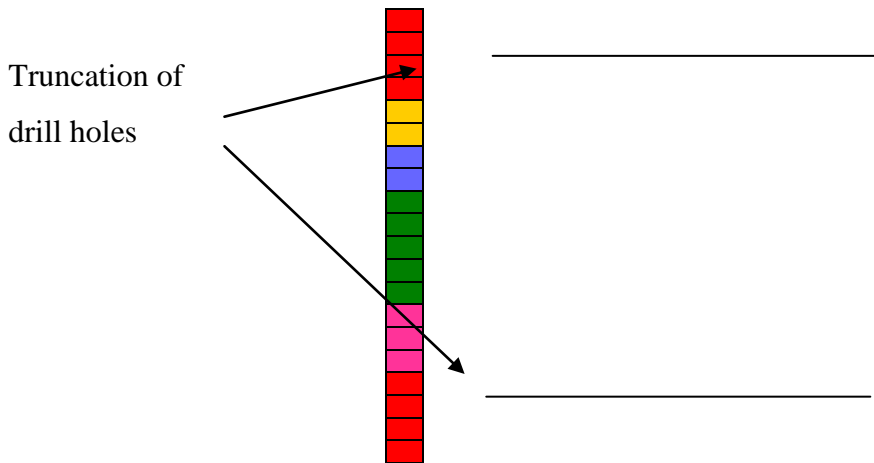
Figure 10   Drill hole truncation.

The data have been created using lithotype 1 as the bounding lithotype for the drill holes, that is each drill hole was truncated at both ends so that only one or two core measurements involving lithotype 1 were kept.
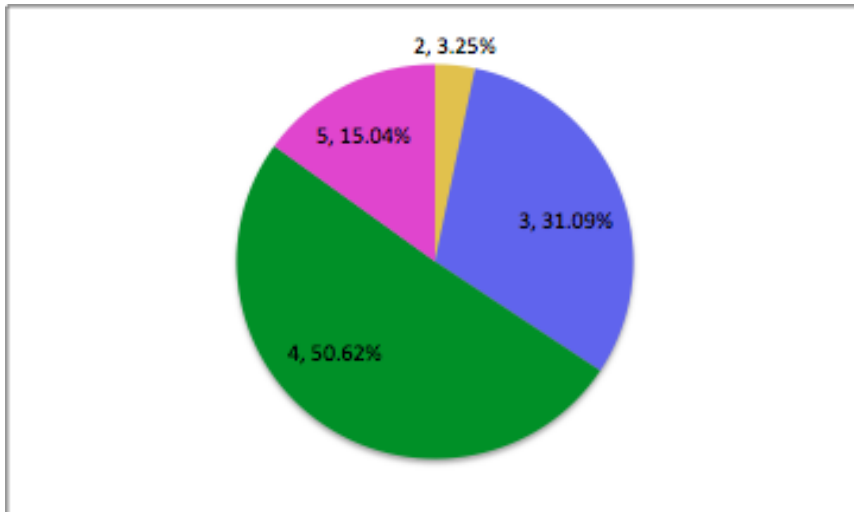
Figure 11. Proportions of lithotypes in the study region in the absence of L1.

Figure 12 shows the lines of the core samples in both perspective and the three projections. The drill holes are not vertical, although they are parallel to the z axis in the y-z projection. The data may be rotated to make the drill holes vertical, although this is not necessary for successful simulation in this algorithm.
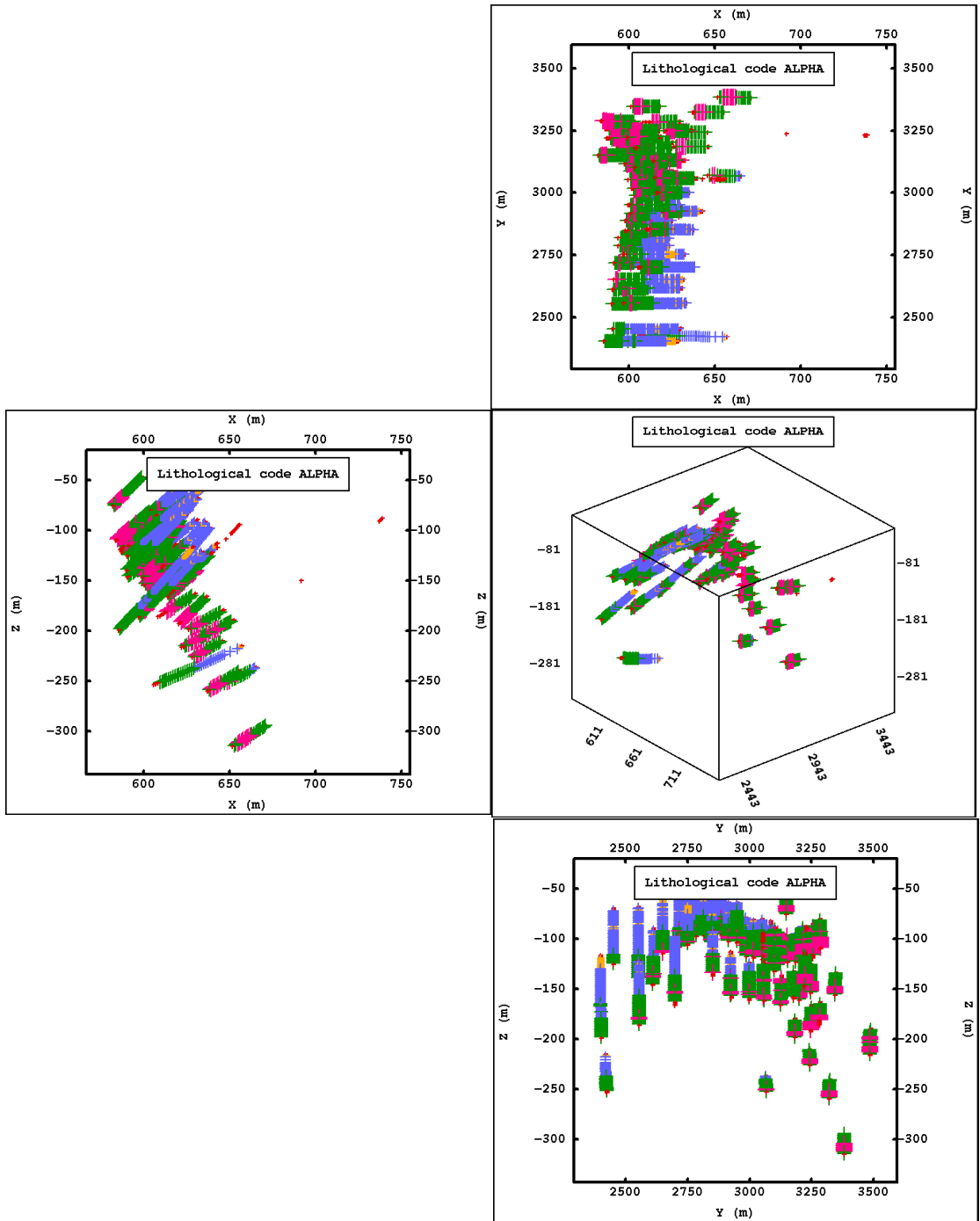
Figure 12   Projections and 3D image of data set. (Lithotype colours from Figure 8)

4.3   Variogram models for IK and SIS

The variogram models for the five categories (lithotypes) used in this thesis for the IK estimations are summarised in Table 3. Each category was modelled with a nested structure consisting of a nugget effect model and two spherical models (Mueller, 2002), the parameters of which are listed in Table 3.

Table 3 Variogram parameters used in IK for categories (Mueller et. al, 2002)

Category	Nugget	Major Continuity Direction	Sill1	Ranges 1	Sill 2	Ranges 2
1	0.25	0.00	0.25	(75,75,7)	0.50	(600,150,45)
2	0.40	0.00	0.30	(150,75,3)	0.30	(500,100,15)
3	0.12	0.00	0.20	(75,75,32)	0.68	(500,400,35)
4	0.25	90.00	0.40	(50,50,20)	0.35	(600,150,30)
5	0.35	90.00	0.50	(40,40,8)	0.15	(75,75,1000)
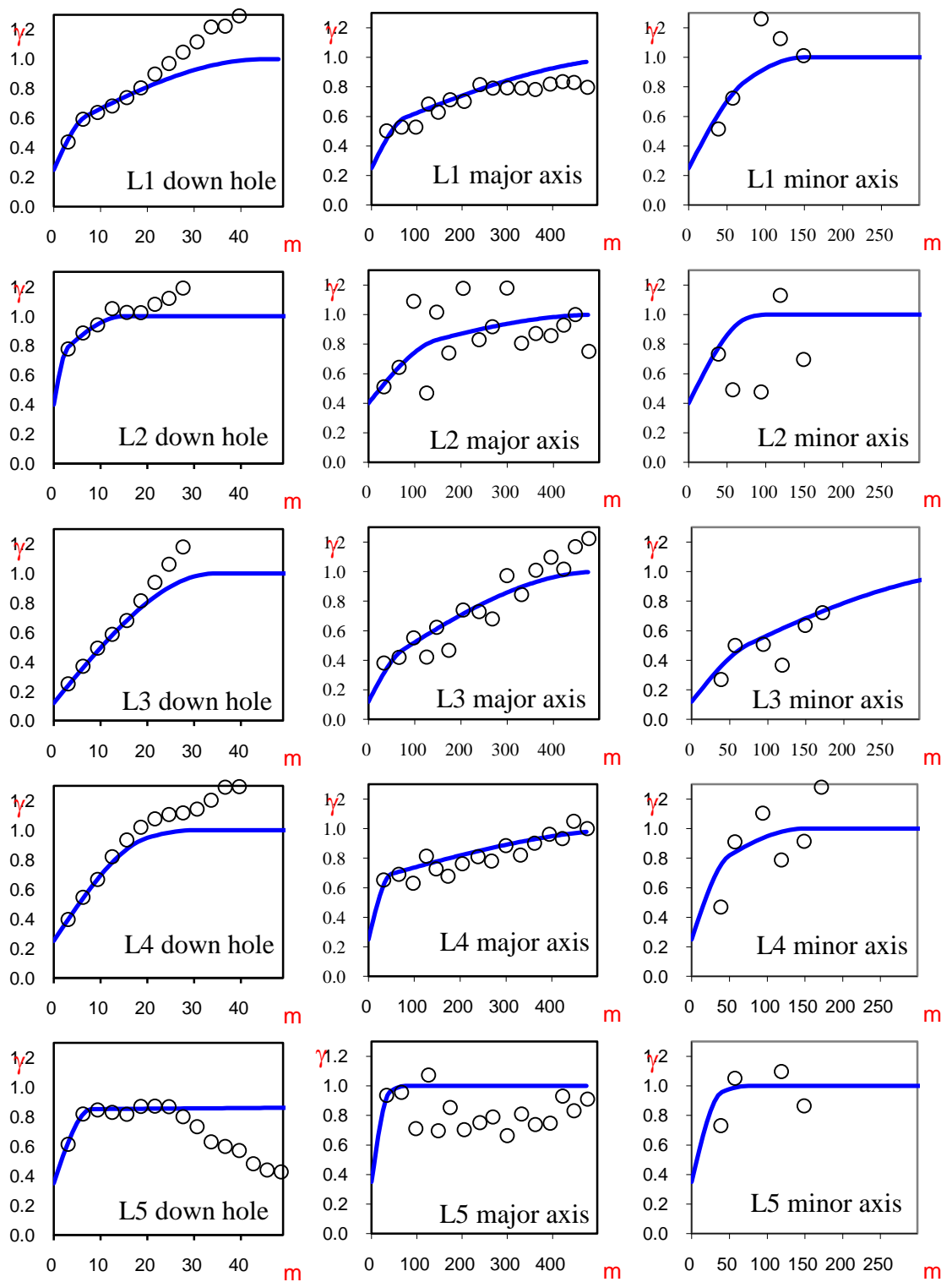
Figure 13 Variogram models for IK of the five lithotypes, L1 to L5, indicating down hole (left), major axis (middle) and minor axis (right). (Mueller, 2002).

The normalised variograms of the lithotypes L1 to L5, shown in Figure 13, all display a very good fit, especially for the first six lags, with the effective sill being modeled for those data that display this feature. The variogram model used in for SIS in this thesis is that of the category with the largest proportion in the data set: L4 (45.57% of data set). This model is a nested structure model consisting of a nugget and two spherical models defined in (54) and (55), with parameters listed in Table 3.

Chapter 5: PluriGaussian simulation

5.1 Overview of the algorithm.

Simulation of the drill hole data from the Big Bell gold mine will be conducted with reference to the gold and silver grade in each of the five lithotypes and also the distribution of lithotypes within the study region. Analysis and simulation of data are conducted using the geostatistical package Isatis, produced by the French company Geovariances.

The plurigaussian algorithm requires a number of sequential steps to be performed. These steps are indicated in the flowchart shown in Figure 14. Detailed elaboration of each of the steps outlined will be done in the following chapter.

Initially, the data set from the drill holes is loaded as a lines file, with a header file indicating the location and name of each drill hole. The drill holes are defined within a three dimensional structural grid of cells, and horizontal surfaces are defined as vertical boundaries of the structural grid. Each drill hole is discretised, or divided into equal length sections, within the structural grid, and the cells are referenced to the lower boundary surface. A number of the discretised lines of data may pass through a single structural grid cell, and so a single lithotype is assigned to each discretised cell within the grid. (Isatis automatically converts the defined structural grid into a working grid for the purposes of calculation). The horizontal plane of the structural grid is divided into polygons and lithotype proportions in each of these polygons are calculated. Vertical proportion curves (VPCs) are also calculated within these polygons and these VPCs are extrapolated (completed to the bounding horizontal surfaces) and smoothed. These VPCs, together with the defined rocktype rule, allow the selection of underlying Gaussian random functions that will produce a model which will be used to simulate Gaussian values for lithotypes at unsampled locations. As a final step, the simulated Gaussian values are back-transformed to the lithotypes to be displayed in the simulation within the structural grid.
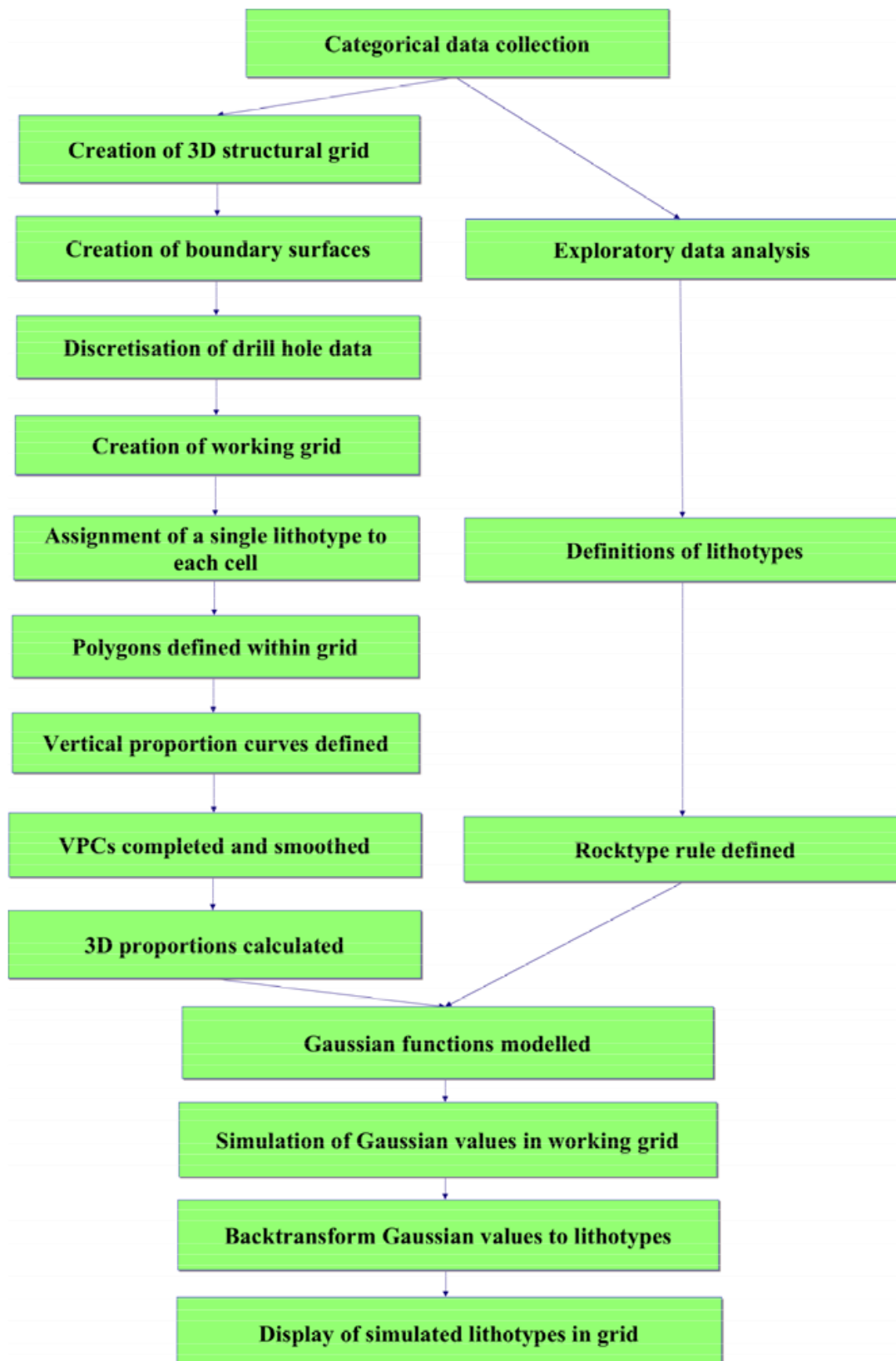
Figure 14 A flowchart of the plurigaussian algorithm.

The first step in the plurigaussian algorithm is the creation of a 3D structural grid in the study region within which the simulation will be displayed. Secondly, the line style drill hole data are discretised along their lines and referenced to a plane. Next, an analysis of proportions of each lithotype in each cell in the 3D grid is undertaken using VPCs, alongside the creation of the rocktype rule.

