Advances in Geostatistical Modeling of Categorical Variables

by

Rafael Barros Ortiz

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

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

In the mining and petroleum industries, subsurface resources are modeled using data sets obtained from widely spaced drilling. It is crucial to optimize the use of available information from these data sets while reducing the amount required to adequately assess risks and uncertainties. Typically, these data sets include spatially correlated categorical and continuous variables. Geostatistics is commonly applied to model these types of spatial variables.

Categorical variables are modeled first to establish stationary domains for continuous variables like ore grades and, therefore, are essential for the accurate modeling of continuous variables. Techniques such as object-based models (Lantuéjoul, 2002), sequential indicator simulation (A. G. Journel & Alabert, 1990), multiple-point statistics (S. Strebelle, 2002), truncated Gaussian and pluri-Gaussian simulation (Armstrong et al., 2011; Matheron et al., 1987), and hierarchical truncated pluri-Gaussian simulation (D. Silva, 2018) have been developed to model and characterize geological uncertainties. Modern approaches in multivariate modeling of continuous variables include principal component analysis (PCA) for dimension reduction and decorrelation. Suro (1988) applied PCA to decorrelate indicators and used kriging, followed by a back transformation, to derive the conditional cumulative distribution function.

This research aims to obtain a greater understanding of the characteristics of categorical indicator variables. Studying the characteristics of categorical indicator variograms led to the demonstration that the nugget effect must be zero. Additionally, correlograms are shown to be a robust alternative in the presence of clustered data. This work also compares multiple-point statistics-based (MPS) conditional probabilities to variogram-based simple kriging (SK), ordinary kriging (OK), simple cokriging (SCK), and standardized ordinary cokriging (SOCK) estimates to evaluate which kriging variant comes closest to the benchmark MPS probabilities. SOCK stands out among the kriging variants when compared to the MPS probabilities. Subsequently, SOCK and OK estimates are compared to reference (training images) for several different cases. The difference in the root mean squared error (RMSE) gives a very small advantage to the SOCK method.

The research work of this thesis also led to the implementation of a method to mitigate extreme weights that result from indefinite or ill-conditioned matrices in linear systems of equations. The method consists of inflating the diagonal of the left-hand side covariance matrix by a small constant. As the constant increases, all weights converge to $\frac{1}{n}$, where n is the number of weights. This technique shows that the extreme weights are mitigated but does not significantly alter the overall estimation across an entire grid.

Lastly, this thesis applies principal component analysis to decorrelate categorical indicator variables and estimate each of the decorrelated components independently. The technique reduces the dimension by $K - 1$, where K is the number of categories, because the last variable is explained through a linear combination of the other variables. However, the estimation of the independent components and their subsequent transformation back to the original values proves to be less effective than ordinary kriging. This could be explained by the mixing of spatial structure when transforming to principal components. The ranges of the principal components' variograms account for the ranges of all the categories' indicator variograms.

Dedication

To my family: Kadija and our little Anna. To my parents: Rafael and Magdalena.

Acknowledgments

I would like to thank my supervisor, Dr. Clayton Deutsch, for his invaluable advice, guidance, and endless ideas. His enthusiasm and motivation throughout my research have been inspiring. His kindness has made this learning process enjoyable, and he serves as a model of professionalism for me.

I would like to acknowledge the Centre for Computational Geostatistics member companies for their financial support. I would also like to thank alumni Diogo Silva and Felipe Pinto for their support and discussions along the way and for their friendship. I want to thank all my colleagues of the Centre for Computational Geostatistics, for your friendship and for sharing your knowledge.

Finally, this would not be possible without the support and love of my family. To Rafael, Maria, Tatiana and Guillermo, and most of all to Kadija and Anna, thank you.

Table of Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description
B	Local accuracy parameter
R^2	Coefficient of determination
ρ	Correlation
$\gamma_{corr}(\mathbf{h})$	Flipped correlogram for lag $\mathbf{h}$
$C_k(\mathbf{h})$	Covariance of category k for lag $\mathbf{h}$
$C(0)$	Covariance for distance 0
Σ	Covariance matrix
$C_{kk'}(\mathbf{h})$	Cross covariance between category k and k' for lag $\mathbf{h}$
$\gamma_{kk'}(\mathbf{h})$	Cross variogram between category k and k' for lag $\mathbf{h}$
$E\{\quad\}$	Expected value operator
$\mathbf{D}$	Diagonal eigenvalue matrix
$\mathbf{V}$	Eigenvector matrix
ϵ	Regularization constant
$i^*(\mathbf{u}; k)$	Estimate at location $\mathbf{u}$ of category k
i_k	Indicator variable
K	Number of categories
μ	Lagrange multiplier
$\mathbf{h}$	Lag distance
p_k	Probability of category k
σ^2	Variance
σ	Standard deviation
Y_k	Standardized indicator variable for category k
$\mathbf{u}$	Location in a domain
$\text{Var}\{\quad\}$	Variance operator
$\gamma_k(\mathbf{h})$	Variogram of category k for lag $\mathbf{h}$
λ	Weight

List of Abbreviations

Abbreviation	Description
2-D	two-dimensional
3-D	three-dimensional
BLUE	best linear unbiased estimator
CCG	Center for Computational Geostatistics
GANs	generative adversarial networks
GRF	Gaussian random function
GSLIB	geostatistical software library
HTPG	hierarchical truncated pluri-Gaussian
IK	indicator kriging
LMC	linear model of corregionalization
MAF	minimum/maximum auto-correlation factors
MAPS	maximum a posteriori selection
MIK	multiple indicator kriging
ML	machine learning
MPS	multiple point statistics
MSE	mean squared error
NE	nugget effect
OBM	object-based modeling
OK	ordinary kriging
PCA	principal component analysis
PPMT	project pursuit multivariate transformation
RF	random function
RMSE	root mean squared error
RV	random variable
SCK	simple cokriging
SGS	sequential Gaussian simulation
SIS	sequential indicator simulation
SK	simple kriging
SOCK	standardized ordinary cokriging
TGS	truncated Gaussian simulation

Abbreviation	**Description**
TI	training image
TPGS	truncated pluri-Gaussian simulation
VL-LMC	very large linear model of corregionalization

Chapter 1

Introduction

This chapter presents an overview of the problem that drives the research topic of this thesis, accompanied by a literature review. Additionally, the thesis statement and the contributions, and the outline of the thesis are presented. Section 1.1 offers an overview of geostatistical methods to model categorical and continuous variables. A literature review is presented in Section 1.2. Section 1.3 introduces the thesis statement and contributions. Section 1.4 presents an outline of the thesis, providing a brief summary of each chapter.

1.1 Problem motivation

Since its conception by Matheron (1962), geostatistics aims to model spatially correlated variables. The cost associated in acquiring samples, both in mining and petroleum industry, leads to relatively few samples. According to Rossi and Deutsch (2013), less than one billionth of the explored mass is sampled before mining. In the petroleum industry, there are even fewer samples according to Chilès and Delfiner (2012). Therefore, taking maximum advantage of the information contained in the sampled data and, perhaps, reducing the number of data to achieve a good understanding of risk and uncertainty is crucial (D. Silva, 2018). It is essential to understand the behavior of the samples to get a good prediction of the rock properties at unsampled locations.

Geostatistics deals with characteristics and properties of the sample data (e.g. mean and correlation) to simulate or estimate a model of the rock properties. Simulation generates multiple realizations of equally probable numerical models (Goovaerts, 1997). The models honor the values observed at the samples locations and reproduce the spatial characteristics observed in the data. Simulation allows assessment of risk and uncertainty.

Geologic variables are divided into two main types: categorical or continuous. Categorical variables are discrete variables defined by a name or label, such as rock types or facies, whereas continuous variables are quantifiable, such as porosity or grade. Categorical variables are modeled first, then continuous variables are modeled within each category. Categorical variables have the critical role of establishing volumes within which the continuous variables are considered stationary (C. V. Deutsch, 2006; Pyrcz & Deutsch, 2014; Rossi & Deutsch, 2013). Hence, it is essential to carefully model categorical variables to ensure accurate modeling of continuous variables.

According to Rossi and Deutsch (2013), the distribution of continuous variables is primarily governed by the stationary domains defined by categorical variables. Numerous techniques exist for the modeling of categorical variables. These methodologies are divided into stochastic and

deterministic. An explicit deterministic approach creates a model by interpreting geologic variables in cross sections and extending the interpretations to three-dimensional (3-D) models. Deterministic geologic models lack a measure of uncertainty, thus, they cannot be used for characterization of risk (Rossi & Deutsch, 2013).

Stochastic techniques are divided into object based and cell based approaches. The object based method replicates morphological patterns like channels in fluvial system. Matching observed data is challenging. Geological events often disrupt the original morphology, resulting in structures that may no longer resemble the original objects (D. Silva, 2018). Cell based techniques can match the data easily.

Among the cell based techniques, multiple point statistics (MPS) stands out as the methodology with the greatest capability of reproducing non-linear features. MPS relies on the use of a reference model known as a training image (TI).

Indicator-based techniques assign an indicator to each categorical geological variable. The indicators can be kriged with multiple indicator kriging (MIK) or simulated in a sequential manner. Indicator kriging calculates the probability of the geologic variable being present. Conditional simulation provides a probabilistic model of the geologic variable (Rossi & Deutsch, 2013). Sequential indicator simulation (SIS) and hierarchical truncated pluri-Gaussian (HTPG) simulation are examples of that method.

An assumption of multivariate Gaussianity and the linear model of coregionalization (LMC) is used when working with multivariate modeling of continuous variables. After univariate transformations of the variables, they are assumed to be multivariate Gaussian, but this is not always correct for geologic variables R. M. Barnett (2015). Poor reproduction of multivariate relationships may occur by incorrectly assuming Gaussianity. Therefore, advances have been made in the multivariate modeling of continuous variables such as principal component analysis (PCA) (Davis, 1987; Hotelling, 1933); minimum/maximum auto-correlation factors (MAF) (Desbarats & Dimitrakopoulos, 2000), projection pursuit multivariate transformation (PPMT) (R. M. Barnett, Manchuk, & Deutsch, 2014; Friedman, 1987) to overcome the limitations of the traditional methods.

Truncated Gaussian simulation (TGS) was introduced by Matheron et al. (1987) using only one Gaussian variable. Galli et al. (1994) generalized the method to truncated pluri-Gaussian simulation (TPGS) for more than one Gaussian variable. Emery and Cornejo (2010) introduced the use of truncated Gaussian simulation for modeling multivariate categorical variables applying LMC to co-simulate the latent variables. For multivariate categorical variables modeling, the most recent and advanced technique was achieved by the work of D. Silva (2018) and the development of the hierarchical truncated pluri-Gaussian technique. Part of this thesis focus on exploring the indicator based approach to see its capability in estimating categorical variables using a linear model of coregionalization. The thesis also analyzes the characteristics of indicator variograms and covariances. Additionally, the thesis develops a regularization technique to mitigate the presence of

extreme weights. Moreover, the thesis studies the decorrelation of the indicators through PCA and the independent estimation of the components.

1.2 Literature review

Geostatical methods rely on statistical models that are based on random function (RF) theory. These techniques model the uncertainty at unsampled locations. In the 1960's, Matheron (1971) developed the theory of regionalized variables, which integrates probability theory, employing random variables (RVs) and random functions (RFs) to address spatial challenges in modelling subsurface resources. A RV refers to a variable with values determined stochastically by a probabilistic mechanism (Isaaks & Srivastava, 1989). RFs, on the other hand, comprise sets of RVs defined within a specifc domain. The roots of geostatistics is also attributed to the groundbreaking work of D. G. Krige in the Witwatersrand gold fields in South Africa in the 1950s when the challenge of estimating grades arised (Krige, 1951). Matheron (1962) offered theoretical support for the unbiased estimation of mineral resources known as Kriging.

Kriging is known as a Best Linear Unbiased Estimator (BLUE) method in spatial statistics as it enables accurate predictions of unknown locations based on nearby data. Matheron (1963) was the first author to publish about kriging. The concept of a RV was developed by Matheron (1971) and forms the foundation of geostatistics.

According to Agresti (2002), there are three main groups in which categorical variables are classified: nominal, ordinal and interval. Nominal variables have no intrinsic ordering of categories. Ordinal variables exhibit a natural ordering, yet the distances between categories are unspecified. Interval variables are both ordered and possess defined numerical distances between categories. Categorical variables may also be binary (dichotomous) or consist of multiple categories (polychotomous).

Geological domains are usually delineated by multiple categories of ordinal and nominal types. For instance, ordinal categories in geology can include sedimentary sequences shaped by depositional environments and domains characterized, for example, by degrees of alteration and weathering. Nominal variables, on the other hand, lack distinct genetic shapes or could have been altered by events such as fractures and intrusions.