5.2 Defining the horizontal boundary surfaces of the structural grid.

The ranges for the co-ordinates of the drill hole data are displayed below. As can be seen from Figure 12, a large proportion of the data lies within the x co-ordinate range 570m – 650m. Upper and lower horizontal surfaces are defined as a first step towards defining a three dimensional grid which encapsulates a large proportion of the available data. The surfaces are at $z = -40$m and at $z = -210m$ and these surfaces bound 91.6% of the drill hole data set. Figure 15 displays these surface files with the line data.
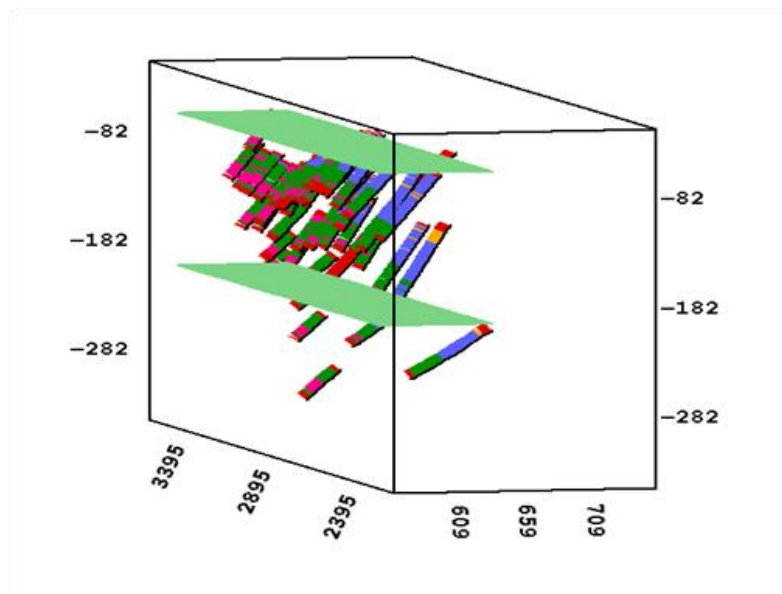
Figure 15   Lines and surfaces

The structural grid is now defined. The horizontal characteristics of the grid *(xo,dx, nx, yo, dy, ny)* are the same as for the surfaces, with the vertical dimensions reaching from surface to surface with a mesh of 2 metres. This creates a grid with 168 000 cells, each of dimensions 4m by 10m by 2m. This structural grid will be transformed into a working grid for the simulation and then the simulation in this working grid will be back-transformed into the structural grid for the display of lithotypes.

5.3 Discretisation and Flattenning

The drill hole data are not regular in their segmentation along each line and so each line is discretised into equal segments of 0.5m, which is appropriate as it is equal to the mesh of the structural grid. As the horizontal lag of the structural grid cells is greater than the vertical lag, a distortion ratio is applied to the cells. The distortion ratio is calculated by a division of the product of the horizontal extension values (dx,dy) by the vertical extension value (dz). Here, $dx = dy = 2$ and $dz = 0.5$, thus the distortion ratio is 8.

The working grid is now created and is the structure within which the analysis of lithotype proportions of the discretised data takes place. The working grid has the same characteristics as the structural grid. In this thesis the working grid has been transformed parallel to the lower surface and has a new proportions Macro variable in a new discretised well file in which the proportion of each lithotype within each cell of the structural grid is stored (and updated subsequently).

The next step is to choose a method of assigning a single lithotype to each cell of the (discretised) working grid. There are three choices available in Isatis, central (the lithotype in the middle of the discretized core), most representative (the lithotype with the greatest proportion within the core) and random (a lithotype is randomly assigned from those present in the discretised core). In this study the central method has been used.

5.4 Analysis and estimation of lithotypes

The proportions of lithotypes in each drill hole can be displayed in pie graphs as shown in Figure 16. Broadly, it is clear lithotypes 4 and 5 occur in larger proportions in the North of the study region ($y > 3000m$) and lithotype 3 occurs in larger proportions in the South of the study region ($y < 2900m$) where lithotype 5 is absent.
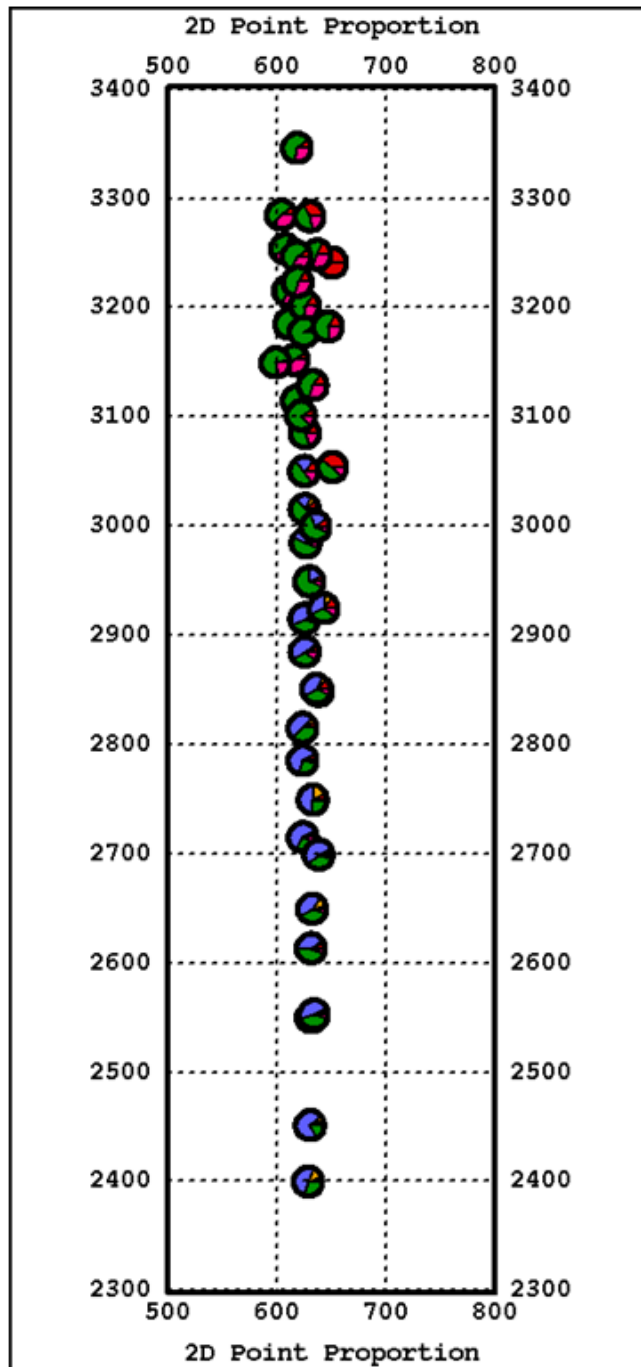
Figure 16   Pie charts indicating regional lithotype proportions.

In the analysis of the data displayed in the working grid in Figure 16, two processes require consideration: *polygon definition* and creation of *proportion curves*.

Polygon definition allows the working grid to be sectioned into a number of polygons that are 3D sections of the structural grid much larger than the horizontal mesh. The balance here is to form enough polygons within the study region so that regional sensitivity of lithotype proportions can be assessed whilst not dividing the 42 drill holes into polygons where similar VPCs are calculated. The longest axis for the data is the in the y direction, and so five polygons were created, splitting this axis into four equal areas in the north and one larger area the more sparsely populated south ($y < 2650m$), allowing each polygon to be populated with at least five drill hole cores. The Proportions of lithotypes can be calculated from VPCs within each of the (five) polygons defined, allowing for the indication of specific lithotypes in particular areas of the working grid and also to update the lithotype Proportions Macro file in a regionally sensitive manner.

The next step in analysis of the lithotype proportions is to consider the VPC that can be generated in each defined polygon of the study region. Each VPC is calculated from surface to surface in the structural grid, so it needs to be completed using a linear extrapolation of the top and/or bottom informed layers.

Once completed, the VPCs are normalised at each layer and also smoothed by running a filtering algorithm a number of times through the data. Five passes of the filter have been run in the examples shown. Figure 17 shows the working grid split into the five polygons, each with its normalised, completed and smoothed VPC, with the global VPC in the bottom right corner. Five polygons were defined to cater for the different lengths and starting heights of the drill holes in the study region.

The individual VPCs are shown in Figure 18. Each VPC is reproduced with raw proportions on the right and with normalised, completed and smoothed proportions on the left. Areas above and below each raw VPC are generally completed with lithotype 1, the bounding lithotype of the study region. This will allow for the production of more realistic simulations.
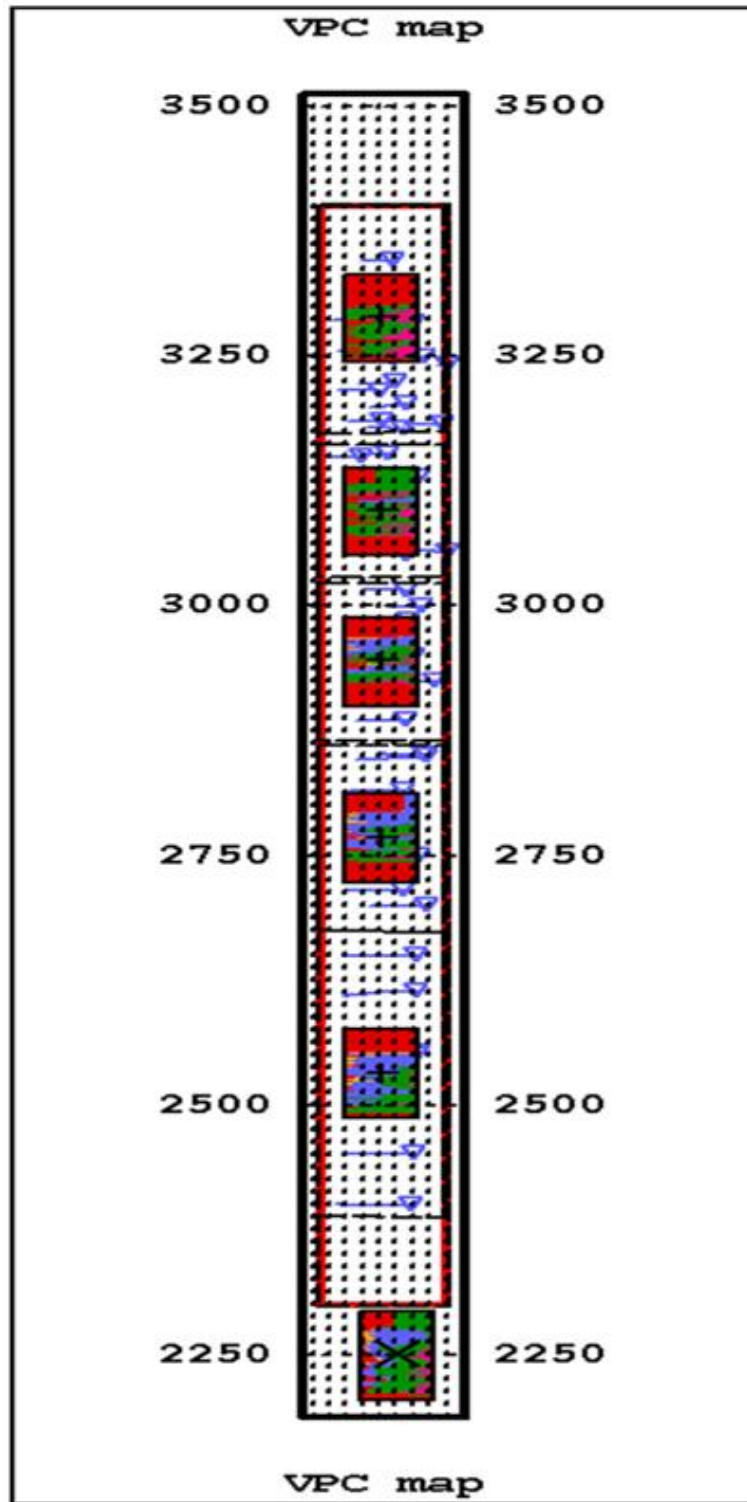
Figure 17 Five polygons and their completed VPCs in the study region.
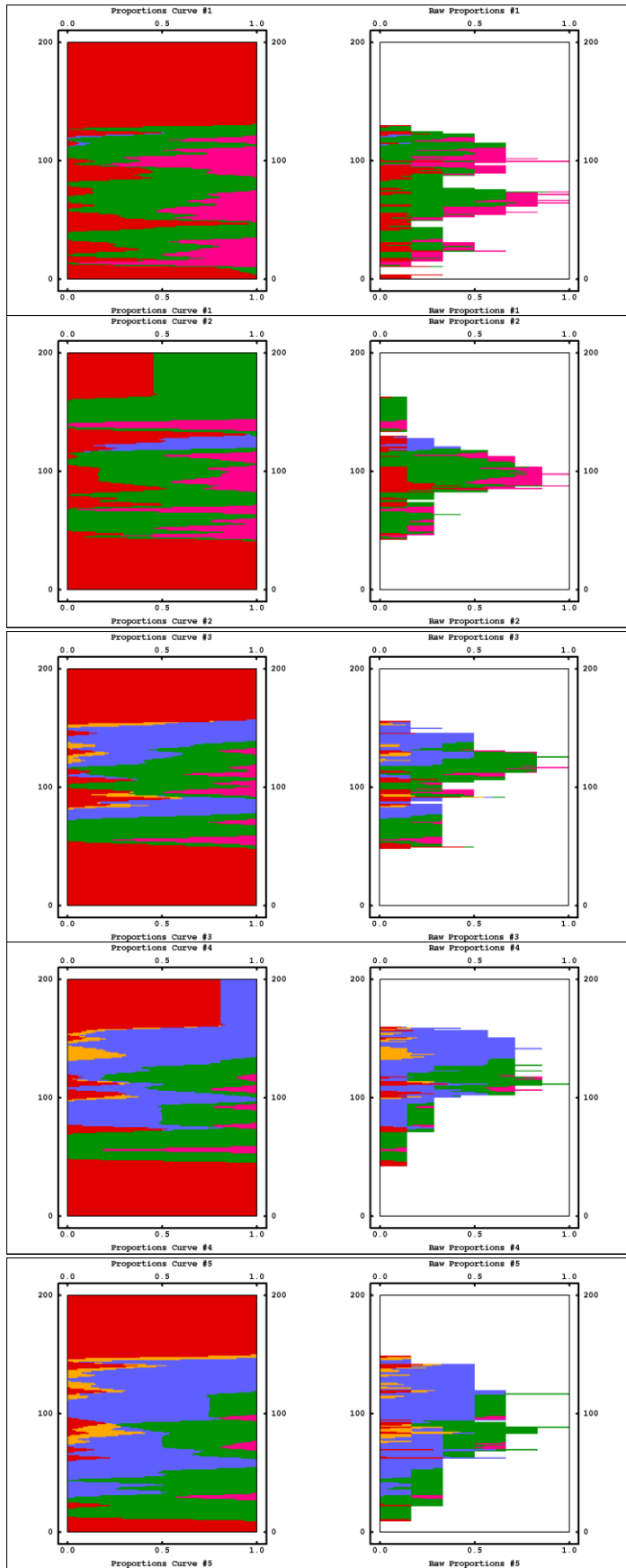
Figure 18   Raw (right) and completed (left) VPCs.

Once the VPCs from each polygon are completed, normalised and smoothed, they can be used in the subsequent interpolation procedures. It should be noted that normalising of the VPCs has caused possible artefacts in diagrams two and four in Figure 18, where the bordering lithotype L1 is not the only lithotype at the top of the completed VPC. This may be the cause of inadmissible contacts occurring in some of the simulations (section 6.3).

An interim step is to calculate the 3D lithotype proportions on the working grid. Initially, these proportions were set to the global proportions. They are now calculated more accurately in the working grid using the inverse squared distance interpolation, using the VPCs as constraining data. For each cell at a particular level in the working grid, the proportion of each lithotype is a linear combination of the proportions in all the VPCs, with the weights of the combination being proportional to the inverse squared distance between each VPC and the target cell. Other possible procedures for the interpolation are a user defined 2D kriging model or a layer-by-layer global proportion calculation.

Displaying each VPC over the entire working grid is difficult, so a horizontal projection (3D proportion map) of a coarse grid of these VPCs is shown. The VPCs are averaged layer by layer in moving windows, with extension for this moving window of 2 units. Each VPC in this grid is an average of 10 VPCs in the x direction with the origin of this axis at rank 5, and an average of 20 VPCs in the y direction with the origin for this axis at rank 10. Figure 19 displays the 3D proportion map, which again shows the absence of lithotype 3 in the North and its appearance in the South, together with the emergence of lithotype 2.
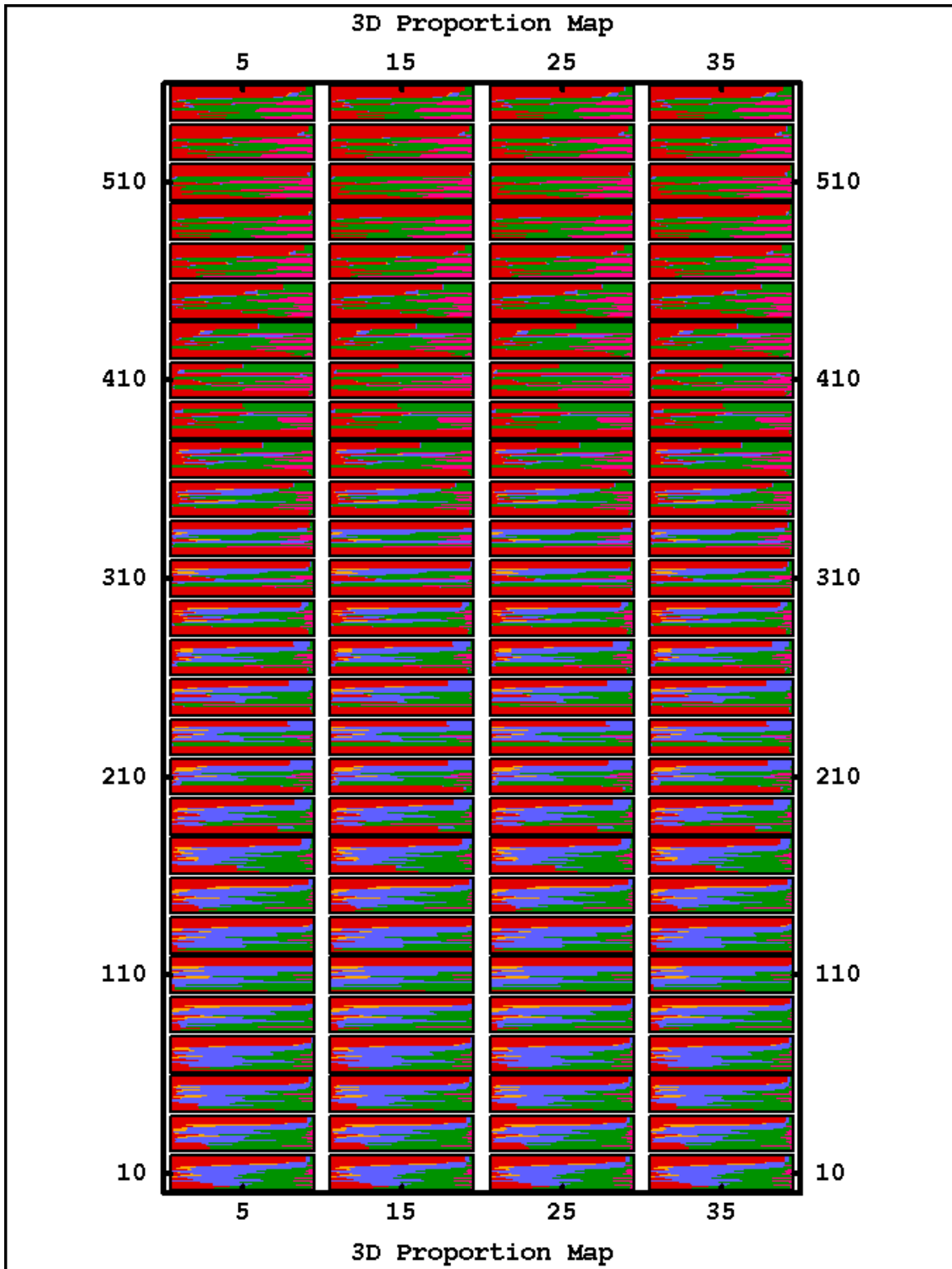
Figure 19   3D proportion map

54

5.5 Rocktype rule and Gaussian random functions.

Once lithotype proportions in each cell have been calculated, determination of the underlying Gaussian random functions and also the allowable transitions between lithotypes must be defined.

If the lithotype contacts are purely sequential, then only one Gaussian random function needs to be truncated to simulate the lithotypes. In this study, whilst lithotypes 2,3, 4 and 5 are generally sequential in their contacts (L2 contacts L3 but not L4 or L5 for example), they also all contact lithotype 1.

Transition matrices (Table 4 and Table 5) indicate the proportions of contacts between the five lithotypes along the drill cores. They assist in the development of the rocktype rule. The indicator variograms between lithotypes are analysed and this information allows for the underlying Gaussian random functions to be modeled subsequently. When all the parameters have been defined, conditional plurigaussian simulations can be run.

Table 4 . Downward matrix of transition proportions.

	Number	L1	L2	L3	L4	L5
L1	158	0.753	0.025	0.120	0.089	0.013
L2	59	0.000	0.542	0.458	0.000	0.000
L3	519	0.000	0.044	0.906	0.050	0.000
L4	781	0.045	0.000	0.004	0.880	0.736
L5	220	0.023	0.000	0.000	0.241	0.736

Table 5   Upward matrix of transition proportions.

	Number	L1	L2	L3	L4	L5
L1	159	0.748	0.000	0.000	0.220	0.031
L2	59	0.068	0.542	0.390	0.000	0.000
L3	519	0.037	0.052	0.906	0.006	0.000
L4	780	0.018	0.000	0.033	0.881	0.068
L5	220	0.009	0.000	0.000	0.255	0.736
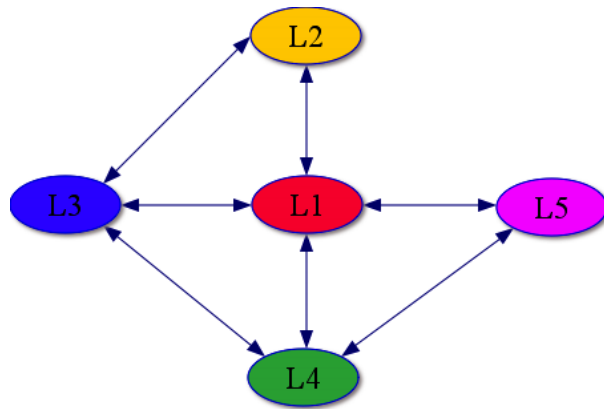
Figure 20 Lithotype contact map

The contact proportions indicatd in the above matrices and the lithotype contact map are summarised in the rock type rule shown in Figure 21 both in terms of relative proportion and physically allowable contacts. The rock type rule has three sections: contacts between lithotypes on the left, the five lithotypes in the centre and their associated summary contact histograms on the right. It can be seen from the matrices that L1 has contact with all other lithotypes and this is indicated as proportions in the bottom right histogram and also in the left hand panel of Figure 21, where L1 (red) is in contact with the other four lithoptypes. L2 (orange) only contacts L1 (red) and L3 (blue), L3 only contacts L1, L2 and L4 (green), L4 only contacts L1, L3 and L5 (pink), and L5 only contacts only L1 and L4.

Figure 21   The rocktype rule definition panel

The horizontal axis is modelled by one Gaussian random function (G1) and is truncated once to model L1 against the combined proportion of the other four lithotypes, whereas the vertical axis defines the second Gaussian random function (G2), which is truncated at three points to model proportions of L2, L3, L4 and L5.

The model for each underlying Gaussian random functions is determined by an iterative process described in section 2.7. Variogram fitting for the lithotypes in the Isatis software is an indirect, iterative process. Underlying Gaussian function models are defined and, together with the rocktype rule and the threshold parameters, theoretical variograms are tested against the experimental lithotype variograms until an appropriate fit for all experimental lithotype variograms is produced. The outcome of the iterative process is shown in Fgures 23, 24 and 25, where the indicator variograms (lines) for the unconditional simulation are compared to the experimental indicator variograms (dotted line) for each lithotype. The two underlying Gaussian models, *G1* (horizontal axis) and *G2* (vertical axis), are defined in Figure 22, where one threshold (t_1) is required for *G1* between lithotype L1 and the other four lithoptyes combined, and three thresholds (t_2, t_3 and t_4) are required for *G2* between lithotypes L2, L3, L4 and L5.
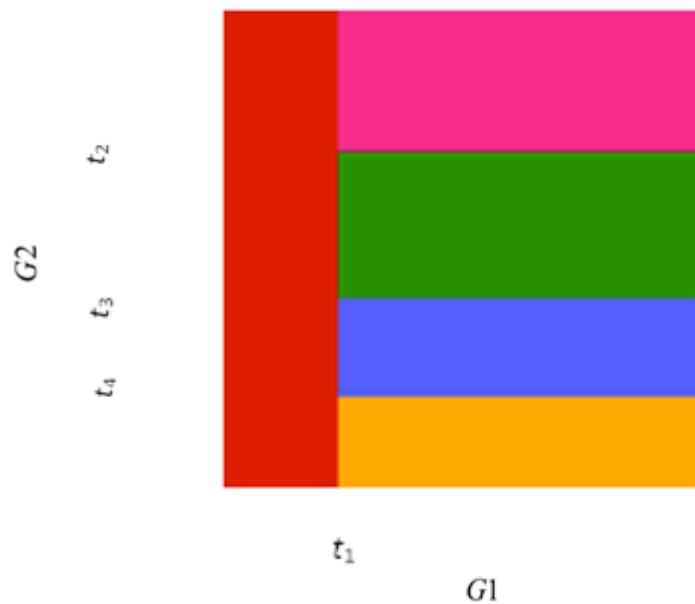
Figure 22   Underlying Gaussians and associated thresholds for PGS simulation.

Table 6 describes the Gaussian model parameters used in the PGS process. Each Gaussian is defined using the spherical model defined in (26). The fit of the experimental variogram models is satisfactory with the exception of L4 (green) in Figure 25. This is the least adequate variogram model of all lithotypes, with the short lag fit being overestimated and the longer range lags being underestimated, however the iterative process demands that a balance be reached in all variograms, and the underlying model parameters (Table 6) allowed a good fit for the remaining lithotype variograms.

Table 6 Parameters for underlying Gaussian models.

Gaussian model	Model	Sill	Range	Anisotropic Scale Factors
G1	Spherical	3.1	110	(7.8,110,20)
G2	Spherical	1.2	8	(7.8,50,16)
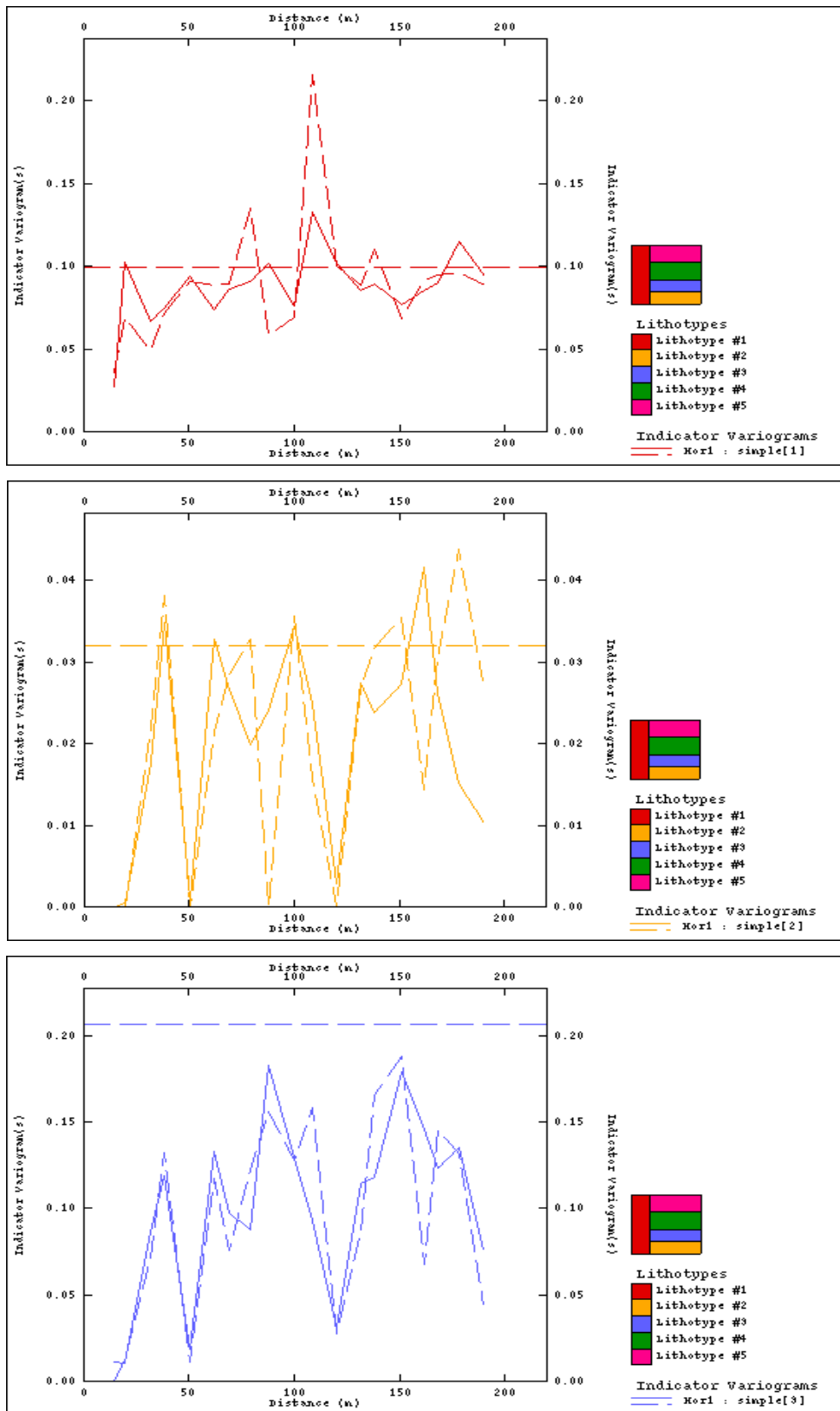
Figure 23   Experimental (dashed) and model (solid) horizontal variograms for lithotypes for L1 (top), L2 (middle) and L3 (bottom).
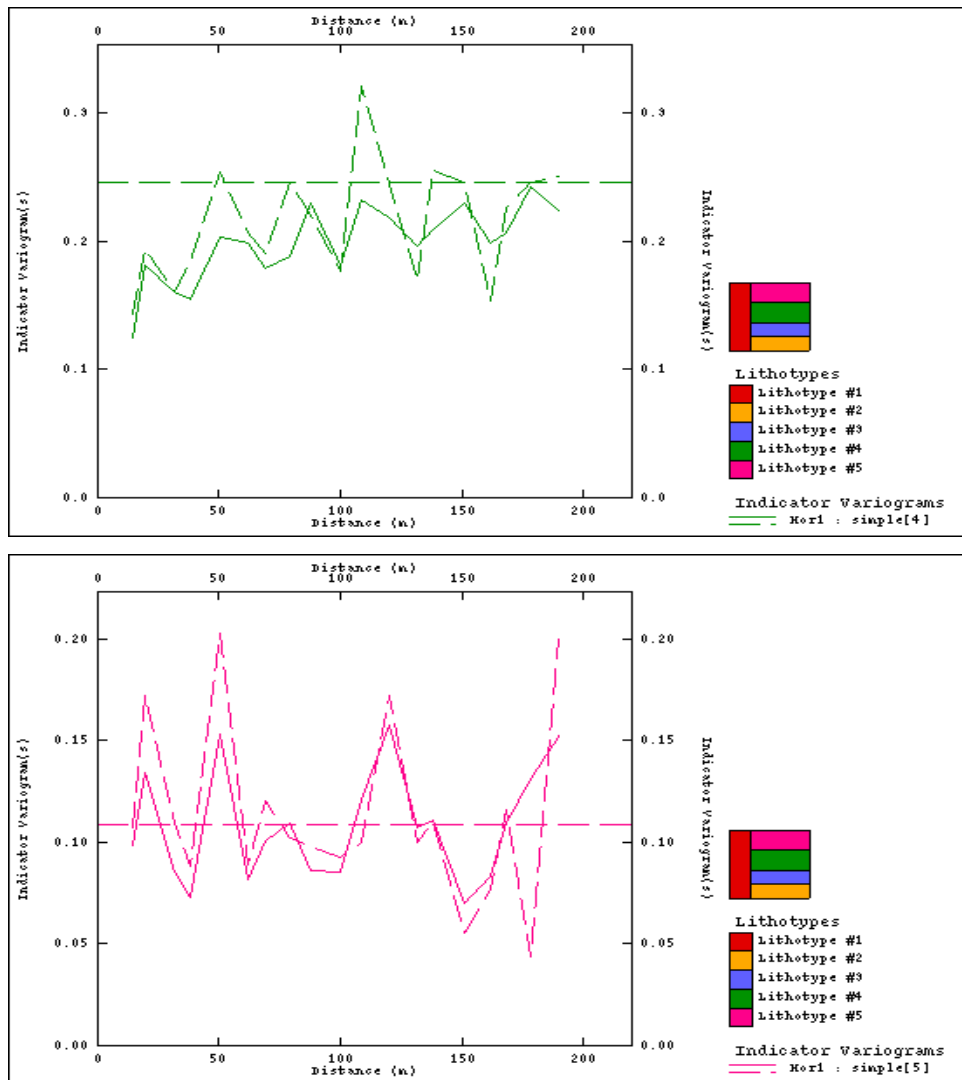
Figure 24   Experimental (dashed) and model (solid) horizontal variograms for lithotypes L4 (top) and L5 (bottom)
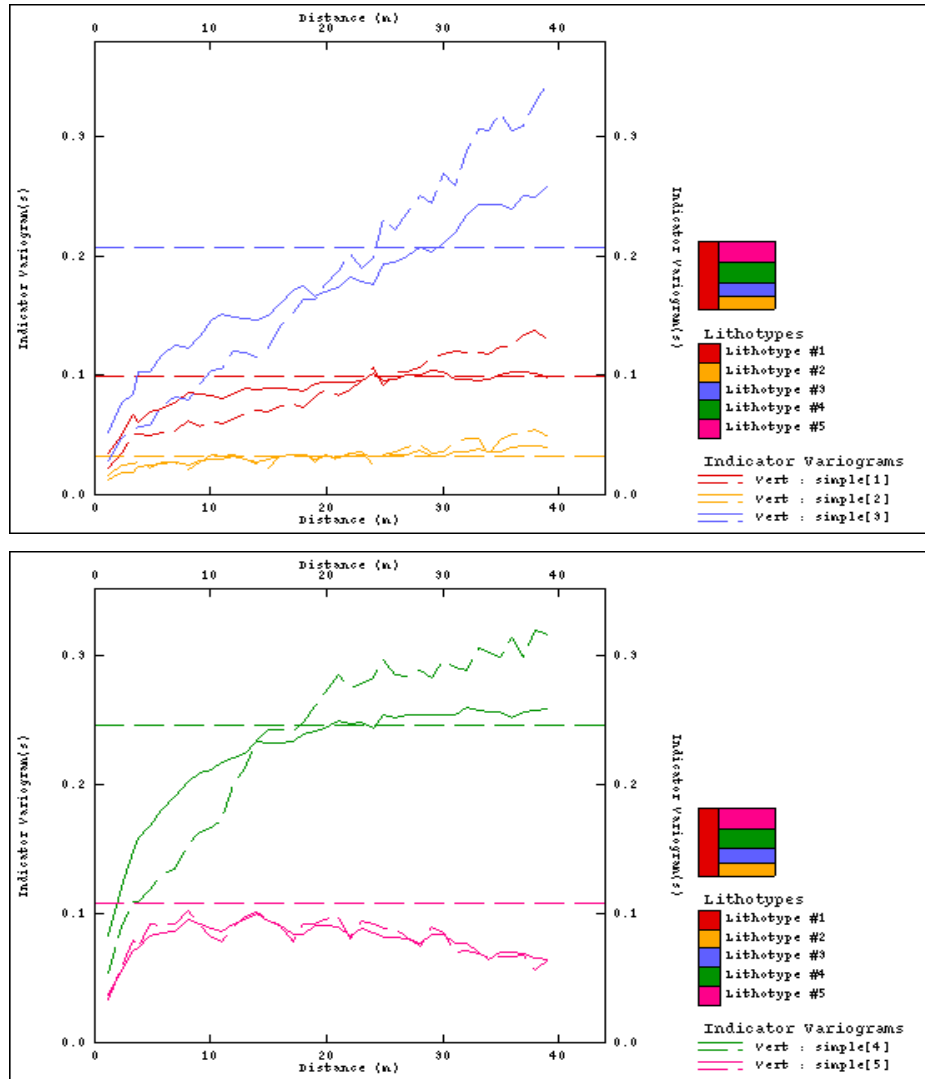
Figure 25 . Experimental (dashed) and model (solid) vertical variograms for lithotypes 1, 2and 3 (top) and lithotypes 4 and 5 (bottom).