Pyrcz and Deutsch (2014); Rossi and Deutsch (2013) state that in the mining industry, categorical variables are mainly used to represent stationary domains; and within the stationary domains, continuous variables like grades are either estimated or simulated using geostatistical methods. There are two main ways of modeling these domains: determinstc and stochastic. Deterministic methods incorporate the parametric wireframing technique (Bezier, 1974), interpolation of volume functions (Cowan et al., 2003), and signed distance functions (McLennan & Deutsch, 2006). These approaches do not account for uncertainty and do not reproduce small scale variability (D. A. Silva, 2015). Numerous stochastic methods for modeling categorical variable have been developed and are

reviewed.

A. G. Journel (1983) proposed the application of indicators in geostatistics for the non-parametric prediction of spatial distributions. Indicators are a useful approach in cases when modeling variables with high coefficient of variation and long-tailed distributions for continuous variables. When modeling continuous variables, binning of the data is a requirement and that leads to information loss. However, this does not applies to categorical variables, which naturally possess binned distributions. The estimation is performed with indicator kriging (IK), and the spatial variability of the indicator variable is through the indicator variogram. This theory is reviewed in Chapter 2.

The indicator formalism is also the base for indicator conditional simulation (C. V. Deutsch, 2006; A. Journel & Isaaks, 1984). Sequential indicator simulation (SIS) (Alabert, 1987; C. V. Deutsch, 2006) is a Monte Carlo simulation technique implemented with sequential steps likewise the sequential Gaussina simulation. The local conditional distribution are non-parametrically estimated using the indicator kriging framework. SIS is well-suited for simulating binary categorical variables, and accommodates multiple categories by employing multiple binary indicator transformations. While SIS simulations generally replicate category proportions and their spatial distribution, they may not consistently reproduce category orderings, particularly in sparsely sampled areas. Consequently, the technique is well-suited for nominal categorical variables (D. Silva, 2018). Also, SIS realizations frequently exhibit exaggerated short-scale variability, leading to noisy categorical models. Moreover, categories with lower proportions tend to be undersampled, while those with higher proportions are oversampled, causing a disparity in global proportion statistics. One solution to address these challenges is the implementation of the maximum a posteriori selection (MAPS) technique (C. V. Deutsch, 1998). The limitation of cleaning the realizations include the potential of hiding the true short scale-variability and the necessity for users to define parameters arbitrarily to ensure reproduction of the conditioning data C. V. Deutsch (1998).

Multiple point statistics is another method for categorical variables modeling that does not rely on variograms. MPS considers relations among multiple points simultaneously, offering an alternative to traditional kriging-based methods that rely on two point statistics. MPS was initially proposed by A. Journel and Alabert (1989) and was then used in simulation by C. V. Deutsch (1992), Guardiano and Srivastava (1993), S. B. Strebelle (2000), Lyster (2009), and others. Unlike conventional geostatistics, MPS bypasses the explicit definition of a RF. MPS can generate categorical models that capture complex nonlinear in mineral deposits (Guardiano & Srivastava, 1993; A. G. Journel, 2004). The method relies on the principle of single normal equations (SNE). Single normal equation simulation (SNESIM) serves as a streamlined MPS algorithm initially introduced by (Guardiano & Srivastava, 1993). S. Strebelle (2002) offers a comprehensive examination of the progression of these MPS algorithms and introduces an efficient non-iterative algorithm (SNESIM).

In MPS, the spatial heterogeneity is deduced from a training image. The TI is seen as a rasterized illustration of the categories being analyzed (Boucher, 2009) and serves as the prior

model of spatial structure (A. Journel & Zhang, 2006). Also, a TI is a facies or rock type model that is exhaustively populated by the the categories of interest (Boisvert, Pyrcz, & Deutsch, 2007; C. V. Deutsch, 1992). Utilizing a dataset as a training image ensures the geological realism of MPS. Nonetheless, obtaining a fully representative dataset (TI) of the deposit of interest is often challenging. MPS-based techniques enable conditioning, yet they remain computationally intensive, particularly regarding memory demands. Furthermore, a key challenge in utilizing MPS is the need for a training image that precisely reflects the characteristics and statistics of the domain of interest. Employing a representative training image is crucial for generating realistic realizations that faithfully capture the studied phenomena (Boisvert et al., 2007; A. Journel & Zhang, 2006).

Truncated Gaussian simulation (TGS), introduced by Matheron et al. (1987), offers an alternative for simulating categorical variables. This method operates on the assumption that categorical variables emerge from truncating an underlying latent continuous variable. The latent variable is a realization of a Gaussian random function (GRF), and the truncation rule governs the ordering and proportion of each category. Galli et al. (1994) proposed the truncated pluri-Gaussian simulation (TPGS), which is an extension of the TGS, to employ an arbitrary number of GRFs as latent variables. This extension enables the simulation of more intricate structures of categories, including ordinal categorical variables such as alteration zones in lateritic nickel or depositional layers in petroleum reservoirs. TPGS aims to replicate categorical proportion, two-point spatial correlation, and transition probabilities.

The truncation rule, also known as the lithotype rule, plays a crucial role in the TPGS approach by regulating the interactions, transitions, and proportions among categories. Typically, the definition of truncation rules relies on transition probabilities observed in data and the conceptual geological model (Mariethoz, Renard, Cornaton, & Jaquet, 2009). The practical implementation of TPGS is frequently constrained to incorporating no more than three Gaussian latent variables. This limitation was primarily due to the prevailing method of defining truncation rules through truncation masks. D. Silva (2018) addresses this constraint by introducing the hierarchical truncated pluri-Gaussian (HTPG) approach. HTPG employs a tree structure to truncate Gaussian latent variables, making it easier to define based on geological expertise. This methodology enables the use of any number of latent variables to model any number of categories, thereby it increases the potential of the truncated Gaussian method (D. Silva, 2018). Moreover, the HTPG framework developed by D. Silva (2018) also accommodates the modeling of multiple categorical variables. This is achieved by enabling correlations among the latent Gaussian variables that define each categorical variable. This enhancement optimizes the utilization of available information by avoiding the consolidation of multiple categorical variables into a single set. HTPG is, perhaps, the most advance method for modeling categorical variables up to date.

Significant progress has been achieved in multivariate modeling of continuous variables in recent years. These modern apporaches are based on multivariate transformations such as principal

component analysis (PCA). PCA was first introduced by Hotelling (1933) and Pearson (1901) and serves as a method for dimension reduction and decorrelation. It transforms a correlated multivariate distribution into orthogonal linear combinations of the original variables. PCA proves valuable in geostatistical modeling for two primary reasons:

1. Multivariate data, comprising several correlated geological variables, are decorrelated by PCA, making them uncorrelated. This allows for independent geostatistical modeling of the decorrelated variables, followed by PCA back-transformation to restore the original correlation to the modeled variables (R. M. Barnett, 2017).

2. PCA can be employed for dimension reduction within this framework. Independent geostatistical modeling proceeds using a subset of the decorrelated variables, and the PCA back-transformation provides models for all original variables (R. M. Barnett, 2017).

Suro (1988) applied kriging the principal component of the original indicator variables for continuous variables. Suro (1988) was able to infer the conditional cumulative distribution function (CDF) by kriging the principal components and then applying a back transformation.

The primary objective of this thesis is to investigate the potential of cokriging with collocated data in situations where an exhaustive secondary dataset is not available by using the full information of covariance and cross-covariance between the indicator variables. Additionally, this research will explore the decorrelation of indicator variables to assess the impact of estimating the components independently. To achieve this goal, the thesis will examine the characteristics of categorical indicator variables. It will compare the performance of Multiple Point Statistics (MPS), indicator cokriging, and indicator ordinary kriging for categorical variables. Furthermore, the thesis will present a regularization method to avoid extreme weights when applying cokriging. Finally, it will explore the use of Principal Component Analysis (PCA) for indicators to decorrelate categorical indicators and estimate the components independently.

1.3 Thesis statement and contribution

This thesis aims to advance the geostatistical modeling of categorical variables and, consequently, improve resource estimation by: a) thoroughly investigating the characteristics of categorical variable indicators to gain a deeper understanding of them; b) assessing the potential of utilizing the full information of covariance and cross-covariances for estimation through cokriging; and c) exploring the potential of decorrelating the indicator variables and estimating them independently. The main contributions of this thesis are as follows:

1. While analyzing the characteristics of the indicator variogram for categorical variables a demonstration is presented showing that the nugget effect is zero. The correlogram is shown to be more robust than the variogram.

2. Although indicator ordinary kriging gives optimal estimates of categorical variables, there are circumstances where the use of the full information of covariance and cross-covariance will improve the estimates through standardized ordinary cokriging.

3. The systems of equations involved in cokriging of categorical variables could lead to system that are close to being non-positive definite. Therefore, the systems can produce extreme solutions weights that could lead to unacceptable estimates. A regularization method is presented in this thesis that mitigates the presence of extreme weights.

4. Lastly, the decorrelation method of PCA was applied to the indicator of categories to decorrelate and independently estimate them. The original estimates are obtained after a back-transformation. Although, being a curious method because of the dimension reduction, the estimates are not better than indicator ordinary kriging.

The insights obtained from this thesis contribute to a greater understanding of categorical variable indicators. An outline of the thesis is presented in the following section.

1.4 Thesis outline

The next chapter presents the necessary theoretical background of the subjects discussed in this thesis. It includes a summary of the theory of the geostatistical methods for modelling categorical variables that are part of this thesis. The chapter explains the indicator formalism, indicator kriging, standardized oridinary cokriging and MPS. Additionaly, it includes the theory behind PCA.

The third chapter describes the characteristics of the spatial correlation of the categorical indicators. It analyzes the variogram, covariance and cross-covariance of the indicator categorical variable. Moreover, it analyzes the behavior of the variogram and correlogram in presence of clustered data.

The fourth chapter compares MPS, indicator ordinary kriging, and standardized ordinary cokriging. These comparisons were done with several different TI's and a summary of the results is presented. The cases where standardized ordinary cokriging outperforms indicator ordinary kriging are presented too.

The fifth chapter describes the regularization method for cases where cokriging systems produce extreme solution weights. The regularization of the cokriging systems mitigates the presence of extreme weights. Locally, the estimates are improved, although it did not make a significant impact globally.

The sixth chapter presents the idea of decorrelating categorical indicator variables. The decorrelation is done through PCA. There is an interesting dimension reduction as the last principal component always have zero variance. The components are kriged independently and then back-transformed to the original unit. This was done on a synthetic case and the results are presented.

The seventh chapter provides the concluding remarks. This chapter summarizes the results and contributions. Chapter 7 also outlines the limitations of the thesis and suggests future work.

Chapter 2

Theoretical Background

This thesis involves the modeling of categorical variables. This chapter presents the theoretical background of the concepts involved. Section 2.1 explains the concept of stationarity. Section 2.2 describes the indicator formalism, including the mean and variance of an indicator variable. Section 2.3 reviews the concepts of the variogram, covariance, cross variograms, and cross covariances of indicators. Section 2.4 presents both simple and ordinary indicator kriging. Simple cokriging and standardized ordinary cokriging are discussed in Section 2.4. Section 2.5 reviews the theory of MPS, and Section 2.6 describes the theory of PCA.

2.1 Stationarity

The concept of stationarity involves the pooling of data for subsequent geostatistical assessment. As stationarity is not a hypothesis, it cannot be tested (Pyrcz & Deutsch, 2014). Once more data is available, or when the geostatistical analysis has started, the decision of stationarity may be reconsidered. For categorical variables, the importance of trend modeling increases significantly, particularly in the field of earth sciences, where categories are almost always non-stationary.

2.2 Indicator formalism

Consider K distinct categories. These categories are mutually exclusive, meaning only one category can be present at any given location. Additionally, they are exhaustive; one of the categories must be present at all locations. The categorical variable is represented by a series of K indicator variables within a domain $\mathcal{A}$ where $\mathbf{u}$ indicates a location (C. V. Deutsch, 2006; A. G. Journel, 1983):

$$i_k(\mathbf{u}) = \begin{cases} 1, & \text{if category } k \text{ prevails at location } \mathbf{u} \\ 0, & \text{otherwise} \end{cases}, \quad k = 1, \ldots, K \tag{2.1}$$

A. G. Journel (1983) states that indicators describe the probability of a specific category k being present at a location. In other words, when the indicator is 1, the category is present, and when it is 0, the category is absent; see Figure 2.1 retrieved from Mizuno and Deutsch (2022).

While hard measurements are encoded as binary values 0s or 1s, soft or imprecise measurements are encoded as continuous probabilities between 0 and 1. Geostatistical inference is performed using indicator data and includes declustering to ensure representative proportions and variography to analyze the spatial continuity of each of the K indicator variables (C. V. Deutsch, 2006).