5.6 Conditional plurigaussian simulation

Once the rocktype rule and the variogram models have been defined, simulations conditioned on drill hole data can now be performed. The moving neighbourhood search ellipsoid is defined at this stage. This defines the region used for selection of neighbourhood samples used for the kriging step. The Gibbs Sampler has been employed to allocate normal score values for the lithotypes at data points. A turning bands algorithm with 100 bands is employed to initially determine an unconditional simulation from the underlying covariance function and then to a conditional simulation is determined from a kriging of the error between the conditioning data at known locations and the unconditional simulation. These simulated Gaussian values are finally back-transformed to lithotypes. A cutaway 3D grid representation is shown in Figure 26, followed by horizontal slices at 10m depth intervals through the study region, beginning from $z = -110m$ (top left) through to $z = -30m$ (bottom right). These slices will be used as a basis for the physical comparison with both the IK and SIS models.
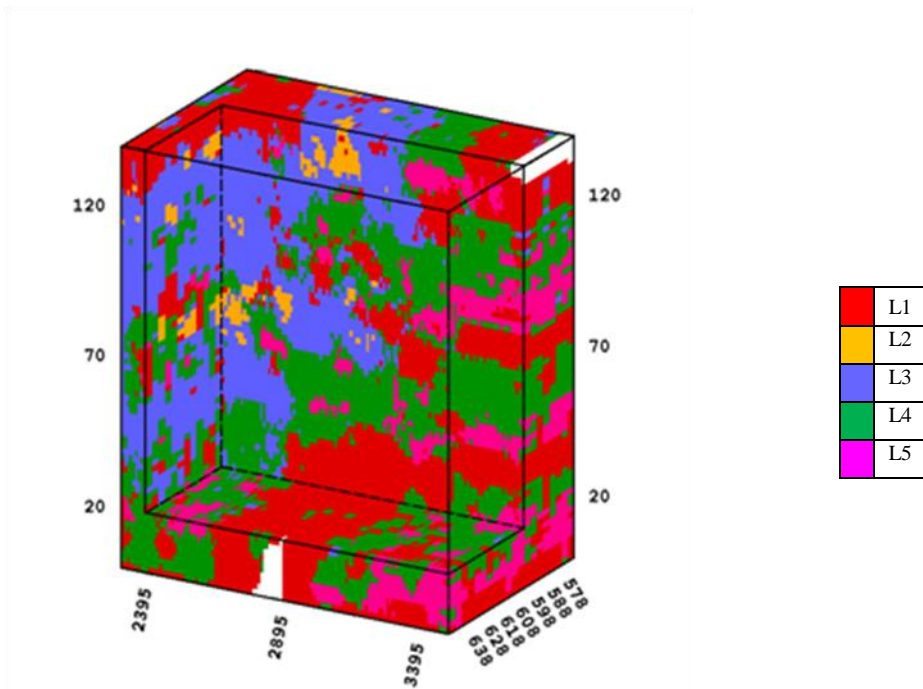
Figure 26   3D cutaway representation of the conditional PGS over the study region.
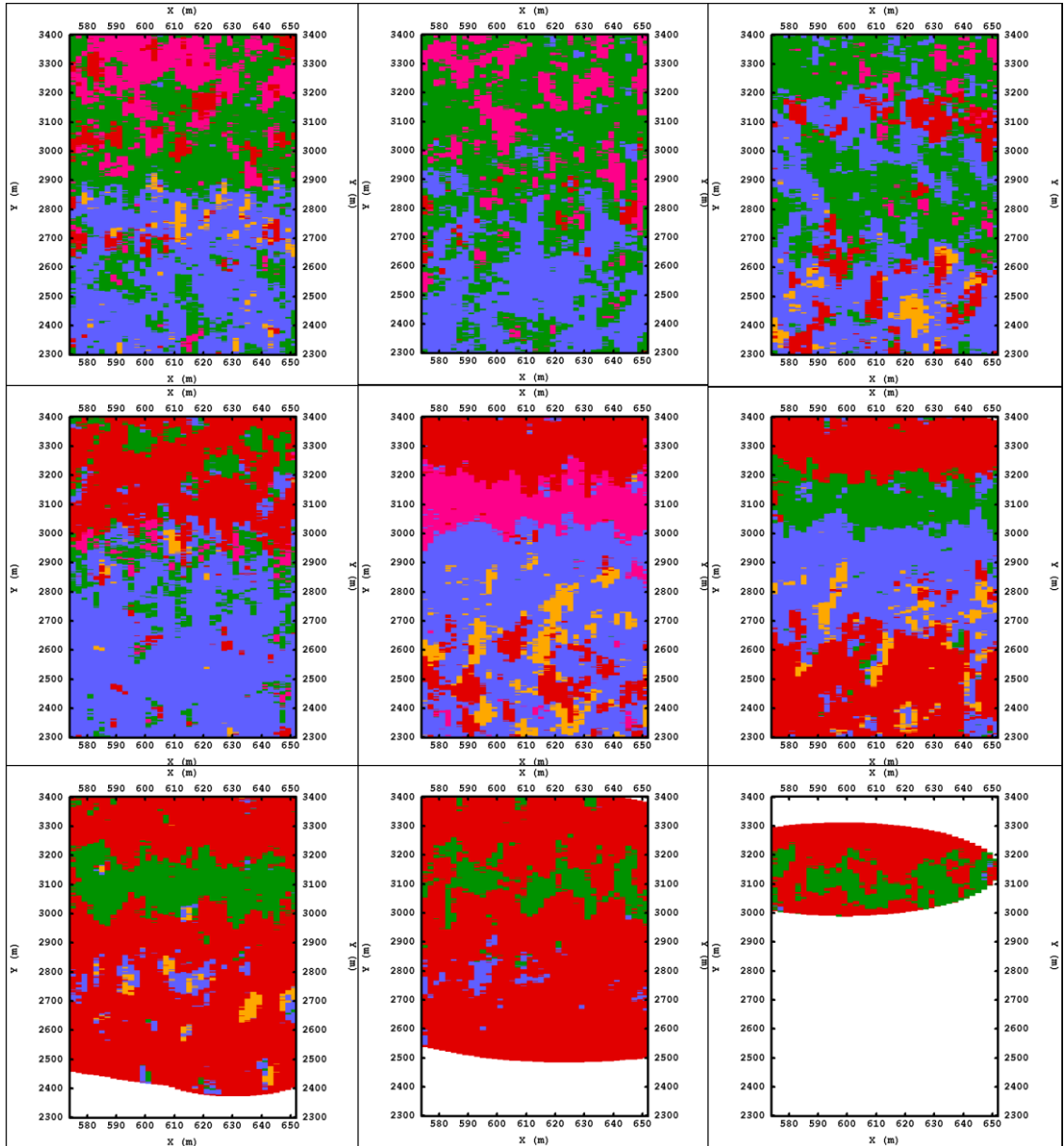
Figure 27 . Horizontal slices at 10m intervals of PGS output in the study region from $z = -110m$ (top left) to $z = -30m$ (bottom right)

The above figure reflects a number of features seen in the original data as displayed in Figure 12. Lithotype 4 (green) predominates in the north (high y values) and in the south lithotype 3 (blue) is more evident at high z values transitioning to lithotype 4 at lower z levels in this region. Lithotype 2 (orange) has the least global proportion (2.93%) in the raw data and importantly it has not been removed from the simulation through under-estimation.

Figure 28 displays the variations inherent in the simulation process. The six diagrams show a horizontal slice at $z = -110m$ from the output of six different simulation runs of the plurigaussian algorithm. There is a consistency in both the proportions of lithotypes present and their relative positions within the slice.
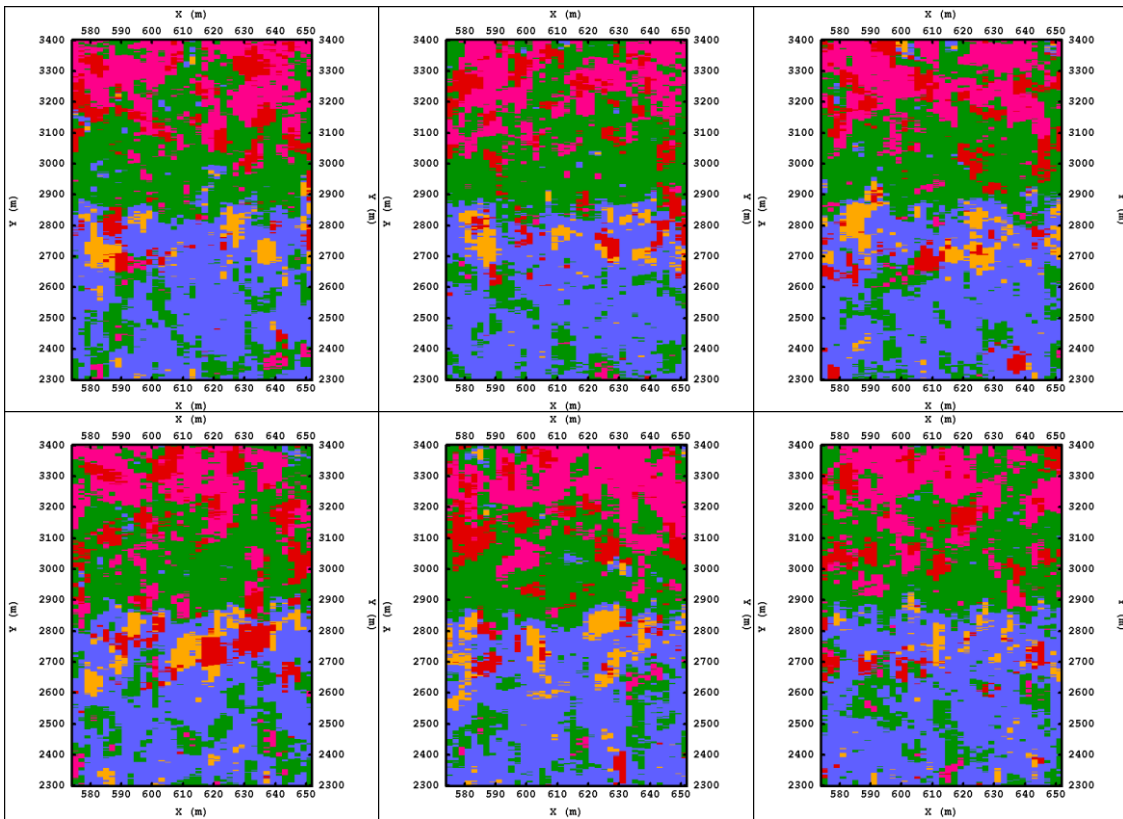
Figure 28. Six different realisations of PGS at depth $z = -110m$

Chapter 6: Comparison of output of models

In order to assess the differences between the three models in this study, three comparisons will be made: visual inspection, regional rock type proportions and a calculation of the numbers of lithotype contacts violating the rocktype rule (a reflection of lithotype contacts in the experimental data).

6.1 Visual comparison between simulations and estimation.
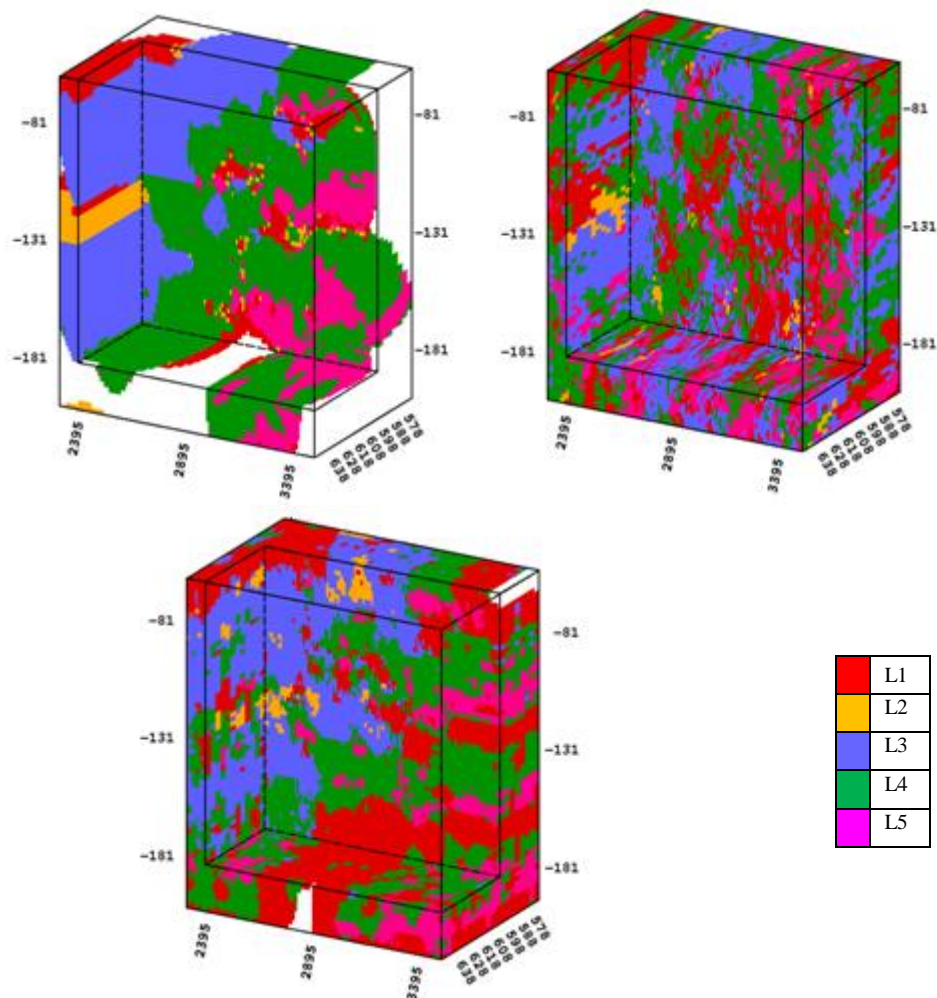
Figure 29 Excavated 3D projections of IK (top left), SIS (top right) and PGS (bottom)

The three models shown in Figure 29 display regionally similar lithotype distributions, with PGS having less of the lithotype clustering effect evident in IK yet also a less fine

65

mosaic effect evident in SIS. This suggests a more balanced and regionally sensitive output on a global scale for PGS although it is clear that IK and PGS have particularly similar spatial distributions of L5 (pink) and L3 (blue). Whilst L2 occurs in the same thin mid-level band in all models, the PGS model places L2 within one distinct upper region of the study area whereas SIS and IK models spread very small pockets of L2 through the entire the study region.

Figures 30 and 31 clearly show the unrealistic clustering of lithotypes from the IK model as well as random allocations of L1 over the study region. The SIS slices show a gradual change of character as z values increase whilst the PGS outputs are more definite in displaying the regional lithotype proportions seen in Figure 12 (lower panel). All models display varying degrees of so-called inadmissible lithotype contacts (contrary to the contacts appearing in the sample data and defined in the rock type rule), generally when lithotype 2 (orange) contacts either lithotype 4 (green) or 5 (pink).
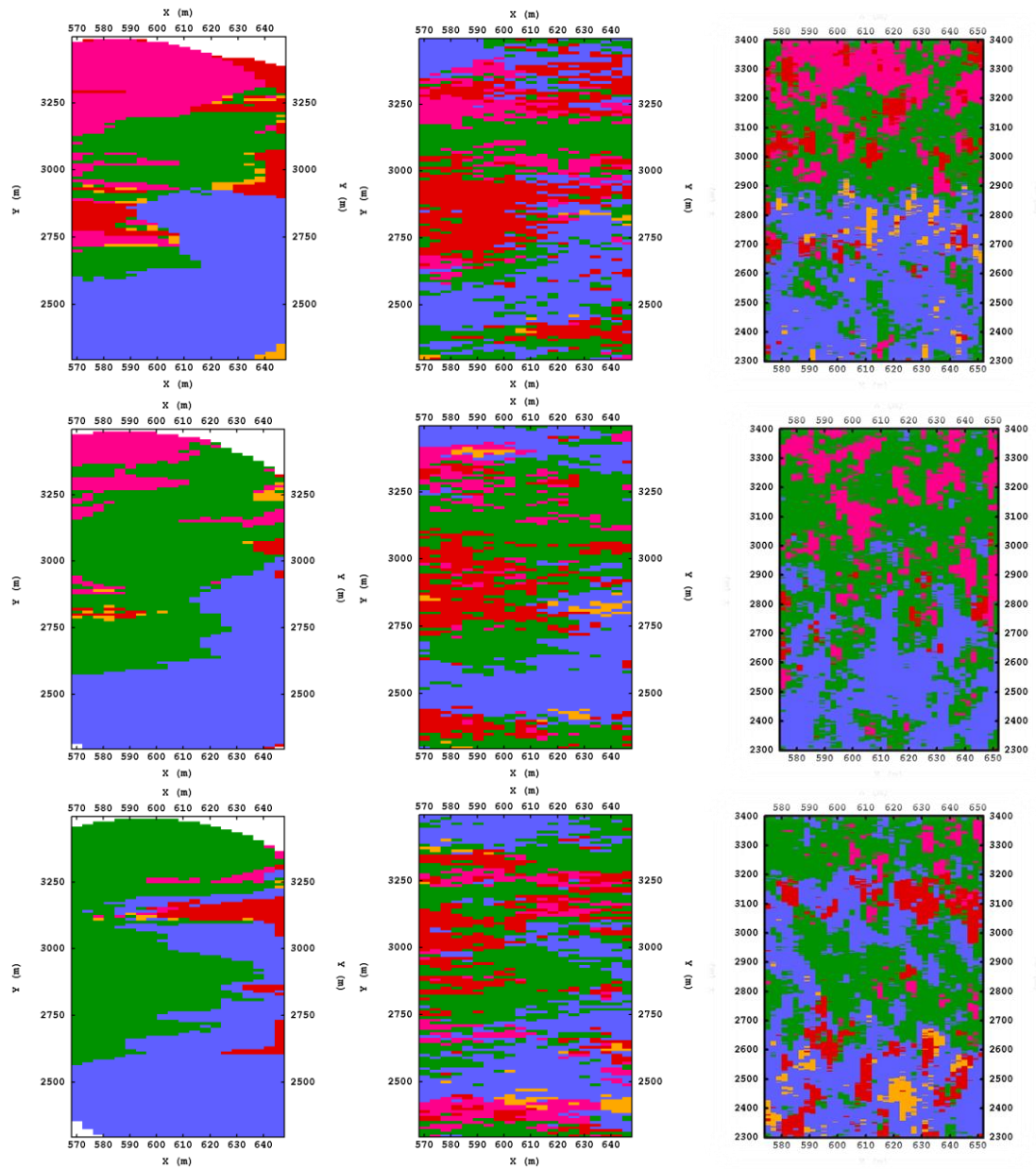
*Figure 30.   IK (left), SIS (middle) and PGS (right) horizontal slices at z = −110m
(top), z = −100m (middle) and z = −90m (bottom).*

67
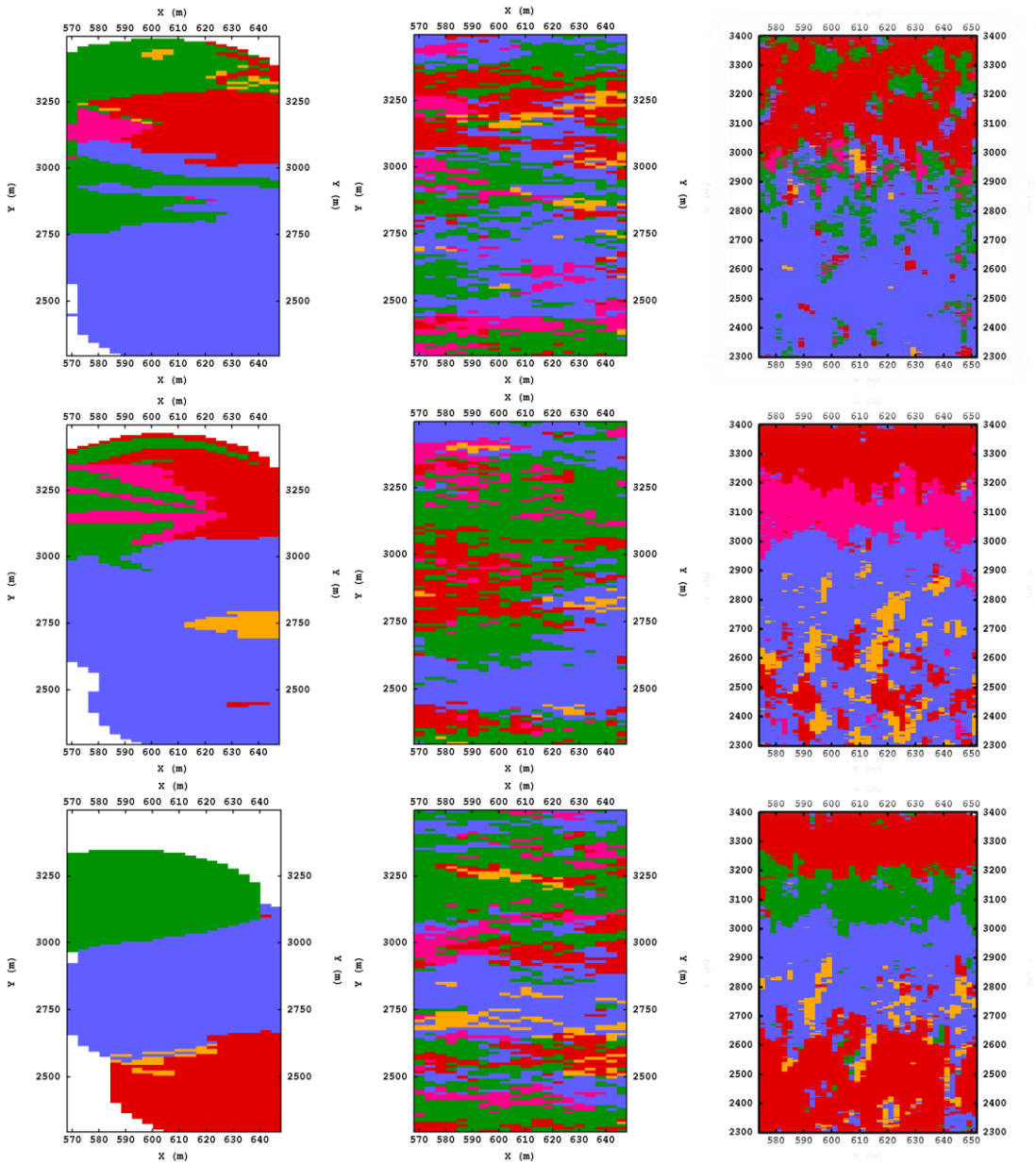
Figure 31. IK (left), SIS (middle) and PGS (right) horizontal slices at $z = -80m$ (top), $z = -70m$ (middle) and $z = -60m$ (bottom).

68

6.2 Global and regional rock type proportions

Table 7 displays global proportions for the drill-hole data set and the three models considered. The IK model has, as expected, reproduced the global lithotype proportions very closely. Both SIS and PGS have much higher proportions of the boundary lithotype L1 and compensatory lower proportions of the other lithotypes. The variation within each of the lithotype proportions for SIS is far greater than for PGS, shown both in summary in Table 7 and over the 100 simulations in Figure 32.

Table 7   Summary lithotype percentage statistics for Data set, IK,
SIS (100 realisations) and PGS (100 realisations).

	Data Set	**IK**	**SIS**				**PGS**			
			min	**mean**	**max**	**st. dev**	**min**	**mean**	**max**	**st. dev**
L1	9.96	9.97	18.19	21.622	28.81	1.73	27.35	28.573	28.59	0.46
L2	2.93	2.4	1.24	2.644	4.45	0.67	1.48	1.875	2.29	0.15
L3	27.99	27.99	22.25	25.348	28.65	1.41	22.78	23.85	24.74	0.43
L4	45.57	46.1	33.05	38.85	44.19	2	33.75	34.77	35.98	0.51
L5	13.54	13.54	8.5	11.536	14.74	1.45	10.28	10.925	11.71	0.33
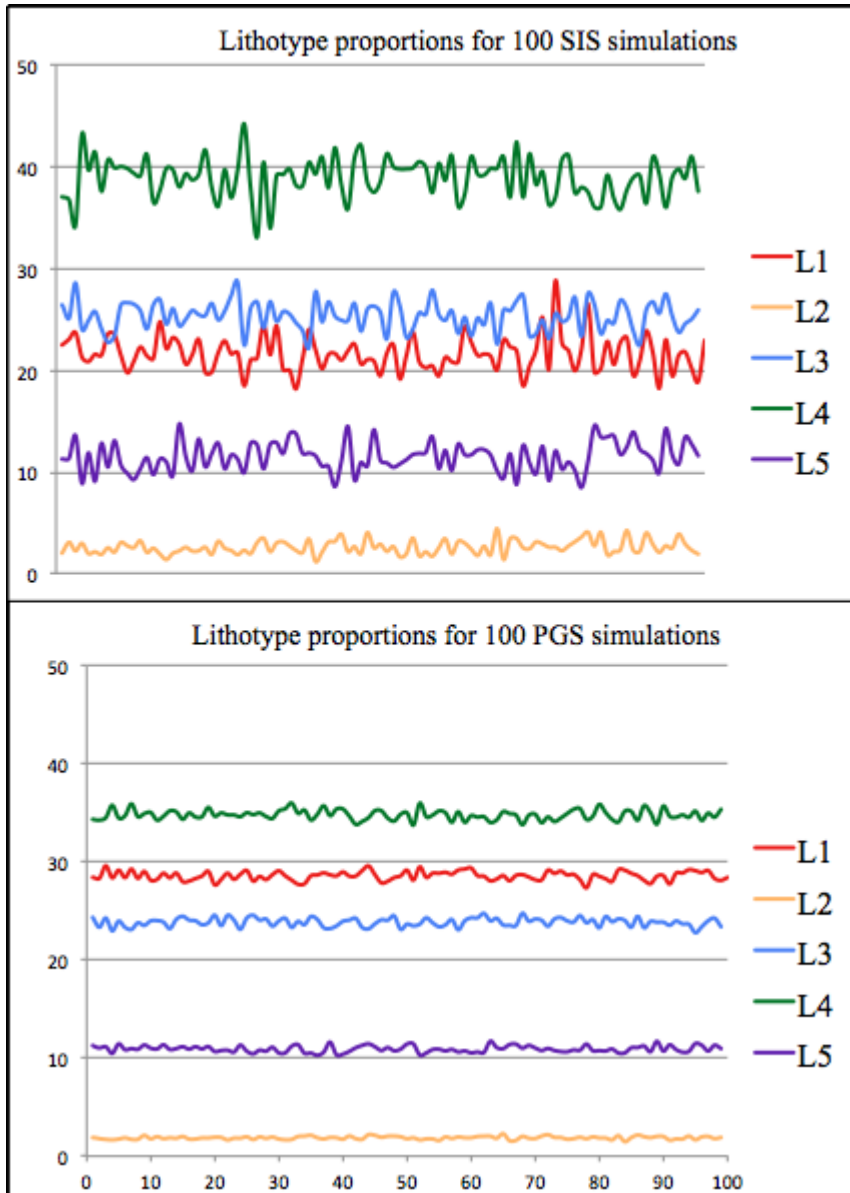
Figure 32   Lithotype proportions for 100 simulations of SIS (top) and PGS (bottom)

The use of the Soares classification algorithm has enforced the proportions from the data set to be reproduced in the IK model (Table 7) . For SIS the relative magnitude of proportions is the same as for IK and the data set. Figure 32 indicates that for SIS (generally) the proportion of L1 (red) is lower than L3 (blue), however for PGS it appears to be different: there appears to be a greater proportion of locations where L1 is allocated rather than L3. The concern that the compensatory balancing required due to the higher proportion of the boundary category L1 has lowered the validity of the PGS output is unfounded. Figure 33 shows the relative proportions of each of L2, L3, L4 and L5 within the SIS and PGS models when removing the L1 proportion from the

calculation. The effect of the variation in the proportions of the boundary lithotype L1 has had little effect with respect to the other lithotypes within each model. Relative proportions are not only very similar in each of the simulations they are also similar to the drill-hole data set and IK proportions, shown in Table 8. Whilst the PGS model has the underlying goal of matching global experimental lithotype proportions, the 'filling in' of the boundary lithotype has had little effect on the consistency of the PGS model in terms of validity of lithotype proportions modelled in the body of the study region.
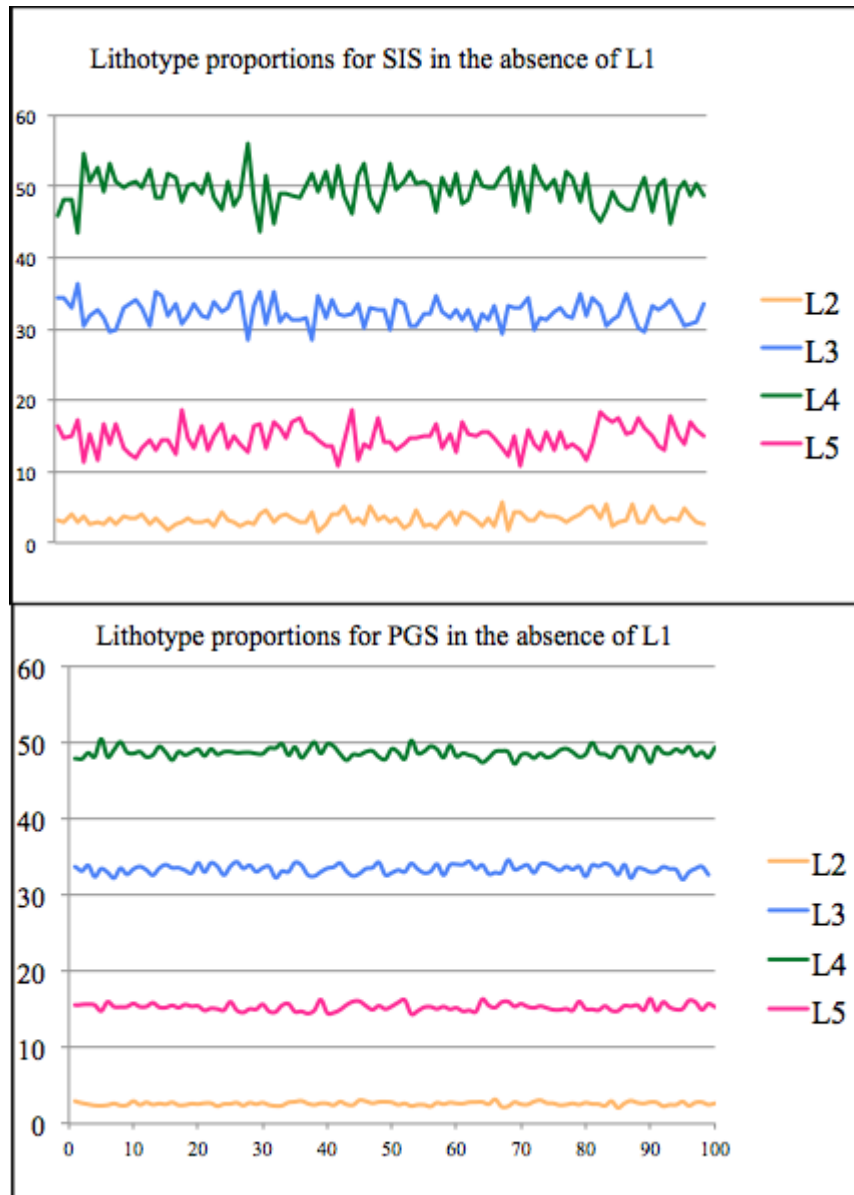
Figure 33　　　*Relative proportions of L2, L3, L4 and L5 with L1 removed for SIS (top) and PGS (bottom).*

In Figure 33 variation in SIS output for the four lithotypes is evident, contrasted against the consistent nature of the lithotype outputs of PGS. Table 8 confirms that SIS has between 3 and 4 times the variation of each lithotype proportion than PGS however mean proportions of L2, L3 and L4 for SIS are closer to the data set proportions than for PGS. In all 100 PGS simulations the proportions of L4 and L2 are below that of the data set whilst the L3 proportion is above the data set.

Table 8   *Summary data of category percentages from SIS (top) and PGS (middle) and IK and Data Set (bottom) in the absence of L1.*

SIS	L2	L3	L4	L5
min	1.55	28.48	43.50	10.81
max	5.78	36.30	55.95	18.71
mean	3.37	32.34	49.57	14.72
st dev	0.86	1.65	2.34	1.80

PGS	L2	L3	L4	L5
min	2.08	32.04	47.22	14.39
max	3.18	34.57	50.44	16.38
mean	2.62	33.39	48.69	15.30
st dev	0.21	0.57	0.64	0.46

	L2	L3	L4	L5
Data Set	3.25	31.09	50.62	15.04
IK	2.67	30.99	51.21	15.04

The 3D PGS excavation box perspective shown in Figure 29 (lower) indicates that the L1 output is concentrated, as expected, at the top and bottom boundaries of the study region. This is confirmed by the histograms of lithotype proportions for PGS, IK and SIS shown in Figures 34 and 35 taken at 10m depth intervals. These histograms are compared to histograms of lithotype proportions of 10m wide sections (in the vertical direction) of the data set (in red), the width of the sections allowing meaningful numbers of data points to be included in each histogram. These data set histograms give a regional lithotype proportion and so allow comparison to the model histograms. The relative proportions of all models correlate with the data set, with the lithotype of greatest frequency matching in each case, with the only exception being the PGS model at $z = -60m$ in Figure 35, where the proportion of L1 is greater than L3, possibly a

result of the completion of the VPCs with L1 in PGS. The histograms also indicate that SIS has less regional variation in terms of the proportions of the five lithotypes. This is of concern when it is clear that such variation does exist in the experimental data (Figure 12).
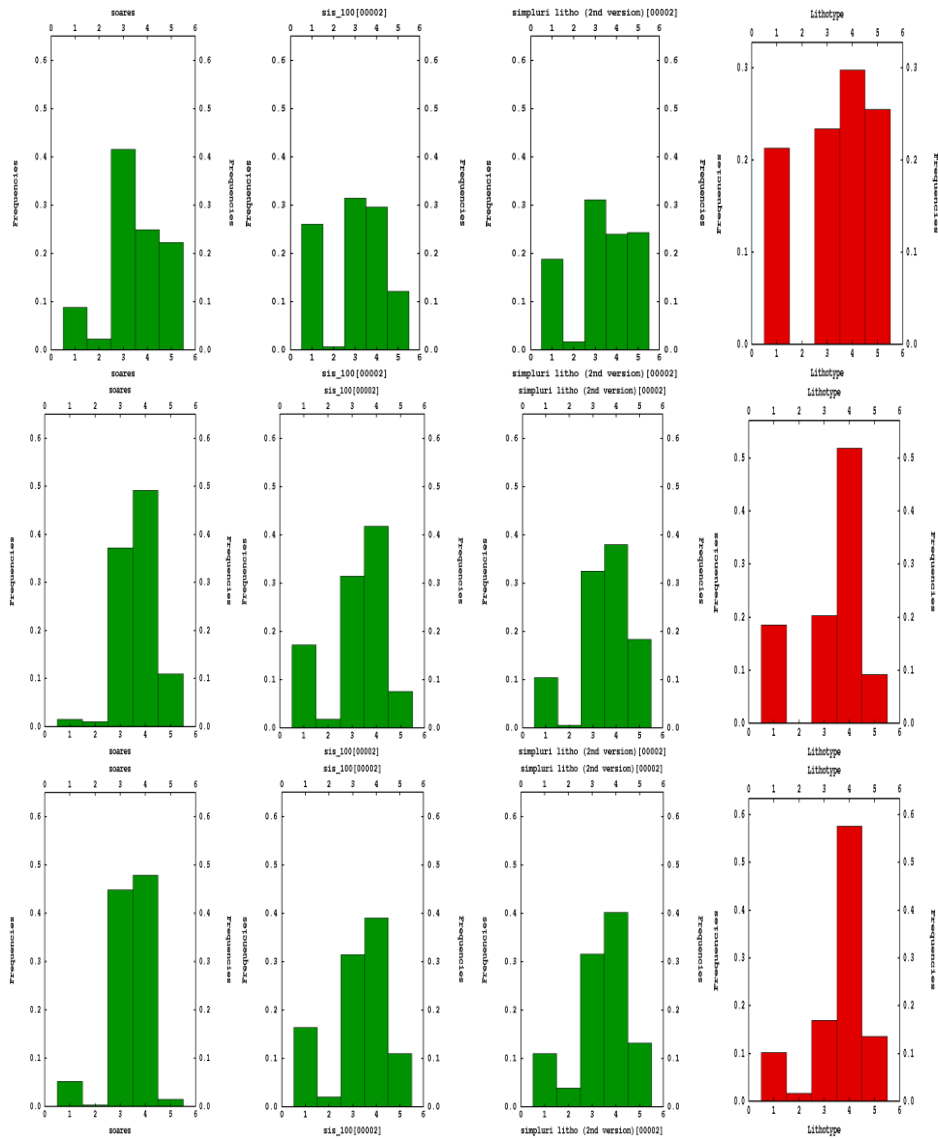
Figure 34   Lithotype proportions at $z = -110m$ (top), $z = -100m$ (middle) and $z = -90m$ (bottom) for (from left to right) IK, SIS, PGS and the data set.
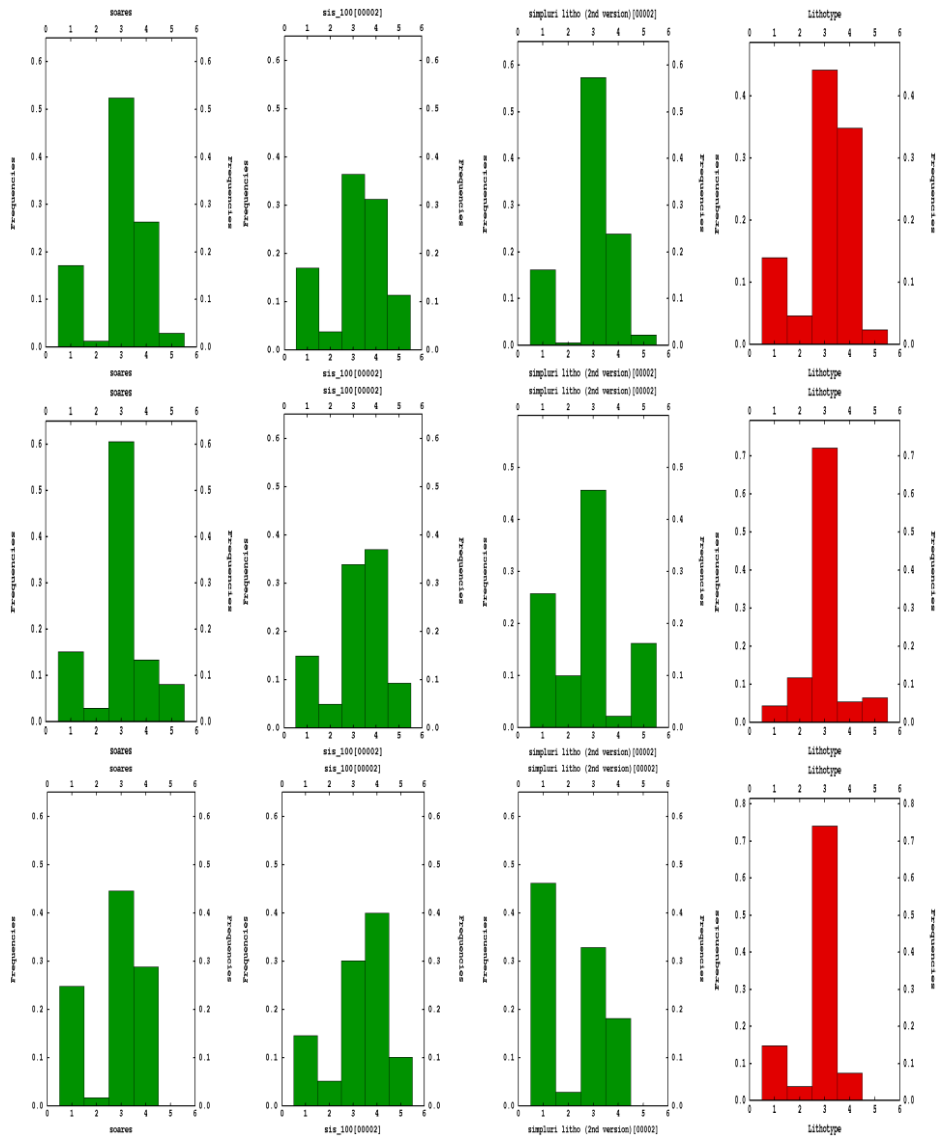
Figure 35   Lithotype proportions at $z = -80m$ (top), $z = -70m$ (middle) and $z = -60m$ (bottom) for (from left to right) IK, SIS, PGS and the data set.