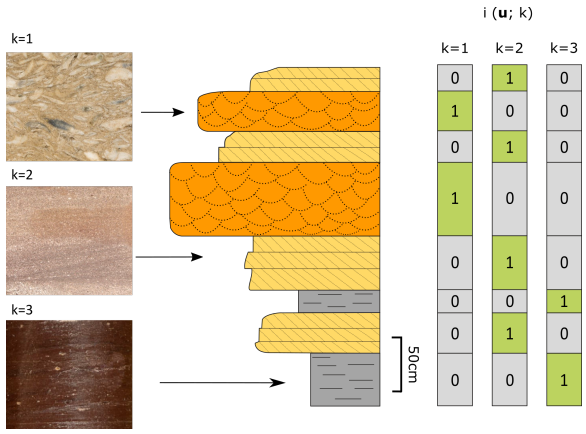


Figure 2.1: Facies description and their respective indicators. Retrived from Mizuno and Deutsch (2022)

The expected value (mean) of an indicator variable is the proportion of that variable, see Equation 2.2 :

$$E\left\{I_k\right\} = p_k \tag{2.2}$$

The variance, also known as the second moment, of an indicator is another crucial parameter. The variance demonstrates the variability of an indicator variable and is also dependent on the proportion. Therefore, the variance is defined as Equation 2.3:

$$\begin{aligned} \text{Var}\left\{I_k(\mathbf{u})\right\} &= E\left[\left(I_k(\mathbf{u}) - p_k\right)^2\right] \\ \text{Var}\left\{I_k(\mathbf{u})\right\} &= E\left\{I_k(\mathbf{u})^2\right\} - \left[E\left\{I_k\right\}\right]^2 \\ \text{Var}\left\{I_k(\mathbf{u})\right\} &= p_k - p_k^2 \\ \text{Var}\left\{I_k(\mathbf{u})\right\} &= p_k\left(1 - p_k\right) \end{aligned} \tag{2.3}$$

Where $E\left\{I_k(\mathbf{u})^2\right\}$ is the expected value of the squared indicator, which is equal to p_k, since the square of a variable that is only 0 or 1 is the same as the variable itself. The estimation is usually performed by kriging, thus the term indicator kriging (IK).

2.3 Indicator variogram and covariance

Two-point statistics, such as variogram and covariance, are used to understand the spatial behavior of the indicator variable. The indicator variogram is used to characterize the spatial heterogeneity of the indicator variable. Spatial statistics play a crucial role in defining uncertainty and heterogeneity. The experimental variogram for category k separated by a lag vector $\mathbf{h}$ is derived from the indicators (C. V. Deutsch & Journel, 1997), see Equation 2.4:

$$2\gamma_k(\mathbf{h}) = E\left\{\left[I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h})\right]^2\right\}, \quad k = 1, \ldots, K \tag{2.4}$$

In the direct indicator variogram for category k, only the average transition from category k to any other category between two locations are taken into account. The cross indicator variogram between categories k and k' is defined as Equation 2.5:

$$2\gamma_{kk'}(\mathbf{h}) = E\left\{\left[I_k(\mathbf{u}) - I_k(\mathbf{u}+\mathbf{h})\right]\left[I_{k'}(\mathbf{u}) - I_{k'}(\mathbf{u}+\mathbf{h})\right]\right\} \quad k' \neq k \quad k, k' = 1, 2, \ldots, K \tag{2.5}$$

The covariance of an indicator variable at two spatial locations separated by a lag vector $\mathbf{h}$ is defined as Equation 2.6:

$$C_k(\mathbf{h}) = E\left\{I_k(\mathbf{u}) \cdot I_k(\mathbf{u}+\mathbf{h})\right\} - E\left\{I_k(\mathbf{u})\right\} \cdot E\left\{I_k(\mathbf{u}+\mathbf{h})\right\} \quad k = 1, 2, \ldots, K \tag{2.6}$$

In geostatistics, the indicator covariance is usually calculated through its relationship with the indicator variogram when stationary. The relationship is given by Equation 2.7:

$$C_k(\mathbf{h}) = C_k(0) - \gamma_k(\mathbf{h}), \quad k = 1, 2, \ldots, K \tag{2.7}$$

where $C_k(0)$ is the covariance at lag zero or the variance.

The indicator cross covariance between categories k and k' is defined as Equation 2.8:

$$C_{kk'}(\mathbf{h}) = E[I_k(\mathbf{u}) \cdot I_{k'}(\mathbf{u}+\mathbf{h})] - E\{I_k(\mathbf{u})\} \cdot E\{I_{k'}(\mathbf{u}+\mathbf{h})\} \quad k' \neq k \quad k, k' = 1, 2, \ldots, K \tag{2.8}$$

The modeling of experimental indicator variograms is carried out using valid models like spherical and exponential. These models are employed to construct the kriging systems necessary for estimating the indicator variable at the model nodes. A separate kriging system is solved for each category at every location. The Gaussian variogram model is not compatible with indicator random functions, even if the criteria of positive definiteness is satisfied (Armstrong, 1992; Christakos, 1984).

2.4 Indicator kriging

Explicit random function models for indicators are not available. Kriging is used to estimate local conditional probability distributions. The optimal choice of kriging is influenced by the decision of stationarity and the utilization of secondary information (Mizuno & Deutsch, 2022). Simple kriging (SK), ordinary kriging (OK), simple cokriging, and standardized ordinary cokriging are reviewed.

2.4.1 Simple kriging

SK calculates the probability for each category using the conditioning data and a declustered global mean. The estimator is defined as Equation 2.9:

$$i^*_{SK}(\mathbf{u}; k) - p_k = \sum_{\alpha=1}^{n} \lambda_\alpha^{SK}(\mathbf{u}; k) \cdot \left[i\left(\mathbf{u}_\alpha; k\right) - p_k\right], \quad k = 1, 2, \ldots, K \tag{2.9}$$

where λ_α^{SK} are the weights calculated from the simple kriging systems of equations for each location and category.

2.4.2 Ordinary kriging

In ordinary kriging, the sum of the weights is constrained to equal one, ensuring that the global mean is not weighted and all the weight is allocated to the sample data. The estimator is defined as Equation 2.10:

$$i^*_{OK}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda_\alpha^{OK}(\mathbf{u};k) \cdot i\left(\mathbf{u}_\alpha;k\right) \tag{2.10}$$

where λ_α^{OK} are the weights calculated from the ordinary kriging systems of equations for each location and category.

2.4.3 Simple cokriging

The estimator for simple cokriging closely resemble that of simple kriging but it incorporates both the direct and cross covariances. The estimator is defined as Equation 2.11:

$$i^*_{\mathrm{SCK}}(\mathbf{u}_0;k_0) - p_{k_0} = \sum_{k=1}^{K}\sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{SCK}}(\mathbf{u}_0;k_0;k)\left[i\left(\mathbf{u}_\alpha;k\right) - p_k\right] \tag{2.11}$$

Where $i^*_{\mathrm{SCK}}(\mathbf{u}_0;k_0)$ is the simple cokriging estimate of the indicator variable at location $\mathbf{u}_0$ for category k_0; p_k is the proportion of category k. The double summation runs over all categories (k from 1 to K) and all sample locations (α from 1 to n). $\lambda_\alpha^{\mathrm{SCK}}(\mathbf{u}_0;k_0;k)$ are the simple cokriging weights. $i\left(\mathbf{u}_\alpha;k\right)$ is the indicator value at location $\mathbf{u}_\alpha$ for category k.

2.4.4 Standardized ordinary cokriging

A commonly preferred method, the standardized ordinary cokriging, involves creating new secondary variables that share the same mean (proportion) as the primary variable. This method imposes a constraint where all the weights must sum to one (Rossi & Deutsch, 2013). The estimator is defined as Equation 2.12:

$$\frac{i^*_{\mathrm{SOCK}}\left(\mathbf{u}_0;k_0\right) - p_{k_0}}{\sqrt{p_{k_0}\left(1-p_{k_0}\right)}} = \sum_{k=1}^{K}\sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{SOCK}}\left(\mathbf{u}_0;k_0;k\right)\frac{i\left(\mathbf{u}_\alpha;k\right) - p_k}{\sqrt{p_k\left(1-p_k\right)}} \tag{2.12}$$

Where $\sum_{K=1}^{K}\sum_{\alpha=1}^{n} \lambda_\alpha^{\mathrm{SOCK}}\left(\mathbf{u}_0;k_0;k\right) = 1;\ k_0 = 1,\ldots,K$, is the category being estimated and k represents the indicator of the data. $\frac{i^*_{\mathrm{SOCK}}\left(\mathbf{u}_0;k_0\right)-p_{k_0}}{\sqrt{p_{k_0}\left(1-p_{k_0}\right)}}$ represents the standardized estimated value for category k_0 at location $\mathbf{u}_0$. $\lambda_\alpha^{\mathrm{SOCK}}\left(\mathbf{u}_0;k_0;k\right)$ are the weights assigned to each data point for each category in the cokriging system. $i\left(\mathbf{u}_\alpha;k\right)$ represents the observed value of the indicator variable for category k at location $\mathbf{u}_\alpha$. $\sqrt{p_k\left(1-p_k\right)}$ is the standard deviation of the indicator variable for category k.

Kriging may result in negative weights being assigned to data that are screened by nearer samples. Typically, this aids in local extrapolation and enhances estimation accuracy. However, occasionally, based on the variogram and the arrangement of local data, this can result in negative estimates for

indicators. A standard practice in such scenarios is to adjust negative estimates to zero and then rescale the remaining estimates by their sum (C. V. Deutsch, 2006).

2.5 Multiple point statistics

Unlike conventional geostatistics, multiple point statistics does not depend on the explicit definition of a random function. It instead extracts the required multivariate distributions from training images. This approach sets it apart from traditional geostatistics, which relies on two point statistics via a covariance or variogram model (Hu & Chugunova, 2008). The advantage of MPS lies in its applicability across diverse geological settings, provided there is a TI that accurately represents the geological heterogeneity.

MPS derives the conditional probability directly from a TI that captures the geological and geometric characteristics of the desired physical property. Instead of formulating a random function model, the primary task lies in developing an authentic TI. The essence of MPS is the process of deducing the conditional distribution from a TI (Guardiano & Srivastava, 1993; Hu & Chugunova, 2008).

Analyzing the relative frequencies in a dataset enables the estimation of probabilities for multiple point events. However, it is crucial to acknowledge that accurate calculation of these frequencies, and thus reliable inference, is only possible when there is an adequate number of repetitions of an event (Ortiz & Deutsch, 2004).

Consider a TI where three facies are present: blue, white, and red, see Figure 2.2. Also, consider an event configuration of three sample data and an unknown location (Figure 2.3). The configuration setting is:

- Sample 1 (red) is +3 grid units northing and -1 easting.
- Sample 2 (blue) is +3 grid units easting.
- Sample 3 (blue) is -2 units northing.

To determine what facies will be estimated at the unknown location, Bayes Law is recalled to calculate the conditional probabilities, see Equations 2.13:

$$
\begin{aligned}
f(0 = \text{ blue } \mid 1 = \text{ red, } 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{f(0 = \text{ blue }, 1 = \text{ red, } 2 = \text{ blue }, 3 = \text{ blue })}{f(1 = \text{ red }, 2 = \text{ blue }, 3 = \text{ blue })} \\
f(0 = \text{ white } \mid 1 = \text{ red, } 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{f(0 = \text{ white, } 1 = \text{ red, } 2 = \text{ blue }, 3 = \text{ blue })}{f(1 = \text{ red }, 2 = \text{ blue }, 3 = \text{ blue })} \\
f(0 = \text{ red } \mid 1 = \text{ red, } 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{f(0 = \text{ red }, 1 = \text{ red }, 2 = \text{ blue }, 3 = \text{ blue })}{f(1 = \text{ red }, 2 = \text{ blue }, 3 = \text{ blue })}
\end{aligned}
\tag{2.13}
$$

In Equations 2.13, the four-variate joint probabilities (numerators) and the marginal probabilities (denominators) must be deduced from the training image (TI). These probabilities are acquired by scanning the TI and directly computing the necessary probabilities, see Figure 2.4.

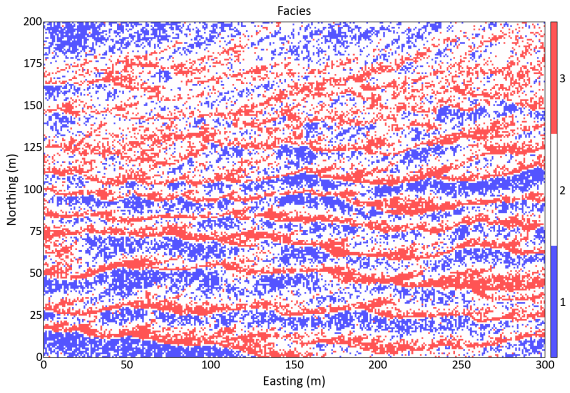


Figure 2.2: TI example.