As well as comparison of lithotype proportions over horizontal slices of the study region, it is useful to consider individual grid node lithotype proportions for PGS and SIS over the 100 simulations of each model. For four grid nodes in the study area frequencies of lithotype allocations over the 100 simulations for each model have been calculated and Figure 36 displays the output histograms. PGS usually has a better defined output mode than SIS.

Location (606m,3250m,-160m)   *Location (606m,2750m,-120m)*

Location (622m,2950m, -140m)   *Location (622m,2550m,-180m)*
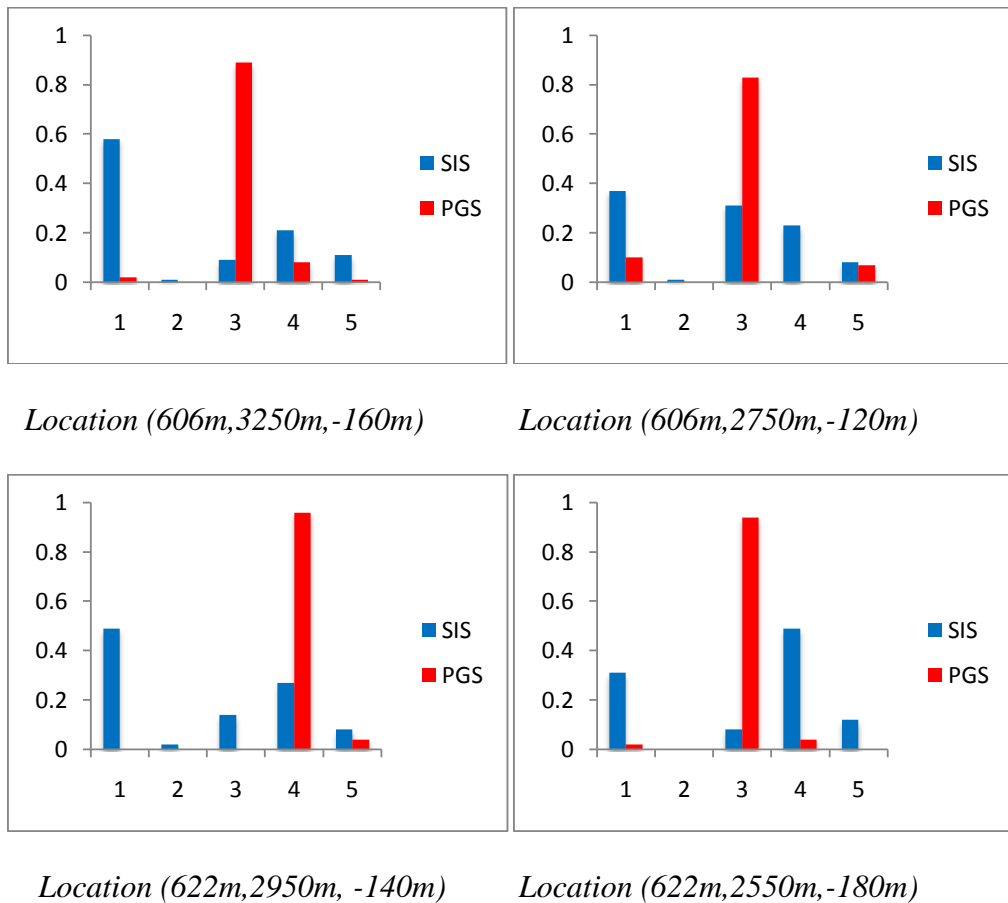
Figure 36 Lithotype proportions at four grid locations in the study region for 100 simulations of PGS (red) and SIS (blue).

6.3 Inadmissible contact numbers for IK, SIS and PGS.

A test of the effectiveness of the PGS output over the IK and SIS output is to count the number of contacts between adjacent cells from each output model that do not match those found in the data set, allowing a quantitative comparison of the three models in terms of their relationship with the reality of lithotype contacts displayed in the reality of the data set. This concept is an adjunct to the concept of connectivity.
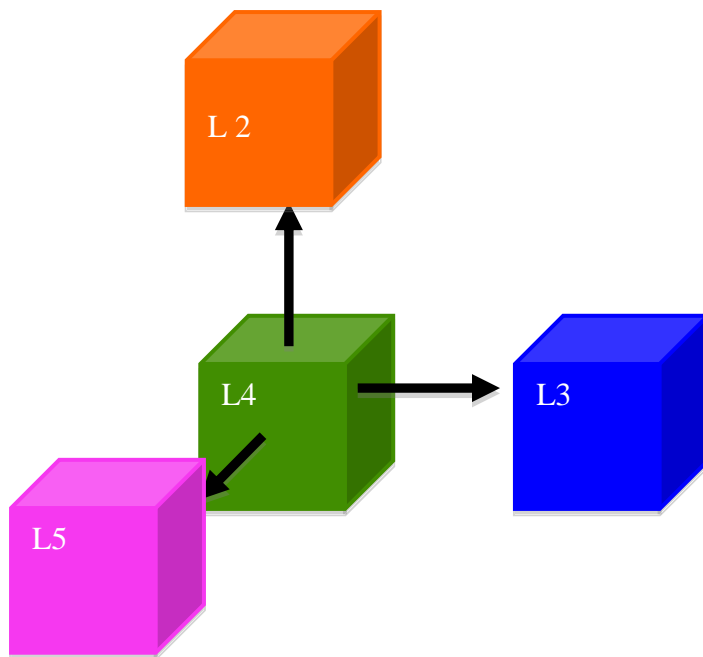
Figure 37. Cells considered for inadmissible contact numbers.

Figure 37 illustrates the process to be considered. For each of the three models the lithotype allocated to a particular cell is compared. A study region grid cell lithotype from all three models is compared to the lithotypes in all cells immediately adjacent. From the rocktype rule, shown in Figure 21, the contact between L4 and L2 in Figure 37 is inadmissible and the inadmissible contact count will be increased by one.

The inadmissible contact process count is carried out by code shown in Appendix 1

The percentages of inadmissible contacts for each simulation are plotted in Figure 38. The realisations of SIS have a greater number of inadmissible contacts (up to 13.624%)

and although PGS has the constraining rocktype rule, the nature of the simulation leads to a small percentage (on average 1.869%) of inadmissible contacts. The IK model (3.364%) has nearly twice the number of inadmissible contacts as PGS on average.
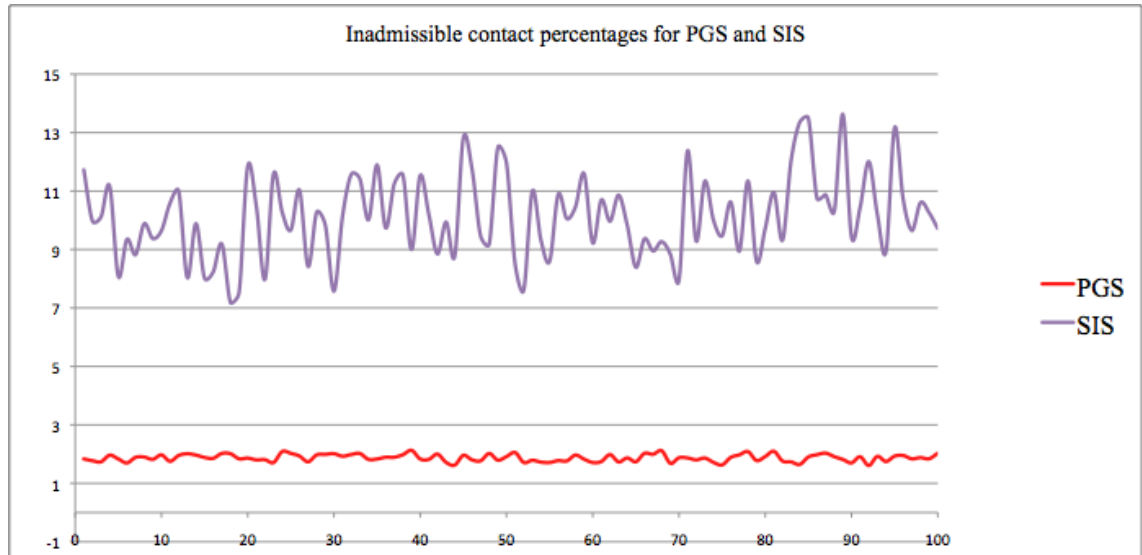
Figure 38   Inadmissible contact percentages for SIS and PGS over 100 simulations.

Table 9   Summary inadmissible contact percentages for IK, SIS and PGS

	Mean	St Dev	Min	Median	Max
IK	3.364	0			
SIS	10.132	1.410	7.183	10.010	13.624
PGS	1.869	0.122	1.610	1.868	2.127

6.4 Transition matrices for IK, SIS and PGS output.

The geostatistical package Isatis provides drill hole transition proportion matrices (downward and upward) for lithotype contacts (Tables 4 and 5). The figures in the matrix are the numerical equivalent of the rocktype rule, where the lithotype occurring in each cell of a discretised core is compared to the adjacent downward or upward cell, and the proportions of transitions between the five lithotypes are calculated. As the downward and upward transition proportion tables are not significantly different, a comparison of transition proportions from cores taken from the three models will be made only with the downward transition proportions from Table 4. Eighteen cores from the output of each model have been taken, with the 100 simulations generated from SIS and PGS all considered. The cores were taken from two central lines in the study region that closely match the drill hole data positions shown in Figure 16. The 18 vertical cores were taken along the two lines $x = 606m$ and $x = 622m$ for y values from $y = 2550m$ to $y = 3350m$ in 100m increments for y, as indicated in Figure 39.
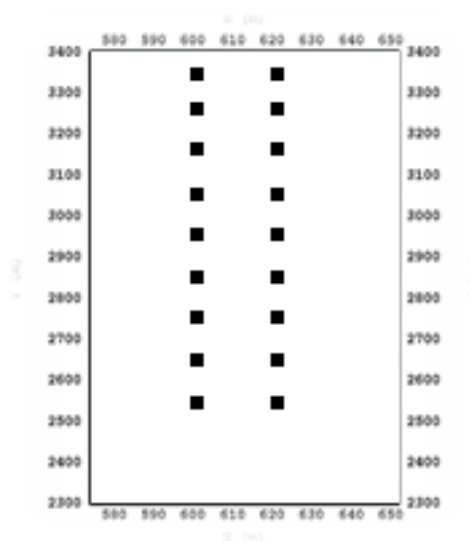
Figure 39   Positions of 18 cores taken from IK, SIS and PGS outputs.

Summary data for each of the 25 possible transitions (L1->L1 to L5->L5) for the 100 simulations of both SIS and PGS have been calculated. Table 10 displays the data downward transition proportions, the IK transition proportions and the average transition proportions for the 100 simulations of both SIS and PGS. Of significance is the mean of the absolute differences for the respective (mean) transitions of each model from the data. PGS differed from the data by an average of 6.9%, compared to IK at

78

10.8% and SIS 11.8%. Each model performed poorly in matching the data L4->L5 transition, and if this figure is removed, PGS performs at a 4.4% average difference for the other 24 comparisons compared to IK at 8.4% and SIS 9.6%. For the inadmissible transitions (l2->L4, L2->L5, L3->L1, L3->L5, L4->L2, L5->L2, L5->L3) PGS averages 1.2% contact, IK 8.1% and SIS 6.0%. These comparisons lend weight to the acceptability of PGS over the other two models.

Table 10 Downward t*ransition matrices for (top to bottom); experimental data, IK, SIS (average of 100 realisations) and PGS (average of 100 realisations).*

Data	L1	L2	L3	L4	L5
L1	0.753	0.025	0.120	0.089	0.013
L2	0.000	0.542	0.458	0.000	0.000
L3	0.000	0.044	0.906	0.050	0.000
L4	0.045	0.000	0.004	0.880	0.736
L5	0.023	0.000	0.000	0.241	0.736

IK	L1	L2	L3	L4	L5
L1	0.841	0.000	0.012	0.059	0.088
L2	0.222	0.185	0.074	0.296	0.222
L3	0.012	0.024	0.925	0.039	0.000
L4	0.023	0.010	0.036	0.888	0.043
L5	0.034	0.028	0.000	0.225	0.713

SIS	L1	L2	L3	L4	L5
L1	0.598	0.009	0.093	0.209	0.090
L2	0.107	0.424	0.286	0.148	0.034
L3	0.099	0.033	0.668	0.159	0.041
L4	0.113	0.009	0.104	0.700	0.074
L5	0.129	0.009	0.078	0.282	0.502

PGS	L1	L2	L3	L4	L5
L1	0.813	0.002	0.027	0.104	0.054
L2	0.055	0.539	0.386	0.019	0.000
L3	0.055	0.036	0.799	0.104	0.006
L4	0.079	0.000	0.084	0.764	0.072
L5	0.063	0.000	0.006	0.312	0.618

The associated boxplots for the transition proportions of the twenty five transitions for SIS and PGS are shown in Figures 40 and 41, together with the experimental drill hole data transitions (red line) and the mean value for the respective model's average transition proportion (blue line). Figure 42 compares the IK transition values with the experimental drill hole data transition values. All models significantly underestimate the L4->L5 transition proportion and both the PGS and SIS models display wide ranges of transition proportions for the L2->L2 and L2->L3 transitions. The IK model displays significant variation from the data transition proportions for L1->L2 through to L2->L5. Of particular interest from the models are two statistics: inadmissible contact transition proportions and internal transitions (downward connectivity of lithotypes).

The inadmissible contacts (0% transition proportion) from the previous section are for the six transitions L2->L4, L2->L5, L3->L5, L4->L2, L5->L2 and L5->L3. Table 10 indicates that the IK model honours two of these (L3->L5 and L5->L3), the PGS model honours three (L4->L2,L5->L2 and L2->L5) while the SIS model has no 0% transitions. For the 6 transitions listed above, PGS performs very well, averaging 1.2% contact, while SIS averages 5.9% and IK 8.1%. Whilst the rocktype rule in PGS should prohibit any inadmissible contacts from occurring, Armstrong et al. (2003) notes that vertical non-stationarity of lithotype proportions, especially in the PGS case, can lead to such anomalies.

For the five within lithotype transitions (L1->L1 for example) the PGS model averages an 8% difference from the downward data transition proportions and the IK model generally performs well, at 9.9% difference, however for L2, the lithotype with the smallest proportion in the study, IK is poorly matched to the data in terms of connectivity. The SIS model matches the downward internal data transition proportions poorly, averaging 18.5% difference.
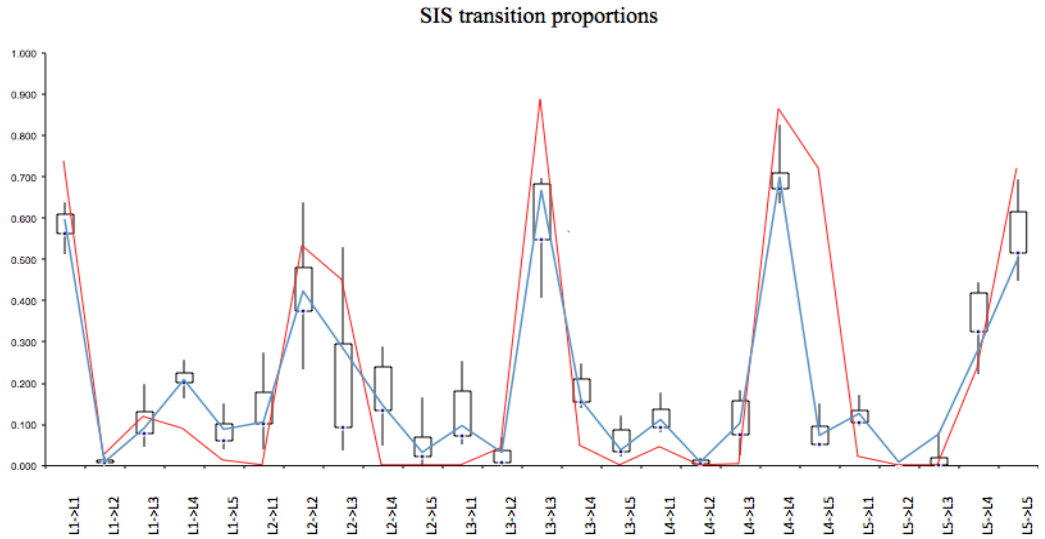
Figure 40   Boxplots of transition proportions for (left to right) L1->L1 to L5->L5 with SIS transition averages (blue) and experimental proportions (red).
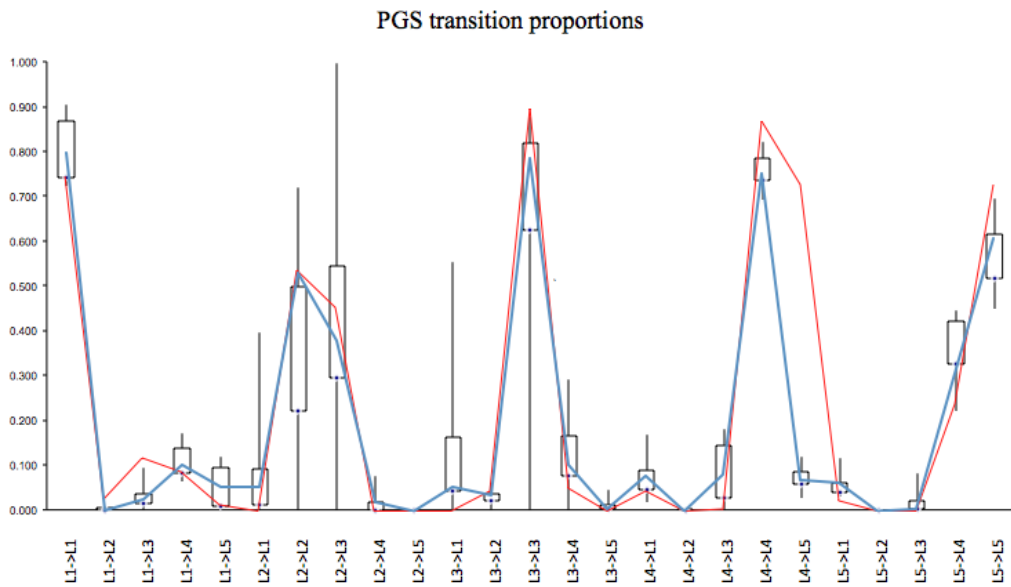
Figure 41   Boxplots of transition proportions for (left to right) L1->L1 to L5->L5 with PGS transition averages (blue) and experimental proportions (red).
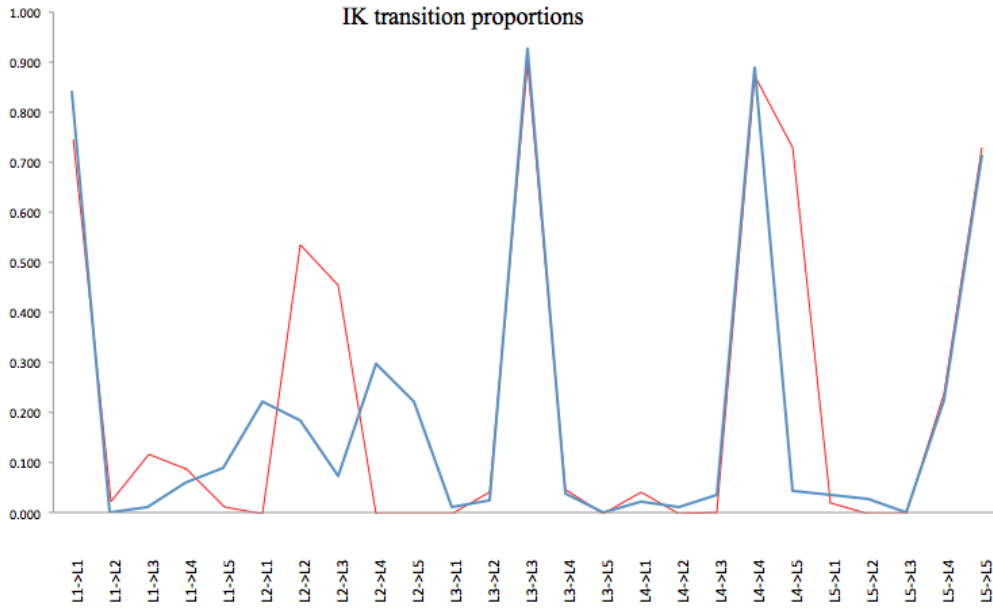
Figure 42   Transition proportions for (left to right) L1->L1 to L5->L5 from IK (blue) and experimental data (red)

6.5   Discussion and conclusion

This thesis sought to explore the PGS algorithm and evaluate its performance against SIS and IK. An exponential increase in computing power over the 25 years since IK was introduced by Journel in 1983 has allowed practitioners access to models of increasing complexity. Do the new models display outputs which have significant advantages over the original, simplistic IK? The evidence in this thesis is that, just as PGS has been developed as computing power has allowed, so the model that this technique produces has measurable advantages over IK and SIS.