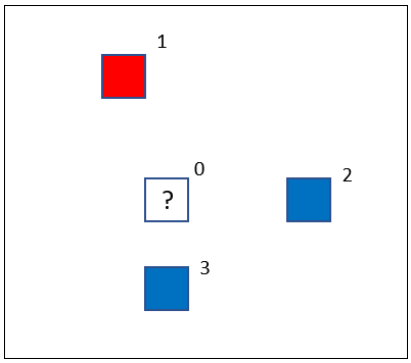


Figure 2.3: Samples configuration.

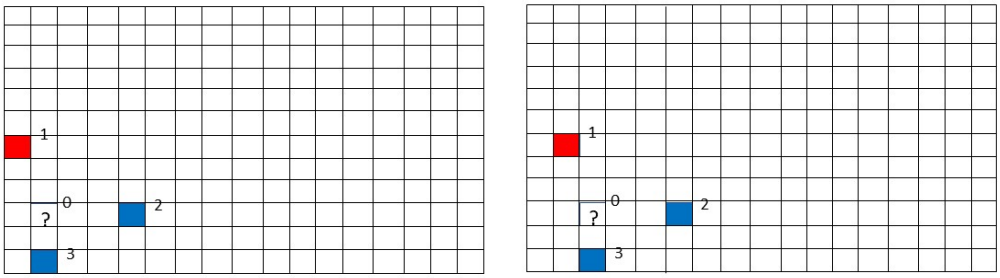


Figure 2.4: TI scanning to obtain probabilities. The image on the right is moving +1 on x axis.

Figure 2.4 illustrates an example of scanning a gridded TI for a particular arrangement of samples, recording the findings, and computing the required probabilities. A challenge with MPS is that the combination of configurations becomes unmanageable with the expansion of the number of categories and grid nodes.

In the given example, after calculating the conditional probabilities using Bayes' Law, the results are as follows:

$$
\begin{aligned}
f(0 = \text{ blue } \mid 1 = \text{ red}, 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{0.014068}{0.021985} = 0.640 \\
f(0 = \text{ white } \mid 1 = \text{ red}, 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{0.007380}{0.021985} = 0.336 \\
f(0 = \text{ red } \mid 1 = \text{ red}, 2 = \text{ blue }, 3 = \text{ blue }) &= \frac{0.000537}{0.021985} = 0.024
\end{aligned}
\tag{2.14}
$$

Note that the sum of these probabilities is equal to one, within a round-off error. The probability of encountering white is higher than red, reflecting the greater proportion of white in the TI. However, problems may arise if the data configuration is not reliably sampled in the TI.

2.6 Principal component analysis (PCA)

PCA is an algebraic method used to transform one vector into another. In two-dimensional (2-D), PCA can be viewed as an axis rotation, where the rotation angle is selected to maximize the spread of the first transformed variable along the first axis and minimize the spread of the second transformed variable along the second axis while making the correlation equal to zero along the rotated axes.

Consider K indicator variables $I_1, \ldots, I_K$ that will be estimated. The conditioning data is represented by a matrix $\mathbf{I}$: $i_{\alpha,k}, \alpha = 1, \ldots, n, k = 1, \ldots, K$, where n is the number of samples. Variables must be standardized to have a mean of zero and a variance of one, as this facilitates the interpretation of the PCA results. Therefore, the standardization of the indicators is done before PCA transformation, see Equation 2.15:

$$\mathbf{Y} : y_{\alpha,k} = \frac{\left(i_{\alpha,k} - p_k\right)}{\sigma_k}, \text{ for } \alpha = 1, \ldots, n, k =, 1, \ldots, K \tag{2.15}$$

where p_k is the proportion of I_k, σ_k^2 is the variance of I_k, and σ_k is the standard deviation. Every standardized variable Y_k has mean of zero and variance of one. PCA centers on the covariance matrix $\mathbf{\Sigma}$ of the standardized data:

$$\mathbf{\Sigma} : C_{k,k'} = \frac{1}{n} \sum_{\alpha=1}^{n} y_{\alpha,k} \cdot y_{\alpha,k'}, \text{ for } k, k' = 1, \ldots, K \tag{2.16}$$

The values in $\mathbf{\Sigma}$ characterize the multivariate system of the $\mathbf{Y}$ data regarding linear variability and interdependence. $C_{k,k}$ are the diagonal elements and represent the variance of each Y_k. The off-diagonal elements $C_{k,k'}$, $k \neq k'$, show the covariance between Y_k and $Y_{k'}$. Given that each Y_k has a variance of one, these covariances are equivalent to correlations.

The PCA tranform starts with the spectral decomposition (Equation 2.17) of the covariance matrix ($\mathbf{\Sigma}$) at $h = 0$ resulting in the eigenvector matrix $\mathbf{V}$ with elements $v_{k,k'}$ where $k, k' = 1, \ldots, K$ and the diagonal eigenvalue matrix $\mathbf{D}$ with elements $d_{k,k}$ where $k = 1, \ldots, K$:

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{2.17}$$

The PCA transformation is then executed by multiplying the matrix $\mathbf{Y}$ with matrix $\mathbf{V}$, see Equation 2.18:

$$\mathbf{P} = \mathbf{Y}\mathbf{V} \tag{2.18}$$

This process rotates the multivariate data, ensuring that the resulting principal components in

$\mathbf{P}$ are uncorrelated. By multiplying $\mathbf{P}$ with the transpose of $\mathbf{V}$, the data rotates back to its original form in $\mathbf{Y}$, providing the back transformation, see Equation 2.19.

$$\mathbf{Y} = \mathbf{P}\mathbf{V}^T \tag{2.19}$$

The eigenvector matrix can be viewed as a rotational matrix that establishes a new basis, transforming the correlated data into an orthogonal form. The product of $Y_1, \ldots, Y_K$ with the k^{th} column of $\mathbf{V}$ yields the P_k principal component. Therefore, each principal component is essentially a linear combination of the original variables, clarifying the term of linear combination.

Each $d_{k,k}$ value in the matrix corresponds to the variance P_k, and it also indicates the amount of variability that P_k accounts for in the $Y_1, \ldots, Y_K$ multivariate system. To specify the percentage of variability that P_k explains in $Y_1, \ldots, Y_K$ is determined as $d_{k,k} / \sum_{k'=1}^{K} d_{k',k'} \cdot 100$. The first component P_1 accounts for the highest amount of variability, followed by P_2 for the second highest, and so on (R. M. Barnett, 2017).

In this chapter, a theoretical foundation relevant to this thesis is presented. The concept of stationarity was introduced then the indicator formalism was explained. Additionally, the chapter discusses variograms and covariance, including both cross variogram and cross covariance. There is also a review of indicator kriging methods that are involved in this dissertation. The chapter showed a section on MPS theory, accompanied by a brief example, and concluded with an explanation of the principles underlying PCA transforms.

Chapter 3

Characteristics of Categorical Indicator Variogram/Covariance

This chapter presents the characteristics of categorical indicator variogram and covariance functions. Section 3.1 explains the link between variograms and geometry. Section 3.2 describes the sill and range of a variogram. Section 3.3 demonstrates that the nugget effect for a categorical indicator variogram must be zero. Section 3.4 deduces a non-stationary relationship between the variogram and covariance. Section 3.5 suggests a method for variogram modeling in cases with clustered samples. Finally, Section 3.6 compares the variogram with the correlogram to determine which experimental measure is more robust in various scenarios.

3.1 Link to geometry

The variogram is crucial in geostatistics for modeling spatial relationships. The spherical and exponential models are the most commonly used due to their conditional negative definiteness which ensures a positive definite covariance matrix. These models are essential for understanding how spatial correlation changes with a certain vector lag ($\mathbf{h}$), a key aspect in predicting unobserved locations based on known data. However these traditional models sometimaes struggle with complex geological data. According to C. V. Deutsch and Journel (1997), linear combinations of spherical and exponential models are valid functions and effective for modeling a wide range of variograms. Despite this, there remain scenarios where even these combined models fail to capture the complexity of the spatial relationships.

Geometric variograms offer a solution for situations where traditional variogram models are insufficient. These variogrmas focus on the geometric properties of spatial features. The key is their foundation in the autocorrelation function of geometric objects. This function measures the similarity of a shape with itself when displaced by a certain vector lag ($\mathbf{h}$) (Pyrcz & Deutsch, 2006). The function is defined as the volume of the intersection $K_v(\mathbf{h})$ of the object V with itself scaled by the object's volume $K_v(0)$, see Equation 3.1.

$$\gamma(\mathbf{h}) = 1 - \frac{K_v(\mathbf{h})}{K_v(0)} \tag{3.1}$$

By normalizing this intersection volume by the object's total volume, a variogram that correctly represents the spatial structure and continuity of the feature is obtained.

The spherical model, $\text{sph}(h) = 1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3$ (for $h \leq a; 1$ for $h > a$); where a is the range, describes the spatial dependence of a variable, assuming that the function has a spherical

shape. This model is often used for isotropic and homogeneous data, such as geological formations. It can also be defined in terms of the volume intersection between two spheres separated by a lag vector $\mathbf{h}$ (Equation 3.2) (Serra, 1967).

$$\gamma(\mathbf{h}) = 1 - \frac{\text{volume}(\mathbf{h})_{\text{int}}}{\text{volume}_{\text{total}}} \tag{3.2}$$

Constructing a variogram model based on an elementary geologic shape, such as spheres, does not automatically ensure that kriged or simulated models will replicate those exact shapes. For instance, when spheres are randomly embedded within a matrix, the resulting function will not have a spherical shape. It will depend on the proportion of spheres present in the matrix ($p_1 = 1 - p_0$) where p_0 is the proportion of indicator 0 falling outside the sphere. $1 - p_0 = p_1$; p_1 is the proportion of indicator 1 falling inside the sphere. The resulting indicator variogram model relates to the spherical model but is not the same as the spherical variogram (Pyrcz & Deutsch, 2006):

$$\gamma(\mathbf{h}) = p_0 \left(1 - p_0^{\text{sph}(\mathbf{h})}\right) \tag{3.3}$$

Equation 3.3 can account for any geometric variogram in the exponent (Pyrcz & Deutsch, 2006).

3.1.1 Sill and range

The indicator variogram $\gamma(\mathbf{h}, k)$, where $k =, 1, \ldots, K$ are the categories' indicators, represents the probability of transitioning from inside a region A at location $\mathbf{u}$ to outside of A at point $\mathbf{u} + \mathbf{h}$, with the transition occurring over the extent of the separation vector $\mathbf{h}$. When the separation lag vector $\mathbf{h}$ is very large, the indicator variables $I(\mathbf{u} + \mathbf{h}; k)$ and $I(\mathbf{u}; k)$ become independent. This means the probability of transitioning becomes the product of the probability of $\mathbf{u}$ being in A and the probability of $\mathbf{u} + \mathbf{h}$ not being in A, which is equal to $p(1 - p)$ and is the sill; where p is the proportion of indicator inside A. The maximum sill is reached when $p = 0.5$, the sill is 0.25.

The range of the indicator variogram represents the distance at which the variogram reaches the sill (variance) value. In other words, the range is the lag distance at which the variogram value reaches the sill. Beyond this distance and without a trend, this value remains constant. The range depends on the proportion of spheres in the matrix because it determines how much data falls within and outside the spheres. With p being high (high proportion of spheres), there are fewer transitions between one category to another, thus, the range of the indicator variogram would be larger. This happens because the category changes over a larger distance, so transitions are less, even at small distances, making the range larger. The opposite is true, when there are more transitions between categories, the variogram range will be smaller.

3.1.2 Nugget effect

The behavior of a categorical indicator variogram at the origin (when lag distance $\mathbf{h} = 0$) is an important characteristic to consider when interpreting the results of a variogram analysis. At the

origin, the two locations being considered are the same, and transitions should not exist; thus, the variogram at the origin (Nugget Effect) must be set to zero.

Consider a large string of size L where intervals of category 1 of length l are distributed over the string. The length l is above the minimum discretization lag of the variogram; also consider lags of size b and $2b$ that are smaller than the minimum discretization lag. For size b lag, there are two transitions per interval, while for size $2b$, there are four transitions per interval. The number of lags when using $2b$ is one less than that when using b, therefore $N_{\text{lag}2b} = N\text{lag}_b - 1$, see Figure 3.1.

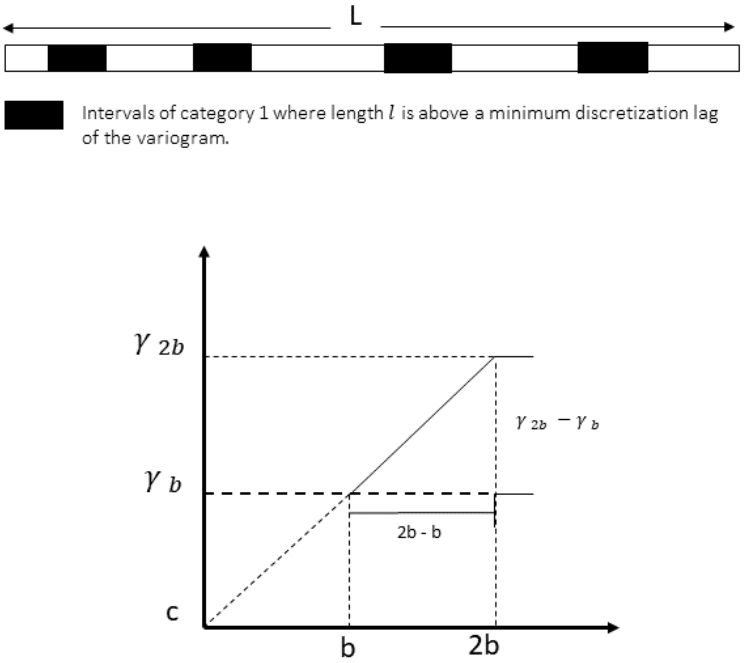


Figure 3.1: String of size L where intervals of category 1 of length l are distributed over the string. Variogram of b and $2b$ and the intercept c.

Analyzing Figure 3.1 we can calculate the slope of the variogram by:

$$\text{slope } = \frac{\gamma_{2b} - \gamma_b}{b} \tag{3.4}$$

Thus, the intercept (c) of that line is:

$$c = \gamma_b - \frac{\left(\gamma_{2b} - \gamma_b\right)}{b} b = 2\gamma_b - \gamma_{2b} \tag{3.5}$$

The variogram can be obtained by the number of transitions divided by 2 times the number of lags. Therefore, the variogram for b is:

$$\gamma_b = \frac{2N_{\text{int}}}{2N_{\text{lag}_b}} \tag{3.6}$$

and the variogram for $2b$ is:

$$\gamma_{2b} = \frac{4N_{\text{int}}}{2N_{\text{lag}_{2b}}} = \frac{2N_{\text{int}}}{N_{\text{lag}_b} - 1} \tag{3.7}$$

Substituting γ_b and γ_{2b} in the intercept (c), the equation would give $c = 2\frac{N_{\text{int}}}{N_{\text{lag}}} - \frac{2N_{\text{int}}}{N_{\text{lag}}-1}$. When the

number of lags is very large, the intercept tends to 0:

$$\lim_{N_{\text{lag}} \to \infty} c = 0 \tag{3.8}$$

As the number of lags for a lag of size b increases for the same total distance L, it means that the size b decreases. Therefore it means that as $N_{\text{lag}} \to \infty$, $b \to 0$. Thus the limit of the intercept c when $b \to 0$:

$$\lim_{b \to 0} c = 0 \tag{3.9}$$

Therefore, at the origin, the nugget effect for a categorical indicator variogram is zero. This conceptualization aligns with the understanding that indicators have linear increments. As such, only linear models like spherical and exponential are applicable. Gaussian variogram models cannot be applied to categorical indicator variables. The size of the facies intervals should be large relative to the block size or discretization interval, which supports the reasoning that the nugget effect on the categorical indicator variograms should be zero. When the intervals are considerably smaller than the block size, this leads to mixing of categories within each block, and treating the variable as categorical becomes invalid (Mizuno & Deutsch, 2022). Figure 3.2 illustrates the behavior at the origin of a standardized variogram (γ_s).

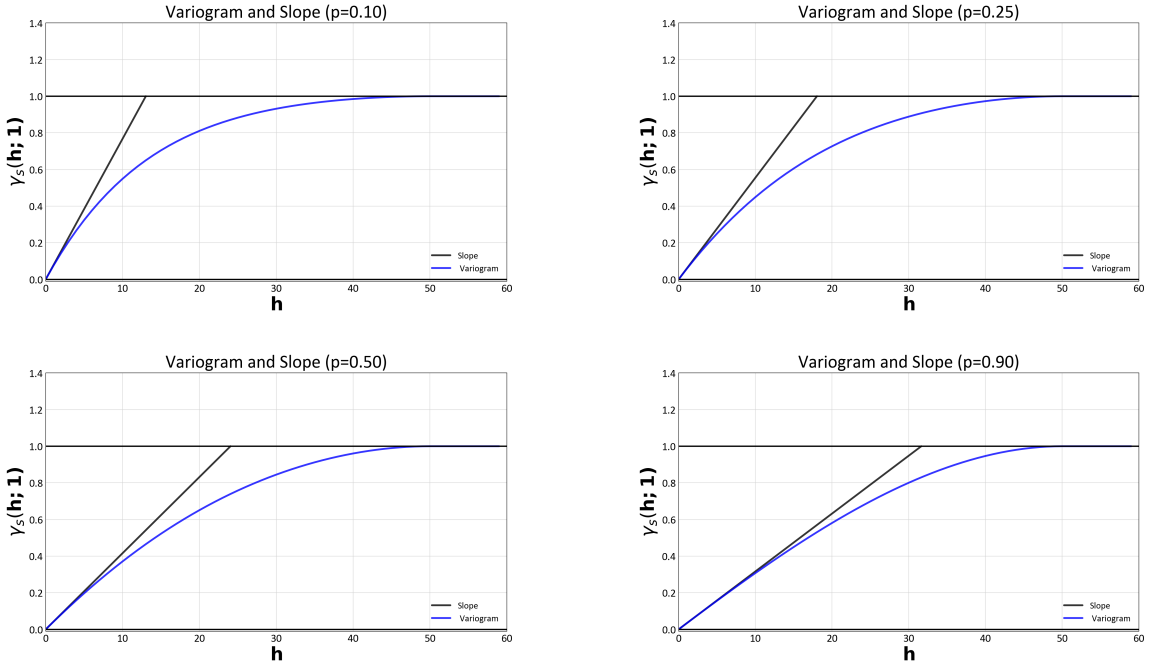


Figure 3.2: Variograms with tangents plotted at the origin, demonstrating that at the origin, the variogram increases linearly, and the slope varies depending on the proportion

At the origin, the variogram increases linearly, the nugget effect is zero, and the slope's inclination varies with the proportion, as shown in Figure 3.2.

3.2 Variogram - Covariance relationship

Consider a binary case (0;1) and a particular lag (**h**). There are only four possible outcomes that a pair of indicators could satisfy: the probability of being a pair of 1's (P_{11}), a pair of 0's (P_{00}), a pair of 1 and 0 (P_{10}), or a pair of 0 and 1 (P_{01}). Thus, the bivariate distribution is fully defined by P_{11}, P_{00}, P_{10}, and P_{01}. This is illustrated in Figure 3.3.

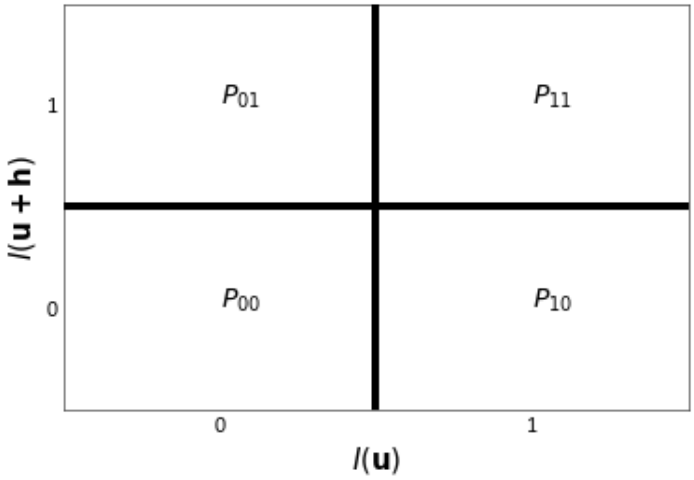


Figure 3.3: Quadrants illustrating the probabilities of indicator pair combinations at lag **h**.

Given this understanding, the variogram and covariance are calculated as follows:

$$2\gamma(\mathbf{h};1) = P_{10} + P_{01} \tag{3.10}$$

$$C(\mathbf{h};1) = P_{11} - P_1^2 \tag{3.11}$$

where $2\gamma(\mathbf{h};1)$ is the variogram at lag **h** for category 1. P_{10} represents the probability of the category being 1 at location **u** and 0 at location $\mathbf{u}+\mathbf{h}$, contributing to the spatial variability. P_{01} represents the probability of the category being 0 at location **u** and 1 at location $\mathbf{u}+\mathbf{h}$, also contributing to the spatial variability. $C(\mathbf{h};1)$ is the covariance at lag **h** for category 1. P_{11} is the probability of both points, separated by the lag **h**, being in category 1 simultaneously. It represents the joint occurrence of category 1 at both locations. P_1, is the probability of the condition (category 1) occurring at any random location. The term P_1^2 is the probability of finding category 1 at both locations independently, assuming the occurrences at the two locations are unrelated (product of individual probabilities).

For the indicator 0, the variogram remains identical to that for indicator 1, and the covariance is given by:

$$C(\mathbf{h};0) = P_{00} - P_0^2 \tag{3.12}$$

this follows the same explanation given for $C(\mathbf{h};1)$ but now for category 0.

The sum of the probabilities for the four possible pairings equals 1:

$$P_{00} + P_{11} + P_{10} + P_{01} = 1 \tag{3.13}$$

We also establish the following relationships:

$$P_{10} + P_{11} = P_1 = P_{01} + P_{11} \tag{3.14}$$

$$P_{10} + P_{00} = P_0 = P_{01} + P_{00} \tag{3.15}$$

These equations are predicated on the assumption of stationarity. Under conditions of stationarity, there is a direct relationship between the variogram and covariance, expressed as $\gamma = \sigma^2 - C$ where σ^2 is the variance. However, this relationship does not hold in practical scenarios involving categorical indicators, as the extent of the domain over which stationarity is assumed remains undefined.

In the case of non-stationarity, the variogram is derived using the same equation as in the stationary case, Equation (3.10). The covariances for indicators 1 and 0 are calculated as follows:

$$C(\mathbf{h};1) = P_{11} - (P_{11} + P_{10})(P_{11} + P_{01}) \tag{3.16}$$

$$C(\mathbf{h};0) = P_{00} - (P_{00} + P_{10})(P_{00} + P_{01}) \tag{3.17}$$

where, $(P_{11} + P_{10})(P_{11} + P_{01})$ represents the non-stationary means for indicator 1 at locations $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$, respectively. This is because these terms sum up the probabilities of finding category 1 at the two respective locations. $(P_{00} + P_{10})(P_{00} + P_{01})$ are the non-stationary means for indicator 0 at locations $\mathbf{u}$ and $\mathbf{u} + \mathbf{h}$, respectively. Expanding both $C(\mathbf{h};1)$ and $C(\mathbf{h};0)$, and applying the relationships established in Equations 3.13, 3.14, and 3.15, we obtain the following results:

$$C(\mathbf{h};1) = P_{11}P_{00} - P_{10}P_{01} \tag{3.18}$$

$$C(\mathbf{h};0) = P_{11}P_{00} - P_{10}P_{01} \tag{3.19}$$

Therefore, if non-stationary mean values are utilized, the relationship $C(\mathbf{h};1) = C(\mathbf{h};0)$ emerges. As these are equivalent, we can consider them simply as $C(\mathbf{h})$. From this, we can rearrange the terms in the covariance equation and isolate both P_{01} and P_{10}:

$$P_{01} = \frac{P_{00}P_{11} - C(\mathbf{h})}{P_{10}} \tag{3.20}$$

$$P_{10} = \frac{P_{00}P_{11} - C(\mathbf{h})}{P_{01}} \tag{3.21}$$

Substituting P_{01} and P_{10} into Equation 3.10, we obtain:

$$2\gamma_I(\mathbf{h}) = \frac{P_{00}P_{11} - C(\mathbf{h})}{P_{10}} + \frac{P_{00}P_{11} - C(\mathbf{h})}{P_{01}} \tag{3.22}$$

This equation establishes a relationship between the variogram and covariance in a non-stationary context where an assumption of stationarity does not hold.

3.3 Cluster of data and variogram model

Calculating a stable experimental variogram for an indicator can be complex, particularly when faced with preferential sampling. It is common for a larger number of samples to be gathered in areas of interest, leading to data clustering. This clustering can cause the short-range variogram values to be non-representative of the overall spatial continuity. Consider the synthetic data case presented in Figure 3.4.