As a technique in resource estimation, IK is over 25 years old. The original intention of IK was the estimation of local uncertainty by the process of derivation of a local ccdf and the appeal was that this technique is non-parametric − it does not rely on the assumption of an underlying model for its implementation. Goovaerts (2001) notes that a ccdf obtained by IK provides a measure of uncertainty related to a single location only, and that a series of these single-point ccdfs does not provide a sense of spatial uncertainty. There are a number of downfalls of IK in terms producing a model of categorical data. The first is that the variations in IK estimations are based on the spatial variation of data at sampled locations only, and variations in estimations at unsampled locations are not taken into account. This important factor is remedied in the simulation algorithms.

IK produces output with unrealistic clustering and distribution of lithotypes. This is also noted by Juang et al (2003) with the contention that kriging estimations present only simplistic spatial patterns and do not catch the detail. An underlying issue with IK is that it usually underestimates those categories with very small proportions (in this thesis, L2). Whilst this has been attended to by the implementation of Soares' post-processing algorithm (which ensures that the proportions of the five categories in the output match those of the data set), L2 has been allocated in an unrealistic manner. Figure 22 shows the mid level band of L2 matching its position in the data set shown in Figure 9 however L2 also appears in a good number of small patches in the centre of the study region, positions not justified when considering the data set. It is however fair to say that with respect to all five lithotypes, IK gives a simplistic model of the lithotype distribution, with the overall picture being consistent with the data set and the output of

the other models. Even with the clustering of lithotypes, IK has managed to be regionally sensitive to lithotype proportions. The regional lithotype values (shown in figures 28 and 29) are similar to SIS but have an even greater degree of similarity to PGS.

A drawback of IK, in terms of this data set, is that it is not sensitive to allowable lithotype contacts. Table 9 indicates that 3.4% of the contacts in the IK output model do not conform to the reality of the data set. Whilst this is surprisingly low considering there is no restriction with IK on the types of contacts that can be produced and also the nature of lithotype allocations from the Soares post-processing algorithm, it is nearly twice the percentage of that produced by PGS.

SIS is an improvement over IK due to the variation in estimations at unsampled locations being taken into account, It is a model which offers some improvements on IK yet it lacks the refinement and balance of PGS. In its favour, it is quick in its implementation and, as for IK, non-parametric. Figure 29 shows that SIS displays a greater degree of global spatial variability (less clustering) over IK, however this global variability is at the expense of regional sensitivity of lithotype proportions. Figures 34 and 35 indicate that a SIS realization displays little variation regionally, as the six SIS histograms of lithotype proportion at the slice levels vary least among the three models (for example the proportion values for L1 varying between 0.24 and 0.37 for SIS whereas for IK the variation is between 0.01 and 0.25, and for PGS between 0.1 and 0.46).

SIS has far greater variation in proportions of all lithotypes over the 100 realizations than PGS, both with L1 present (Figure 32) and L1 absent (Figure 33). Interestingly, SIS and PGS have varied most in their proportional allocation of L1 and L3 when L1 is present although this is rectified when L1 is removed from the calculation. The high proportion of inadmissible lithotype contacts for SIS (10.1%) was more than five times that of PGS (1.9%) and also the degree of variation in this proportion (more than ten times greater than PGS) indicated in Table 9 are evidence towards the strength of PGS over SIS in terms of consistency of results.

In all areas of model output PGS stands as the model of choice for this data set. It is initially a time consuming algorithm in terms of variography, definition of polygons and the subsequent completion and smoothing of the VPCs, the rocktype rule and the implementation of the non-conditional and then conditional simulations. Unlike PGS, neither IK nor SIS consider the cross correlations between indicators. All evidence leans towards this effort being of great value. The PGS realizations honour the topological constraint of allowable contacts between lithotypes to a far greater extent than either of the other two models considered, they honour the spatial continuity (indicator variograms) of the data set, and the non-stationarity of category proportions with respect to depth (VPCs). Figures 34 and 35 reflect the sensitivity of the PGS models to regional lithotype proportions whilst figure 36 indicates that the output models of PGS also display a remarkable consistency of lithotype proportions over the 100 simulations.

Table 7 indicates that PGS allocates L1 at 28.57% on average against the data set figure of 9.96% and this is most probably as a result of completion of the study region with the L1 bounding lithotype and there is a subsequent under-allocation of other lithotypes in the study region. When L1 is removed from the calculation (Table 8) PGS does adequately mirror the data set percentages for the other four lithotypes.

IK and SIS offer some advantages: they are easily computed, they reflects the global proportions of the data set (although SIS also has great variation in lithotype proportions) and provide visually appropriate global models, although IK lacks the fine detail and SIS produces too patchy a model, and both have relatively high inadmissible lithotype contact percentages. Evidence is strong that the computation time required for PGS is more than adequately compensated for by a number of spatial advantages. Global and regional lithotype proportions are well matched to the data set, spatial variability is well presented and admissible lithotype contacts are minimized. The PGS simulations also display a consistency of lithotype allocation.

PGS is a model worthy of consideration in this non-sedimentary environment. It outperforms both IK and SIS in many key areas of comparison. Table 8 summarises the advantages and disadvantages of each model.

Table 11   Advantages and disadvantages of the three algorithms

Algorithm	Major Advantages	Disadvantages
IK	• Implementation is inexpensive • Matches initial lithotype proportions (with Soares post-processing)	• Unrealistic clustering of lithotypes • Not sensitive to allowable lithotype contacts
SIS	• Implementation is inexpensive	• Simulations highly variable • Not sensitive to allowable lithotype contacts • Not as regionally sensitive as PGS to lithotype proportion non-stationarity • Simplistic variography
PGS	• Sensitive to allowable lithotype contacts • Consistency of lithotype allocations across 100 simulations • Regionally sensitive to lithotype proportion non-stationarity • Balanced output of model: greater detail than IK but less mosaic than SIS. • Models connectivity well • Models inadmissible transition proportions well	• Implementation is time consuming

References

Armstrong, M., Galli, A., Le Loc'h, G., Geffroy, F., and Eschard, R. (2003) *Plurigaussian Simulations in Geosciences*. Springer.

Aniekwena, A. (2003). *Integrated Characterisation of the thin-bedded 8 reservoir, Green Canyon 18, Gulf of Mexico*.Retrieved 29th April 2006 from http://www.earthdecision.com/news/white papers/spe84051.pdf

Armando, Z., and Bardossy, A. (n.d.*) Introduction to Geostatistics*. Retrieved on 3rd May 2006 from http://www.warem.unistuttgart.de/study/program03/downloads/download702e/geostatistics_print.pdf

Beucher,H. And Renard, D. (2005). *Reservoir characterisation*. Retrieved 28th January 2006 from http://cg.ensmp.fr/bibliotheque/2005/BEUCHER/Rapport/DOC 00318/BEUCHER_Rapport_00318.pdf

Biver, P., Haas, A., and Bacquet, C. (2002*). Efficient Techniques to include uncertainties on global parameters in a geostatistical procedure*. Retrieved 17th April 2006 from http://web.fu-berlin.de/iamg2002/download/Biver Pierre.pdf.

Brett, M., Penny, W., Kiebel, S. (2003). *An introduction to random field theory*. Retrieved on April 16 2006 from http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/pdfs/Ch14.pdf

Carrasco, P., Carrasco, P., Ibarra, F., Le loc'h, G., Rojas, R., and Seguret, S. .(n.d.) *Application of the truncated Gaussian Simulation Method to the MM deposit at Codelco Norte, Chile*. Retrieved on 20th March 2006 from http://www.geovariances.com/IMG/pdf/EAGE_Madrid Leloch.pdf

Chiles, J.P. and Delfiner, P. (1999) *Geostatistics: modelling spatial uncertainty*. Wiley, New York.

Chown, E.H., Hicks, J., Phillips, G.N.& Townend, R. (1984). The disseminated Archean Big Bell gold deposits, Murchison province, Western Australia: an example of premetamorphic hydrothermal alteration. In Proceedings of Gold '82: the Geology, Geochemistry and Genesis of Gold Deposits (R.P Foster, ed.) *Geol. Surv. Zimbabwe, Spec. Publ*. 1, p 305-324.

Cressie, N. A. C. (1990), "The Origins of Kriging," *Mathematical Geology*, v. 22, p. 239-252.

Delbari, M., Loiskandl, W., and Afrasiab, P. (2010) Uncertainty assessment of soil organic carbon content spatial distribution using geostatistical stochastic simulation. *Australian Journal of Soil Research*, 48, 27-35

De Marsily, G., (2005). *Dealing with spatial heterogeneity*. Retrieved 28th April 2006 from http://www.fast.u-psud.fr/~cargese/papers/marsily/marsily1.pdf.

Deraisme, J and Farrow, D. (2004). *Geostatistical simulation techniques applied to kimberlite orebodies and risk assessment of sampling strategies*. Retrieved on 8th February 2006 from http://www.geovariances.com/IMG/pdf/Banff2004.pdf

Deutsch, C.V, (1996) Hierarchical Object-Based Stochastic Modeling of Fluvial Reservoirs. *Mathematical Geology*, v.28, no.7, 857-880

Deutsch, C.V., and Journel, A. G. (1998), *GSLIB - Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 338 pp

Dimitrkopoulos, R. and Dagbert, M. (1992) Sequential modelling of relative indicator variables: dealing with multiple lithology types. Geostatistics Troia '92, Soares, A. Editor. Kluwer 1993.

Dowd, P.A. (1996) Structural controls in the geostatistical simulation of mineral deposits, in Baafi E.Y. and Schofield, N.A. (eds) Geostatistics Wollongong '96 pp647-657.

Dowd, P.A., Pardo-Iguzquiza, and E ,Xu, C. (2003) Plurigau: a computer program for simulating spatial lithotype using the truncated plurigaussian method. *Computers and Geosciences*, 29,123-141.

Galli, A., Beucher, H., and Leloc'h, G. (1994) *The pros and cons of the truncated Gaussian method.* M. Armstrong and P.A. Dowd (Eds), Geostatistical Simulation, Kluwer Academic Publishing: Netherlands, p 217-233.

Girard, A. (2004) *Approximate Methods for propagation of Uncertainty with Gaussian Process Models*. Thesis submitted to University of Glasgow. Retrieved on April 16 2006 from http://www.dcs.gla.ac.uk/~rod/publications/Gir04.pdf

Gomez-Hernandez, J. and Srivastava, R.M., (1990) ISIM3D: an ANSI-C three dimensional multiple indicator conditional simulation program. Computers and Geosciences, v.16, no.4, p395-440

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

Goovaerts, P. (2001). Geostatistical modelling of uncertainty in soil science. *Geoderma* 103 (2001) 3-26.

Gringarten, E. and Deutsch, C.V (2001). Teacher's Aide Variogram Interpretation and Modeling.. *Mathematical Geology*, 33 (4) 507-534.

Handley, G.A and Cary, R. (1990) Big Bell Gold Deposit. *Geology of the mineral deposits of Australia and Papua New Guinea*. 211-218

Isaaks, E.H. (1984) Indicator simulation:application to the simulation of a high grade mineralisation. *Geostatistic for natural resources,* Part2, D. Reidel Publishing Company, 1057-1069.

Isaaks, E. H., and Srivastava, R. M. (1989), *An Introduction to Applied Geostatistics*, Oxford University Press, New York, 561 pp

Journel, A.G. (1974). Geostatistics for conditional simulation of ore bodies. *Economic Geology* 69, p 673-687.

Journel, A.G., and Huijbregts, C. (1978), *Mining Geostatistics*, Academic Press, 600 pp.

Journel, A.G. (1983), Nonparametric estimation of spatial distributions, *Mathematical Geology*, 15(3) p445-468

Journel, A.G. (1989), Fundamentals of Geostatistics in Five Lessons, American Geophysical Union, Washington D.C.

Juang, K.W, Chen, Y,S. Lee, D,Y (2003) Using sequential indicator simulation to assess the uncertainty of delineating heavy-metal contaminated soils. *Environmental Pollution 127* (2004) p 229-238.

Krige, Danie G. (1951). "A statistical approach to some basic mine valuation problems on the Witwatersrand". *J. of the Chem., Metal. and Mining Soc. of South Africa* **52** (6): 119-139

Lantuéjoul C. (1994) *Non Conditional Simulation of Stationary Isotropic Multigaussian Random Functions*. In M. Armstrong & P.A. Dowd eds., Geostatistical Simulations. Dordrecht: Kluwer, 1994. pp.147-167

Lantuéjoul C. (1996) *Iterative Algorithms for Conditional Simulations*. In E.Y Baafi & N.A. Schofield eds., Geostatistics Wollongong, vol.1. Dordrecht: Kluwer,. 1997 pp. 27-40.

Lantuejoul, C. (2002). Conditional *simulation of spatial stochastic models.* Retrieved on 14th March 2006 from http://www.cmis.csiro.au/ismm2002/proceedings/PDF/29 lantuejoul.pdf.

Le Loc'h, and G., Galli, A., (1996). *Truncated plurigaussian method: theoretical and practical points of view.* Proceedings of the Fifth International Geostatistics Congress, Wollongong '96. Australia, pp. 211-222.

Le Ravalec-Dupin, and M., Roggero, F.. (2004) Conditioning truncated Gaussian Realisations to Static and Dynamic Data. *SPE Journal*. 9(4) 475-480

Li, W. and Zhang, C. (2007) Comparing a fixed-path markov chain geostatistical algorithm with sequential indicator simulation in categorical variable simulation from regular samples. *Web of Science*, 44(3) 251-266

Lloyd, C.D. and Atkinson, P.M.(2001). Assessing uncertainty in estimates with ordinary and indicator kriging. *Computers and Geosciences*, 27(8), 929-937.

Matheron, Georges (1962). *Traité de géostatistique appliquée*. Editions Technip

Moysey, S., and R.J. Knight (2003), Full-inverse statistical calibration: A Monte Carlo approach to determining field-scale relationships between hydrologic and geophysical variables, *Eos Trans. AGU*, *84*(46), Fall Meet. Suppl., Abstract H21F-03.

Mueller, A.G and Groves, D.L (1991) *The classification of Western Australian greenstone-hosted gold deposits according to wall-rock alteration mineral assemblages*. Ore Geology Reviews, Volume 6 issue 4, July 1991 pp. 291-331.

Mueller, U.A, Goovaerts, P. and Mueller, A.G. (2002). Geostatistical Modelling of rock type: *A Comparason of the Performance of Classification Schemes Based on Data from the Big Bell Gold Deposit, Western Australia*. Proceedings of IAMG02, Berlin, September 15-29, 2002

Pebesma, E.J., (2004). Multivariate geostatistics in S: the gstat package. *Computers and Geosciences* 30 (2004) 683-691.

Phillips, G.N, Powell, R. (1985) Formation of gold deposits: Review and evaluation of the continuum model. *Earth Science Reviews* , V.94, p 1-21.

Remacre, A.Z, & Zapparolli, L.H. (2003). Application of the plurigaussian simulation techniques in reproducing litholithotype with double anisotropy. *Revista Brasileira de Geociencias*, 33, 37-42.

Seifert, D. And Jensen, J.L., (1999). Using sequential indicator simulation as a tool in reservoir description, issues and uncertainties. *Mathematical Geology*, 31 p527-550.

Soares, A. (1992) Geostatistical Estimation of Multi-Phase Structures. *Mathematical Geology*, V. 24, No 2, p 149-160.

Srivastava, R.M., (1994) Thoughts and comments on conditional simulation algorithms. *Geostatistics*, V.7, no. 2, p9-10.

Walsh, B. (2004) *Markov chain Monte Carlo and Gibbs Sampling*. Retrieved 17th April 2006 from http://nitro.biosci.arizona.edu/courses/EEB581-2004/handouts/Gibbs.pdf

Vann, J., Bertoli, O. and Jackson, S. (2002). *An Overview of Geostastical Simulation for Quantifying Risk.* Paper published at a Geostatistical Association of Australasia symposium "Quantifying Risk and Error'. March 2002. Retrieved on 26th January 2006 from http://www.quantitativegeoscience.com/images/pdf/vann bertoli jackson simulation for risk distribution.pdf

Woodward, H.P. 1914, 'A geological reconnaissance of a portion of the Murchison Goldfield', *Geological Survey, Bulletin No. 57*, Government Printer, Perth.

Xu, C., Dowd, P.A., Mardia, K.V., Fowell, R.J. (2006) A flexible true plurigaussian code for spatial lithotype simulations. *Computers and Geosciences*, v.32 (10) 1629-1645.

Xu, C. And Dowd, P.A. (2008) Plurigaussian simulation of rock fractures. *Proceedings of the VIII International Geostatistics Congress* (pp 41-50). Santiago, Chile.