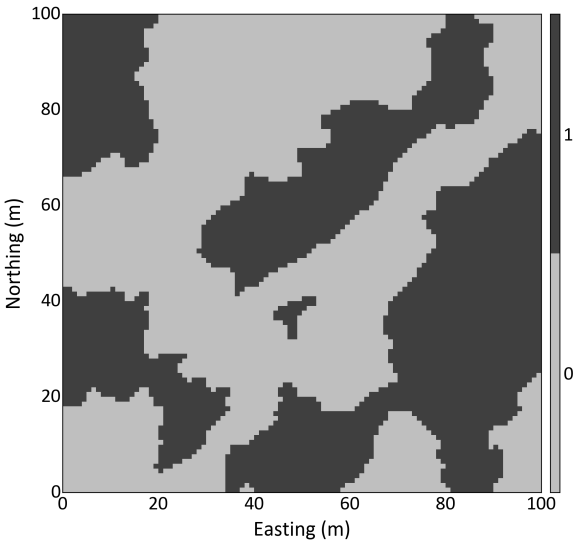


Figure 3.4: Location map of the synthetic data case.

The sampling in this data creates a cluster of samples near the southeastern corner, illustrating the preferential sampling often seen in high-interest areas, as shown in Figure 3.5.

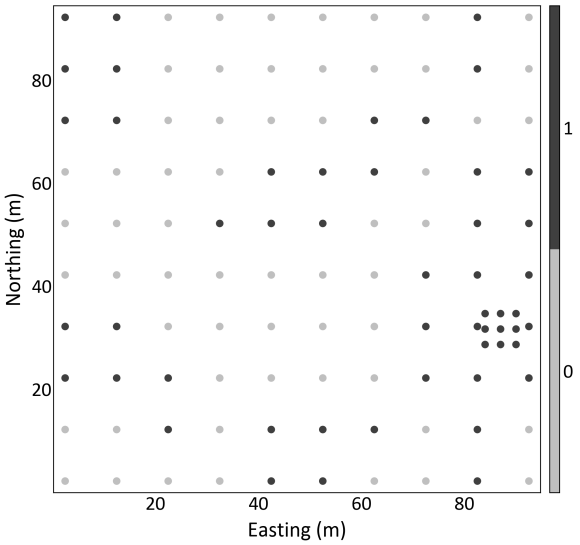


Figure 3.5: Sampling of synthetic data illustrating a cluster of samples.

First, let us verify the equality $C_I(\mathbf{h};1) = C_I(\mathbf{h};0)$ derived in Section 3.2, according to Equa-

tions 3.18 and 3.19. This verification is presented in Figure 3.6.

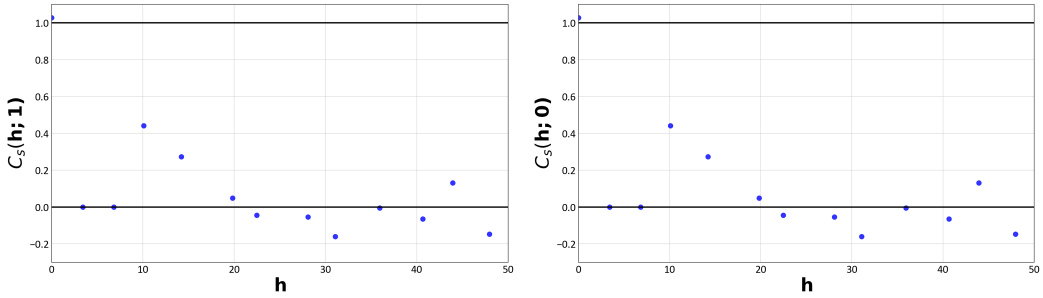


Figure 3.6: Covariance comparison showing $C_I(\mathbf{h};1) = C_I(\mathbf{h};0)$.

Next, the variogram is calculated using the non-stationary relationship deduced in Equation 3.22 and it is shown in Figure 3.7.

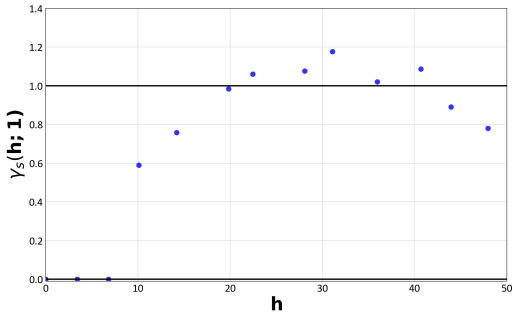


Figure 3.7: Variogram calculated using the non-stationary relationship with covariance.

The non-stationary relationship is verified for category 1. Note the initial variogram points in Figure 3.7, resulting from the sample clustering in the southeastern area. As indicated in Equation 3.10, $2\gamma_I(\mathbf{h};1) = P_{10} + P_{01}$, the variogram at these short lags is zero due to the absence of category transitions. This applies when analyzing the variogram for category 0 as well.

In cases exhibiting this behavior, the modeling should account for a zero nugget effect and a linear increase at the origin, thus disregarding these non-representative points.

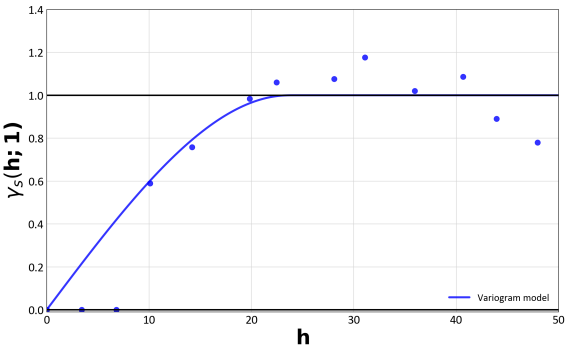


Figure 3.8: Variogram model with a zero nugget effect and linear increase at the origin, excluding non-representative points caused by sample clustering.

The practical approach involves ignoring these non-representative points when fitting a variogram model. The points to be ignored are the ones that in short lags distances are not following the linear increment at the origin, departing from the nugget effect equal to zero. As it is seen in Figure 3.8, the points ignored are the two points between 0 and 10 lag units.

Starting with a flatter slope to match the points between 0 and 10 lag units and then transitioning to a steeper slope might appear to be a better fit; however, this is invalid, as demonstrated in Figure 3.9. The sketch provided illustrates such an invalid variogram model.

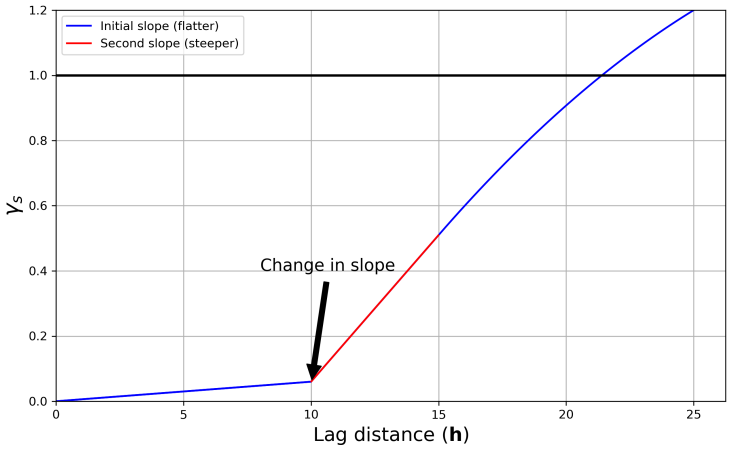


Figure 3.9: Sketch of an invalid variogram fit.

The red line represnts an increase in slope, which is counter-intuitive. The variogram should be steeper at the origin to indicate the possibility of a categorical variable changing abruptly over short distances. Moreover, a slope that becomes steeper away from the origin may violate the Triangular Inequality Principle, as set out by Matheron (1989), and is described in Equation 3.23:

$$\gamma\left(\mathbf{h1}+\mathbf{h2}\right) \leq \gamma\left(\mathbf{h1}\right)+\gamma\left(\mathbf{h2}\right) \tag{3.23}$$

In cases where non-representative points are present, the model should ensure that the initial slope is steeper than any subsequent slopes, and that the nugget effect is set to zero.

3.4 Correlogram

The correlation relative to the lag distance $\mathbf{h}$ is derived by standardizing the covariance $C(\mathbf{h})$. The expression 'one minus the correlation' resembles a variogram and is termed the correlogram (Bai & Deutsch, 2020). The correlogram $\gamma_{\text{Corr}}(\mathbf{h})$ is defined as in Equation 3.24 (C. V. Deutsch & Journel, 1997):

$$\gamma_{\text{Corr}}(\mathbf{h}) = 1 - \frac{C(\mathbf{h})}{\sigma_{I(\mathbf{u})}\sigma_{I(\mathbf{u}+\mathbf{h})}} \tag{3.24}$$

where $\sigma_{I(\mathbf{u})}$ and $\sigma_{I(\mathbf{u}+\mathbf{h})}$ represent the standard deviations of the indicator values at locations $\mathbf{u}$ and $\mathbf{u}+\mathbf{h}$, respectively. The term $\sigma_{I(\mathbf{u})}\sigma_{I(\mathbf{u}+\mathbf{h})}$ is the product of these standard deviations, used to standardize the covariance.

A study was conducted to analyze and compare the behaviors of the variogram and correlogram across various scenarios, with the aim of determining which experimental measure demonstrates more stability. The first scenario is illustrated in Figure 3.10.

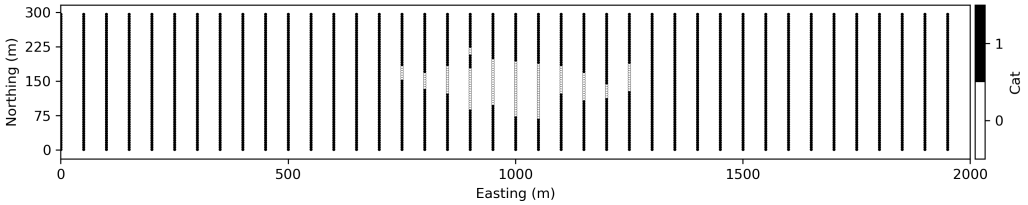


Figure 3.10: First scenario with drillholes every 50m and category 0 concentrated in the center.

This scenario includes drillholes spaced every 50 meters, with the region of interest, category 0, located in the center. The variogram and correlogram for this scenario are shown in Figure 3.11.

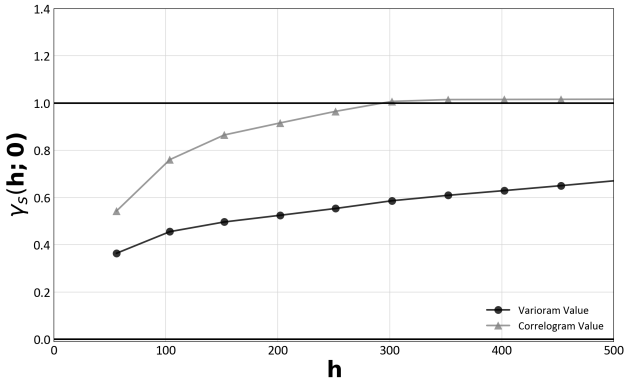


Figure 3.11: Variogram and correlogram for the first scenario.

This setup serves as a benchmark for subsequent scenarios. In the second scenario, a smaller search window is applied, as seen in Figure 3.12. The variogram and correlogram results with this smaller search window are presented in Figure 3.13.

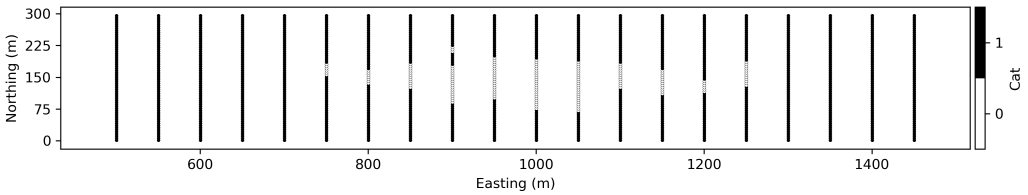


Figure 3.12: Second scenario with a smaller search window.

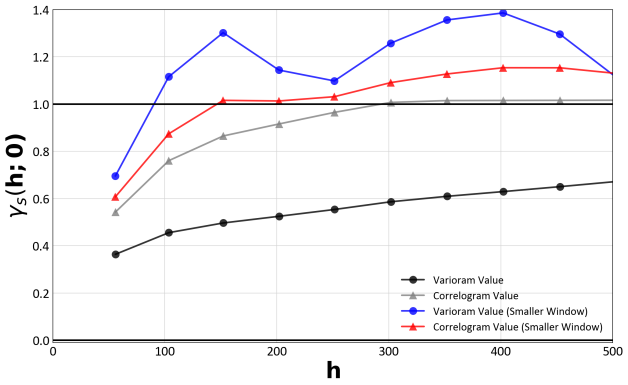


Figure 3.13: Comparative variograms and correlograms of the first and second scenarios.

The correlogram is observed to be more stable than the variogram, showing less fluctuation across both scenarios.

In the third scenario, additional drillholes are placed in the area of interest (category 0), creating a preferential sampling scenario, as depicted in Figure 3.14. The variogram and correlogram for this scenario are displayed in Figure 3.15.

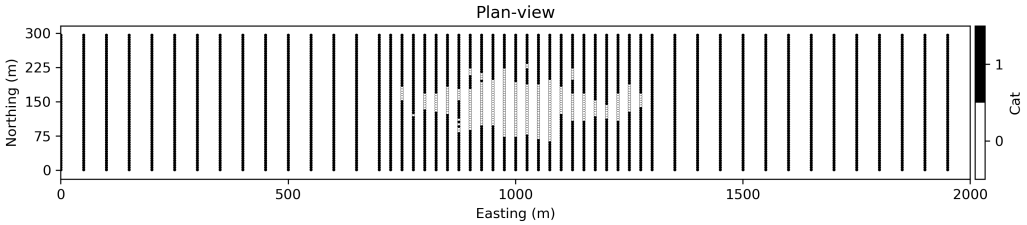


Figure 3.14: Third scenario emphasizing preferential sampling.

Figure 3.15 demonstrates that the correlogram remains more stable even in this third scenario. Therefore, in cases of categorical data with preferential sampling, the correlogram may offer a more robust analytical option as it showed to be more stable than the variogram throughout different scenarios. For instance, in the third scenario, the correlogram was practically identical to that of the first scenario.

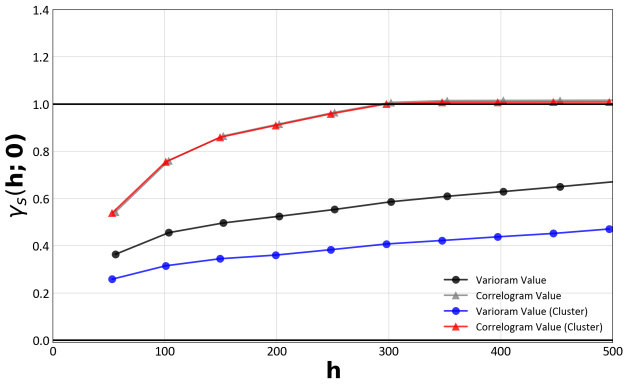


Figure 3.15: Variograms and Correlograms comparing the first and third scenarios.

This chapter explored the characteristics of categorical indicator variograms and covariance. It explained the relationship between variograms and geometry. A review of the variogram's sill and range was presented. A key focus was on demonstrating why the nugget effect for a categorical indicator variogram must be 0 and should have linear increments. The chapter also presented the deduction of a non-stationary relationship between variograms and covariance and provided insights into variogram modeling in the presence of clustered samples. Moreover, the chapter offered a comparative analysis between the variogram and correlogram, assessing their robustness in various scenarios.

Chapter 4

Comparison Between Multiple Point Statistics and Kriging Algorithms

This chapter presents a study comparing multiple point statistics (MPS) conditional probabilities to different kriging algorithms. The kriging methods are Simple Kriging (SK), simple cokriging (SCK), ordinary kriging (OK), and standardized ordinary cokriging (SOCK). The MPS probability results are used as a benchmark for the study as this method considers the most information. The results from the kriging algorithms are compared to the MPS results. The mean squared error (MSE), coefficient of determination (R^2), and correlation (ρ) metrics are used to establish what methodology gets the best estimates. Section 4.1 covers the methodology of an initial study with and the results of the first case. Sections 4.2, and 4.3 present the second and third case with their results for the initial study. Subsequently, the cases are tested for an entire grid. The methodology for the whole grid is presented in Section 4.4, and the cases, including one where the variograms and cross variograms are fitted with the very large linear module of corregionalization method (VLLMC), are shown in Sections 4.5, 4.6, and 4.7. Section 4.8 analyzes the results of OK and SOCK against the reference. Section 4.8 extends the comparison of OK and SOCK estimates against the reference for images from the Data Validation Project by Mokdad, Binakaj, and Boisvert (2022) and summarizes the results.

4.1 Case 1

The study consists in getting the MPS conditional probabilities for one hundred different sample configurations. Subsequently, for the same sample configurations, SK, SCK, OK and SOCK estimates are calculated. The covariances and cross covariances required for the kriging algorithms are calculated directly from the training images (TI). Cross plots comparing the MPS probabilities to each of the kriging algorithms are plotted. MSE and R^2 are calculated for each case. The study is conducted using three different TI's.

Figure 4.1 represents the TI used in the first case, which is obtained from the Center for Computational Geostatistics (CCG) TI libray. Blue represents facies 1, orange indicates facies 2 and green is facies 3.

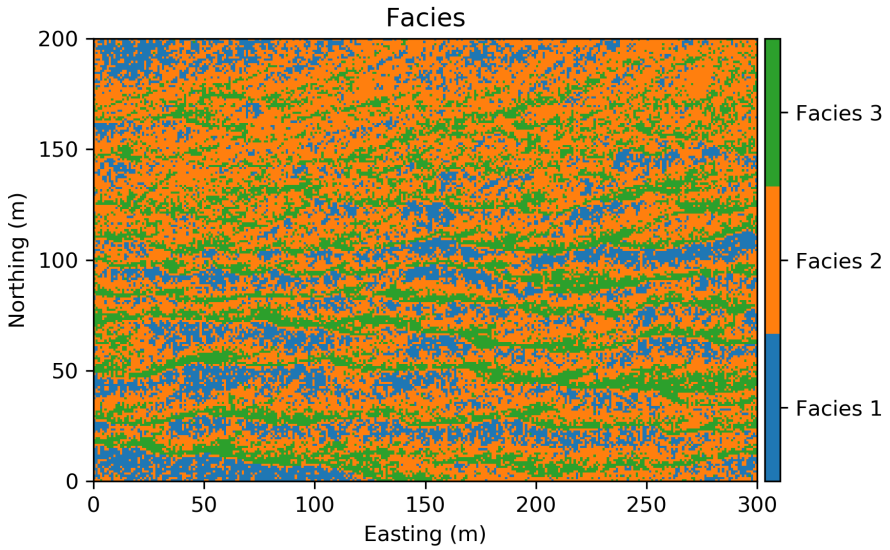
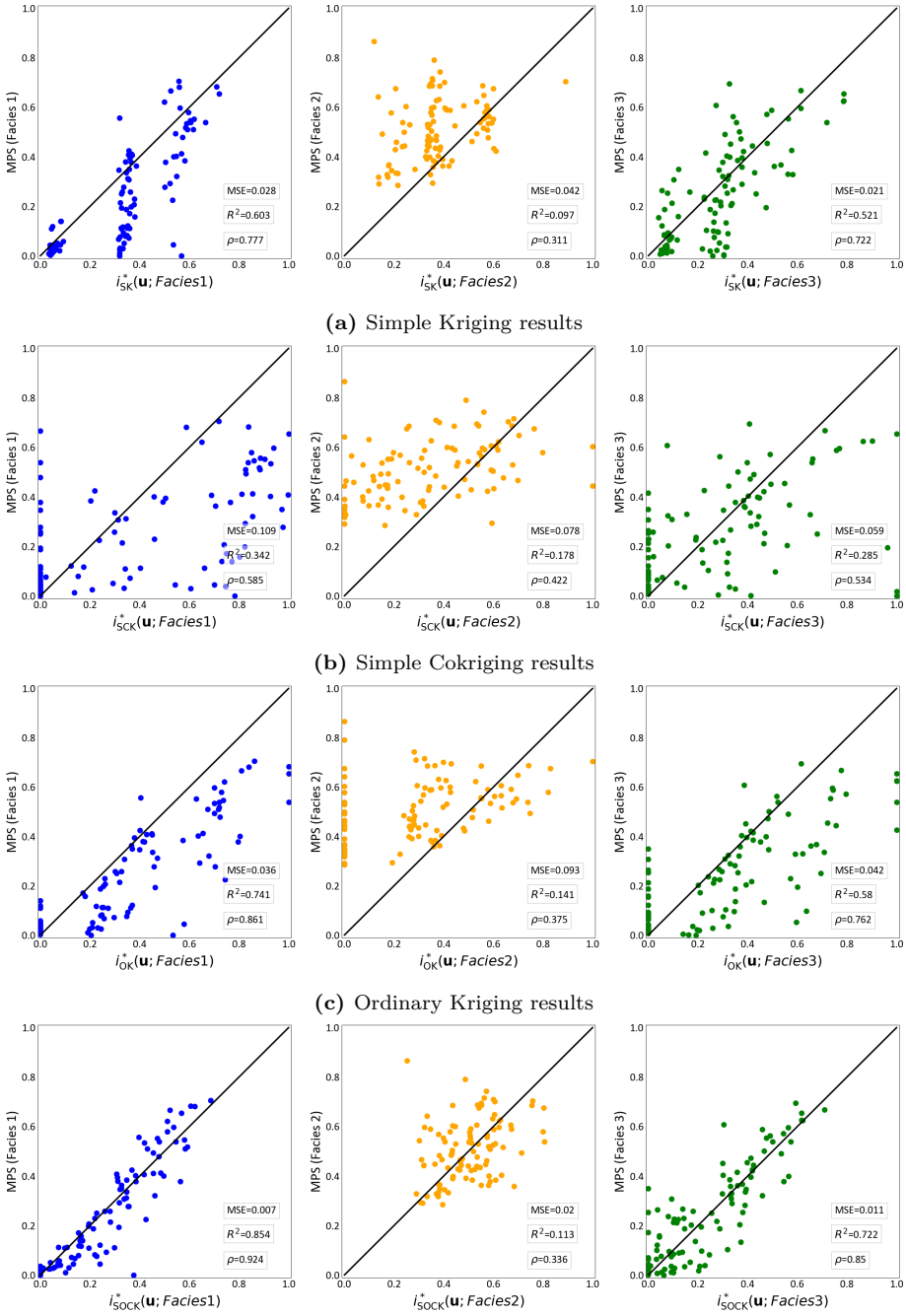


Figure 4.1: Case 1 - TI.

MPS probabilities are calculated and saved for one hundred different sample configurations. The sample configurations are randomly selected from the possible combinations, up to a limit of 50 lag units of distance, without repetition. Covariances and cross covariances are obtained by scanning the TI, and they are used to calculate the weights and estimates for the kriging algorithms. The results are presented in Figure 4.2, which displays cross plots of the kriging estimates against the MPS probabilities for the three categories. The results are also tabulated and shown in Table 4.1. This table summarizes the results obtained for the first case, showing the MSE, R^2, and ρ for each kriging algorithm's estimates compared to the MPS probabilities.



(a) Simple Kriging results

(b) Simple Cokriging results

(c) Ordinary Kriging results

(d) Standardized Ordinary Cokriging results

Figure 4.2: Comparison of Kriging methods in Case 1, showing cross plots versus MPS probabilities across three categories.

Table 4.1: Case 1 - Comparison of MSE, R^2, and ρ for each Kriging algorithm and each category.

Facies 1	SK	SCK	OK	SOCK
MSE	0.028	0.109	0.036	0.007
R^2	0.603	0.342	0.741	0.854
ρ	0.777	0.585	0.861	0.924
Facies 2	SK	SCK	OK	SOCK
MSE	0.042	0.078	0.093	0.020
R^2	0.097	0.178	0.141	0.113
ρ	0.311	0.422	0.375	0.336
Facies 3	SK	SCK	OK	SOCK
MSE	0.021	0.059	0.042	0.011
R^2	0.521	0.285	0.580	0.722
ρ	0.722	0.534	0.762	0.850

Between SK and SCK, the former demonstrates better overall performance. SCK often adjusts many estimates to 0 due to issues with order relations. In comparison, OK performs better overall than both SK and SCK, but it also shows order relation issues. The best overall performance is observed in the SOCK method. Specifically, for facies 1, the R^2 value is 0.854, a significant improvement over the second-best performer, OK, which has an R^2 value of 0.741. The MSE values for the SOCK algorithm are also lower compared to other methods.

4.2 Case 2

In a second case, the same methodology is applied using a TI created by the `FLUVSIM.exe` program, a fluvial object-based algorithm form `GSLIB`. This TI will be referred as TI2 and it is shown in Figure 4.3.