Appendix

Indicator Kriging Parameter File

```
                Parameters for IK3D
                *******************

START OF PARAMETERS:
0                    \1=continuous(cdf), 0=categorical(pdf)
2                    \option: 0=grid, 1=cross, 2=jackknife
gridbb1.dat              \file with jackknife data
4   5   6   0        \ columns for X,Y,Z,vr
5                    \number thresholds/categories
1  2  3  4  5        \ thresholds / categories
0.11  0.03  0.28  0.45  0.13   \ global cdf / pdf
BB3d.dat                 \file with data
0  4  5  6  7            \ columns for X,Y,Z and variable
direct.ik               \file with soft indicator input
1   2   0    3  4  5  6      \ columns for DH X,Y,Z and indicators
-1.0e21   1.0e21         \trimming limits
2                    \debugging level: 0,1,2,3
ik3d.dbg                \file for debugging output
BBIKok1.out              \file for kriging output
50   570    2.0      \nx,xmn,xsiz
40   2350   25.0        \ny,ymn,ysiz
30  -200.0   5.0        \nz,zmn,zsiz
2   16               \min, max data for kriging
200.0  50.0  20.0        \maximum search radii
 0.0   0.0   0.0         \angles for search ellipsoid
0                    \max per octant (0-> not used)
0   2.5              \0=full IK, 1=Median IK(threshold num)
1                    \0=SK, 1=OK
2   0.25             \One   nst, nugget effect
1   0.25 180.0   0.0   0.0   \    it,cc,ang1,ang2,ang3
        75.0  75.0   7.0   \    a_hmax, a_hmin, a_vert
1   0.50 180.0   0.0   0.0   \    it,cc,ang1,ang2,ang3
        600.0  150.0  45.0   \    a_hmax, a_hmin, a_vert
2   0.40              \Two   nst, nugget effect
1   0.30 180.0  0.0   0.0   \    it,cc,ang1,ang2,ang3
        150.0  75.0   3.0   \    a_hmax, a_hmin, a_vert
1   0.30 180.0  0.0   0.0   \    it,cc,ang1,ang2,ang3
        500.0  100.0  15.0   \    a_hmax, a_hmin, a_vert
2   0.12              \Three nst, nugget effect
1   0.2  180.0   0.0   0.0   \    it,cc,ang1,ang2,ang3
        75.0  75.0   32.0   \    a_hmax, a_hmin, a_vert
1   0.68 180.0  0.0   0.0   \    it,cc,ang1,ang2,ang3
        500.0  400.0  32.0   \    a_hmax, a_hmin, a_vert
2   0.25              \Four  nst, nugget effect
1   0.40  90.0   0.0    0.0 \    it,cc,ang1,ang2,ang3
        50.0   50.0   20.0 \    a_hmax, a_hmin, a_vert
1   0.35  90.0   0.0    0.0 \    it,cc,ang1,ang2,ang3
        600.0  150.0   30.0 \    a_hmax, a_hmin, a_vert
2   0.35              \Five  nst, nugget effect
1   0.50  90.0   0.0    0.0 \    it,cc,ang1,ang2,ang3
        40.0   40.0    8.0 \    a_hmax, a_hmin, a_vert
1   0.15  90.0   0.0    0.0 \    it,cc,ang1,ang2,ang3
        75.0   75.0  1000.0 \    a_hmax, a_hmin, a_vert
```

IK post-processing parameter file

```
c----------------------------------------------------------------------
c              Amilcar Soares classification
c              ***************************
c
c Allocate each grid node to the category with the largest probability of
c occurrence under the constraint of reproduction of marginal proportions
c
c INPUT/OUTPUT Parameters:
c
c  - Name of a data file with probabilities of occurrence of categories
c  - number of categories
c  - marginal proportions to be reproduced
c  - An output file with classification statistics and results
c    (may be overwritten)
c
c
c
c Original: P. GOOVAERTS                    Date: April 1997
c----------------------------------------------------------------------
      program amilcar
      parameter (MAXDAT=170000, MAXVAR=17,MAXPROB=9,MAXITER=40)
      real     prob(MAXDAT,MAXVAR),num(MAXDAT,MAXVAR),
     +         nbre(MAXVAR),sum,
     +         tvec(MAXDAT),tnum(MAXDAT),count(MAXDAT),
     +         emploi1, emploi2, emploi3, emploi4, emploi5,
     +         margprob(MAXVAR),npcl(MAXVAR),iter(MAXVAR),
     +         test(MAXVAR)
      integer  amic(MAXDAT)
      character outfl1*40,str*40,datafl*40, outfl2*40
      logical  testfl
      data     lin/1/,ncut/0/, lout/2/


      write(*,*) 'Which parameter file do you want to use?'
      read (*,'(a40)') str
      if(str(1:1).eq.' ')str='amilcar.par                    '
      inquire(file=str,exist=testfl)
      if(.not.testfl) then
         write(*,*) 'ERROR - the parameter file does not exist,'
         write(*,*) '        check for the file and try again  '
         stop
      endif

      open(lin,file=str,status='OLD')

c Find Start of Parameters:
c
 201  read(lin,'(a4)',end=198) str(1:4)
      if(str(1:4).ne.'STAR') go to 201
c
c Read Input Parameters:
c
      read(lin,'(a40)') datafl
      write(*,*) ' data file:       ',datafl
      read(lin,'(a40)') outfl1
      write(*,*) ' output file:       ',outfl1
      read(lin,'(a40)') outfl2
      write(*,*) ' output file:       ',outfl2
      read(lin,*)    nv
      write(*,*) ' number of category:    ',  nv
      read(lin,*) (margprob(j), j=1,nv)
      write(*,*) 'p_k :',(margprob(j), j=1,nv)
      close(lin)

c ****************************************************************
c
c Check to make sure the data file exists, then either read in the
c data or write an error message and stop:
```

```
c
      inquire(file=datafl,exist=testfl)
      if(.not.testfl) then
        write(*,*) 'ERROR data file ',datafl,' does not exist!'
        stop
      endif
c
c Open the file with updated probabilities
c
      open(lin,file=datafl,status='OLD')
      do 3 i=1,(nv+2)
 3    read(lin,'(a40)',err=193) str

c
c Read the updated probabilities
c

      nd = 1.0
 4    read(lin,*,end=5,err=193) (prob(nd,j),j=1,nv)
      nd = nd + 1.0
      if(nd.gt.MAXDAT) then
          write(*,*) ' ERROR: Exceeded available memory for data'
          stop
      end if
      go to 4
 5    close(lin)
      nd = nd -1
      write(*,*) 'number of data readed', nd

c
c Correct for order relation deviations
c
      do i=1, nd
        sum=0.0
        do j=1, nv
          if(prob(i,j).gt.1) prob(i,j)=1.0
          if(prob(i,j).lt.0) prob(i,j)=0.0
          sum = sum + prob(i,j)
        end do
        do j=1, nv
          prob(i,j) = prob(i,j)/sum
        end do
      end do

C *************************************************************
C
C     AMILCAR SOARES CLASSIFICATION
C


C
C     Sort the vectors of probabilities and index vector accordingly
c

      do j=1, nv
        do i=1, nd
          tvec(i) = -prob(i,j)
          tnum(i) = i
        end do
      call sortem(1,nd,tvec,1,tnum,c,d,e,f,g,h)
        do i=1, nd
          num(i,j)=tnum(i)
        end do
      end do

C
C     Check the proportion of double, triple...
C     allocations of the same grid nodes
C

      do i=1, nd
        count(i) = 0
        amic(i)  = -1
      end do
      do j=1, nv
```

94

```
       nbre(j)=nd*margprob(j)
       do i=1,nbre(j)
         k=num(i,j)
         count(k)=count(k)+1.0
       end do
     end do

     emploi1=0.0
     emploi2=0.0
     emploi3=0.0
     emploi4=0.0
     emploi5=0.0
   do i=1,nd
     if(count(i).eq.0) emploi1=emploi1+1
     if(count(i).eq.2) emploi2=emploi2+1
     if(count(i).eq.3) emploi3=emploi3+1
     if(count(i).eq.4) emploi4=emploi4+1
     if(count(i).eq.5) emploi5=emploi5+1
   end do
     emploi1=emploi1*100/nd
     emploi2=emploi2*100/nd
     emploi3=emploi3*100/nd
     emploi4=emploi4*100/nd
     emploi5=emploi5*100/nd

C
C    Perform the classification
C

   do j=1, nv
      npcl(j) = 0.0
      iter(j) = 0.0
      test(j) = 0.0
   end do
      nshift=0.0
      nshift1=0.0
   do 35 i=1,nd
      do 36 j=1,nv
         if(npcl(j).lt.nbre(j)) then
            iter(j)= iter(j)+1
            is=iter(j)
            k=num(is,j)
            if (amic(k).eq.-1) then
              amic(k)=j
              npcl(j)= npcl(j)+1
              go to 36
            endif
            if (amic(k).gt.-1) then
              nshift= nshift + 1.0
              jloc= amic(k)
              if (prob(k,jloc).ge.prob(k,j)) go to 36
              if (prob(k,jloc).lt.prob(k,j)) then
                 nshift1= nshift1 + 1.0
                 amic(k)=j
                 npcl(j)= npcl(j)+1
                 npcl(jloc)= npcl(jloc)-1
              endif
            endif
         endif
36    continue
35  continue


c
c    Check the reproduction of target proportions
C

c    do i=1, nd
c      j = amic(i)
c      test(j) = test(j) + 1
c    end do
c    do j=1, nv
c      test(j) = test(j)/nd
c    end do
c    write(*,*) 'p*_k :', (test(j), j=1,nv)
```

```fortran
      open(lout,file=outfl1,status='UNKNOWN')
C
C     Write results to outputfile
C

      write(lout,*) 'Results for :', outfl
      write(lout,*) '************************'
      write(lout,*) 'Statistics about the classification '
      write(lout,*) 'nber of samples considered :', nshift
      write(lout,*) 'nber of samples actually shifted :', nshift1
      write(lout,*) 'percentage of no classif:', emploi1
      write(lout,*) 'percentage of double classif:', emploi2
      write(lout,*) 'percentage of triple classif:', emploi3
      write(lout,*) 'percentage of quadr. classif:', emploi4
      write(lout,*) 'percentage of quint.. classif:', emploi5
      write(lout,*) ' '
        close(lout)
      open(lout,file=outfl2,status='UNKNOWN')
      write(lout,*) 'Results for :', outfl
      write(lout,*) '1'
        write(lout,*) 'Unit'
      do i=1, nd
        write(lout,*) amic(i)
      end do
      stop
 193  stop 'ERROR in data file!'
 198  stop 'ERROR in parameter file!'
      end

      subroutine sortem(ib,ie,a,iperm,b,c,d,e,f,g,h)
c-----------------------------------------------------------------------
c
c                Quickersort Subroutine
c                **********************
c
c This is a subroutine for sorting a real array in ascending order. This
c is a Fortran translation of algorithm 271, quickersort, by R.S. Scowen
c in collected algorithms of the ACM.
c
c The method used is that of continually splitting the array into parts
c such that all elements of one part are less than all elements of the
c other, with a third part in the middle consisting of one element.  An
c element with value t is chosen arbitrarily (here we choose the middle
c element). i and j give the lower and upper limits of the segment being
c split.  After the split a value q will have been found such that
c a(q)=t and a(l)<=t<=a(m) for all i<=l<q<m<=j.  The program then
c performs operations on the two segments (i,q-1) and (q+1,j) as follows
c The smaller segment is split and the position of the larger segment is
c stored in the lt and ut arrays.  If the segment to be split contains
c two or fewer elements, it is sorted and another segment is obtained
c from the lt and ut arrays.  When no more segments remain, the array
c is completely sorted.
c
c
c INPUT PARAMETERS:
c
c   ib,ie      start and end index of the array to be sorteda
c   a          array, a portion of which has to be sorted.
c   iperm      0 no other array is permuted.
c              1 array b is permuted according to array a
c              2 arrays b,c are permuted.
c              3 arrays b,c,d are permuted.
c              4 arrays b,c,d,e are permuted.
c              5 arrays b,c,d,e,f are permuted.
c              6 arrays b,c,d,e,f,g are permuted.
c              7 arrays b,c,d,e,f,g,h are permuted.
c             >7 no other array is permuted.
c
c   b,c,d,e,f,g,h  arrays to be permuted according to array a.
c
c OUTPUT PARAMETERS:
c
c   a    = the array, a portion of which has been sorted.
c
```
96

```
c   b,c,d,e,f,g,h  =arrays permuted according to array a (see iperm)
c
c NO EXTERNAL ROUTINES REQUIRED:
c
c----------------------------------------------------------------------
      dimension a(*),b(*),c(*),d(*),e(*),f(*),g(*),h(*)
c
c The dimensions for lt and ut have to be at least log (base 2) n
c
      integer   lt(64),ut(64),i,j,k,m,p,q
c
c Initialize:
c
      j     = ie
      m     = 1
      i     = ib
      iring = iperm+1
      if (iperm.gt.7) iring=1
c
c If this segment has more than two elements  we split it
c
 10   if (j-i-1) 100,90,15
c
c p is the position of an arbitrary element in the segment we choose the
c middle element. Under certain circumstances it may be advantageous
c to choose p at random.
c
 15   p   = (j+i)/2
      ta  = a(p)
      a(p) = a(i)
      go to (21,19,18,17,16,161,162,163),iring
 163    th  = h(p)
        h(p) = h(i)
 162    tg  = g(p)
        g(p) = g(i)
 161    tf  = f(p)
        f(p) = f(i)
 16     te  = e(p)
        e(p) = e(i)
 17     td  = d(p)
        d(p) = d(i)
 18     tc  = c(p)
        c(p) = c(i)
 19     tb  = b(p)
        b(p) = b(i)
 21   continue
c
c Start at the beginning of the segment, search for k such that a(k)>t
c
      q = j
      k = i
 20   k = k+1
      if(k.gt.q)    go to 60
      if(a(k).le.ta) go to 20
c
c Such an element has now been found now search for a q such that a(q)<t
c starting at the end of the segment.
c
 30   continue
      if(a(q).lt.ta) go to 40
      q = q-1
      if(q.gt.k)    go to 30
      go to 50
c
c a(q) has now been found. we interchange a(q) and a(k)
c
 40   xa  = a(k)
      a(k) = a(q)
      a(q) = xa
      go to (45,44,43,42,41,411,412,413),iring
 413    xh  = h(k)
        h(k) = h(q)
        h(q) = xh
 412    xg  = g(k)
        g(k) = g(q)
```

97

```
      g(q) = xg
411   xf   = f(k)
      f(k) = f(q)
      f(q) = xf
41    xe   = e(k)
      e(k) = e(q)
      e(q) = xe
42    xd   = d(k)
      d(k) = d(q)
      d(q) = xd
43    xc   = c(k)
      c(k) = c(q)
      c(q) = xc
44    xb   = b(k)
      b(k) = b(q)
      b(q) = xb
45  continue
c
c Update q and search for another pair to interchange:
c
      q = q-1
      go to 20
50  q = k-1
60  continue
c
c The upwards search has now met the downwards search:
c
      a(i)=a(q)
      a(q)=ta
      go to (65,64,63,62,61,611,612,613),iring
613    h(i) = h(q)
      h(q) = th
612    g(i) = g(q)
      g(q) = tg
611    f(i) = f(q)
      f(q) = tf
61     e(i) = e(q)
      e(q) = te
62     d(i) = d(q)
      d(q) = td
63     c(i) = c(q)
      c(q) = tc
64     b(i) = b(q)
      b(q) = tb
65  continue
c
c The segment is now divided in three parts: (i,q-1),(q),(q+1,j)
c store the position of the largest segment in lt and ut
c
      if (2*q.le.i+j) go to 70
      lt(m) = i
      ut(m) = q-1
      i = q+1
      go to 80
70  lt(m) = q+1
      ut(m) = j
      j = q-1
c
c Update m and split the new smaller segment
c
80  m = m+1
      go to 10
c
c We arrive here if the segment has  two elements we test to see if
c the segment is properly ordered if not, we perform an interchange
c
90  continue
      if (a(i).le.a(j)) go to 100
      xa=a(i)
      a(i)=a(j)
      a(j)=xa
      go to (95,94,93,92,91,911,912,913),iring
913    xh   = h(i)
      h(i) = h(j)
      h(j) = xh
```

```
912   xg  = g(i)
      g(i) = g(j)
      g(j) = xg
911   xf  = f(i)
      f(i) = f(j)
      f(j) = xf
 91   xe  = e(i)
      e(i) = e(j)
      e(j) = xe
 92   xd  = d(i)
      d(i) = d(j)
      d(j) = xd
 93   xc  = c(i)
      c(i) = c(j)
      c(j) = xc
 94   xb  = b(i)
      b(i) = b(j)
      b(j) = xb
 95 continue
c
c If lt and ut contain more segments to be sorted repeat process:
c
100  m = m-1
     if (m.le.0) go to 110
     i = lt(m)
     j = ut(m)
     go to 10
110  continue
     return
     end


C*************************************************************************
C
C    THIS SUBROUTINE EVALUATES THE CPU TIME IN SECONDS USED SO FAR
C
C*************************************************************************
C
     SUBROUTINE GETCPU(CPUSEC)
C
C    OUTPUT:
C        CPUSEC    CPU TIME USED SO FAR.
C
C    THIS SUBROUTINE WORKS ONLY ON ABQAIQ AND PANGEA.
C
     REAL*4 TARRAY(2),ETIME
     REAL*8 CPUSEC
C
     CPUSEC = ETIME(TARRAY)
C
     RETURN
     END
```

The following code performs the inadmissible contact count in four steps:

1) Determine the size and dimensions of the matrix.

2) Perform misclassification procedure in the *x* direction:

```
# Do X
if (xsize > 0)
        b = 0.*a;                    Create a zero matrix b of size equal to a.
```

b(:,2:(xsize+1),:) = **a**;	Import values from **a** into the 2nd to n+1th row of **b.**
c = **a**;	
c(:,xsize+1,:) = **a**(:,xsize,:);	Create a matrix **c** of size equal to **b** with entries from **a** with the final row of **a** repeated
d = (**c-b**)(:,2:xsize+1,:);	Perform the subtraction
e = (**a** != 1).*∗**d**;	Product of **d** and **a** where unit entries of **a** are set to 0
xres = (abs((**c**(:,2:xsize+1,:) != 1).*∗**e**) > 1);	Set values in **e** to 1 if greater than 1, else set to 0.
xc = sum(xres);	Sum (unit) entries in matrix

b(:,2:(xsize+1),:) = a; — Import values from **a** into the 2nd to n+1th row of **b.**

```
endif
```

3) Similarly for *y* and *z* directions:

```
# Do Y
if (ysize > 0)
     b = 0.*a;
     b(2:(ysize+1),:,:) = a;
     c = a;
     c(ysize+1,:,:) = a(ysize,:,:);
     d = (c-b)(2:ysize+1,:,:);
     e = (a != 1).*d;
     yres = (abs((c(2:ysize+1,:,:) != 1).*e) > 1);
     yc = sum(yres);
endif


# Do Z
if (zsize > 0)
     b = 0.*a;
     b(:,:,2:(zsize+1)) = a;
     c = a;
     c(:,:,zsize+1) = a(:,:,zsize);
     d = (c-b)(:,:,2:zsize+1);
     e = (a != 1).*d;
     zres = (abs((c(:,:,2:zsize+1) != 1).*e) > 1);
     zc = sum(zres);
endif
```

4) Summing the (three) directional misclassifications

```
if (zsize > 0)
     ret = sum(sum(xc + yc + zc));
else
     ret = sum(xc + yc);
endif

endfunction
```