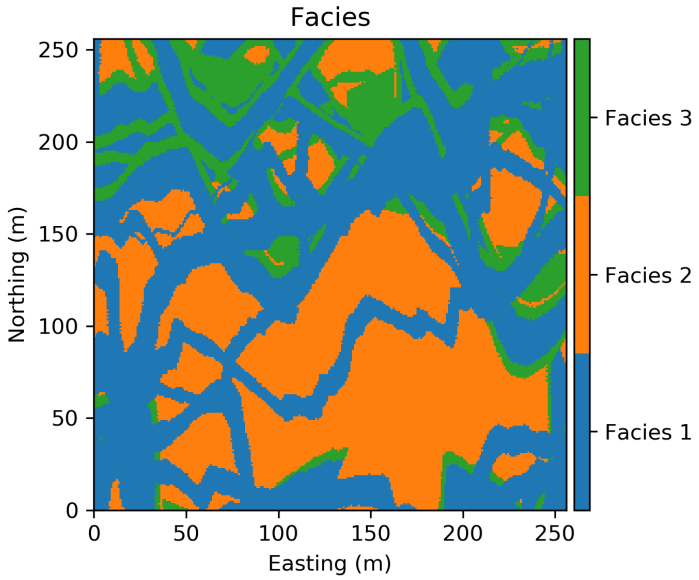
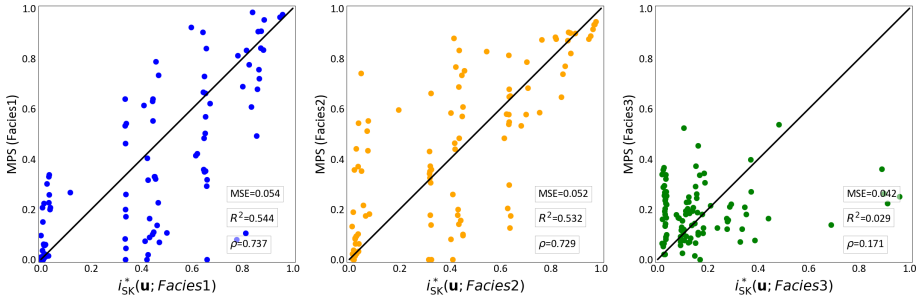
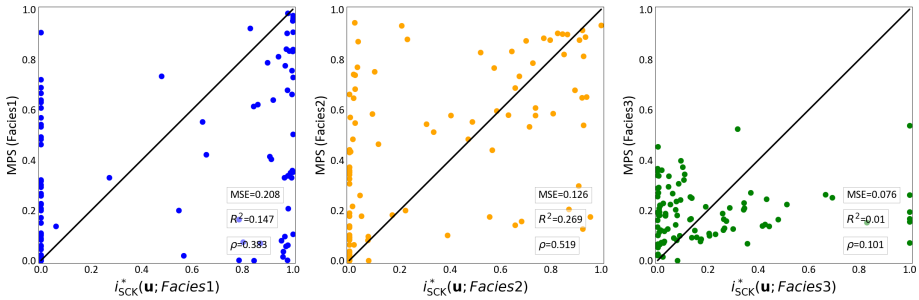


Figure 4.3: Case 2 - TI2.

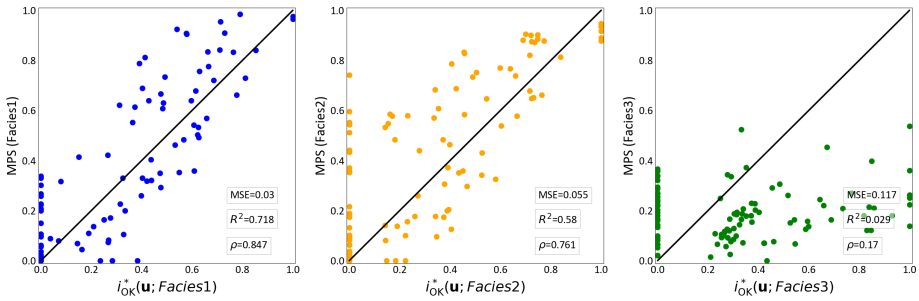
Like the first case, MPS probabilities are computed and saved for one hundred different sample configurations. TI2 is scanned to calculate the covariances and cross covariances that are used to obtain the kriging estimates. The cross plots for each kriging estimate and each category are displayed in Figure 4.4. The results for the second case are tabulated (Table 4.2), summarizing the MSE, R^2, and ρ values for each kriging algorithm's estimates compared to the MPS probabilities.



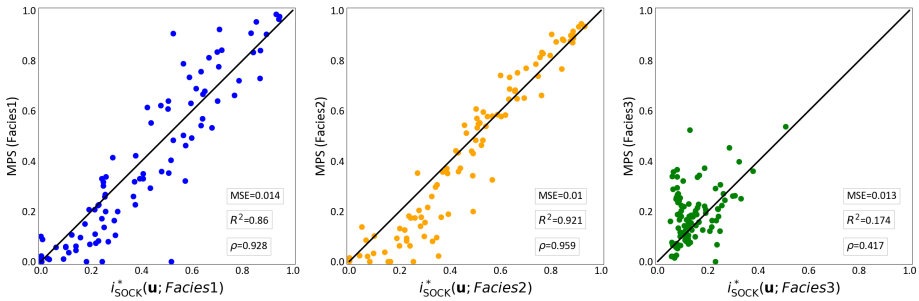
(a) Simple Kriging results for Case 2



(b) Simple Cokriging results for Case 2



(c) Ordinary Kriging results for Case 2



(d) Standardized Ordinary Cokriging results for Case 2

Figure 4.4: Comparison of Kriging methods in Case 2, showing cross plots versus MPS probabilities across three categories.

Table 4.2: Case 2 - Comparison of MSE, R^2, and ρ for each Kriging algorithm and each category.

Facies 1	SK	SCK	OK	SOCK
MSE	0.054	0.208	0.003	0.014
R^2	0.544	0.147	0.718	0.860
ρ	0.737	0.383	0.847	0.928
Facies 2	SK	SCK	OK	SOCK
MSE	0.052	0.126	0.055	0.010
R^2	0.532	0.269	0.58	0.921
ρ	0.729	0.519	0.761	0.959
Facies 3	SK	SCK	OK	SOCK
MSE	0.042	0.076	0.117	0.013
R^2	0.029	0.01	0.029	0.174
ρ	0.171	0.101	0.17	0.417

In this case, similar to the previous one, SK outperforms SCK, though both exhibit weaker performance compared to MPS. SCK continues to face numerous issues with order relations. OK surpasses both SK and SCK but also demonstrates challenges with order relations. Notably, the best performance is observed in the SOCK algorithm. This method shows a significant improvement in the R^2 values, with R^2 for facies 1 and 2 reaching 0.86 and 0.921, respectively. While the results for facies 3 are not as impressive due to its smaller global proportion, an improvement is still evident when utilizing the SOCK algorithm.

4.3 Case 3

A third case is presented to check and validate whether SOCK is getting the best estimates compared to MPS. Figure 4.5 shows the third case TI which will be reffered as TI3 for simplicity.

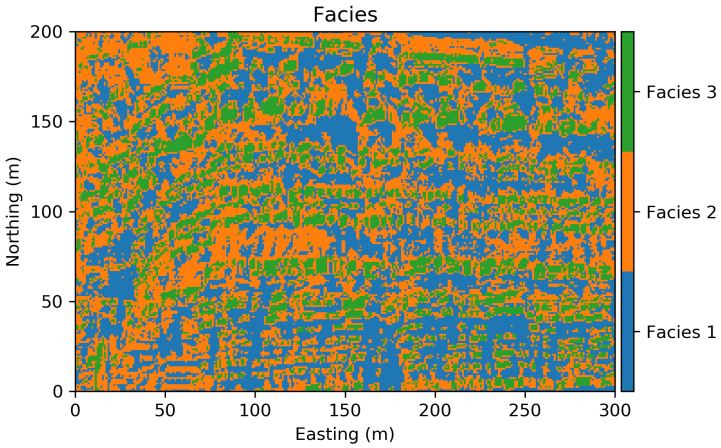
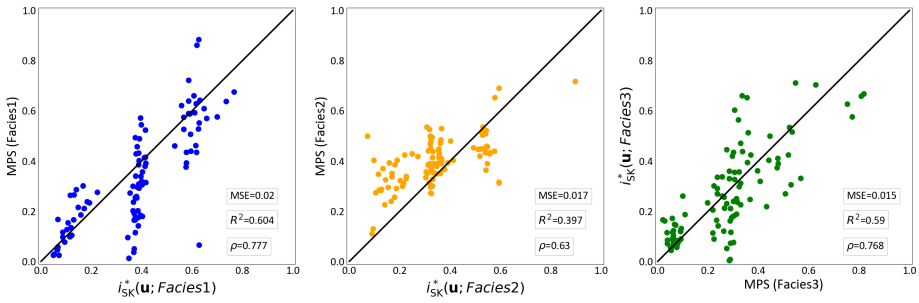


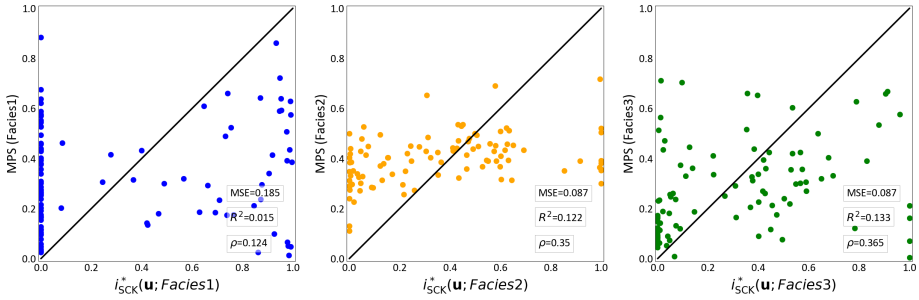
Figure 4.5: Case 3 - TI3.

Like in the previous examples, blue represents facies 1, orange indicates facies 2, and green

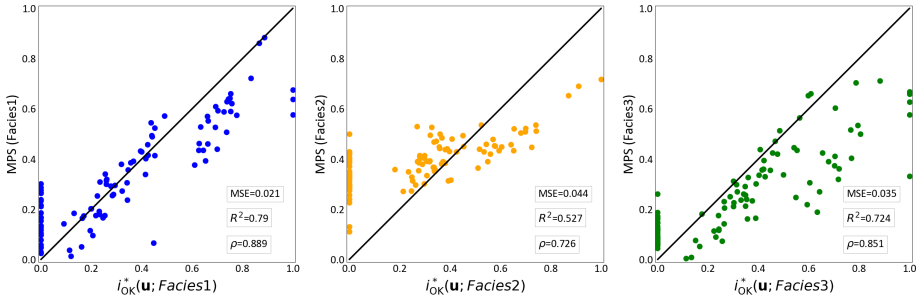
represents facies 3. The image is scanned for a hundred different sample configurations. Sample configurations are randomly chosen from available combinations, with a maximum of 50 lag units of distance and no repetitions. MPS probabilities are calculated and saved. Covariances and cross covariances are also calculated by scanning the image. The cross plots for each kriging method and each category are presented in Figure 4.6. Table 4.3 summarizes the MSE, R^2, and ρ values for each method in this third case.



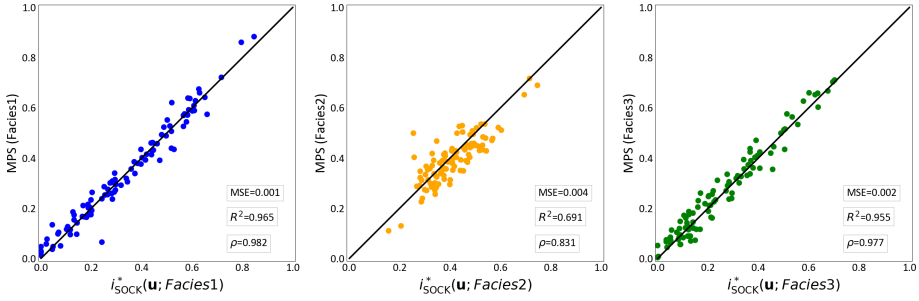
(a) Simple Kriging results for Case 3



(b) Simple Cokriging results for Case 3



(c) Ordinary Kriging results for Case 3



(d) Standardized Ordinary Cokriging results for Case 3

Figure 4.6: Comparison of kriging methods in Case 3, showing cross plots versus MPS probabilities across three categories.

.

Table 4.3: Case 3 - Comparison of MSE, R^2, and ρ for each Kriging algorithm and each category.

Facies 1	SK	SCK	OK	SOCK
MSE	0.020	0.185	0.021	0.001
R^2	0.604	0.015	0.790	0.965
ρ	0.777	0.124	0.889	0.982
Facies 2	SK	SCK	OK	SOCK
MSE	0.017	0.087	0.044	0.004
R^2	0.397	0.122	0.527	0.691
ρ	0.630	0.350	0.726	0.831
Facies 3	SK	SCK	OK	SOCK
MSE	0.015	0.087	0.035	0.002
R^2	0.590	0.133	0.724	0.955
ρ	0.768	0.365	0.851	0.977

As in previous cases, SK shows better results than SCK, yet both fall short in performance compared to MPS. SCK continues to struggle with order relation issues. In contrast, OK surpasses both SK and SCK in performance but is not free from order relation issues. However, the best method is SOCK, which significantly exceeds other approaches, as indicated by its R^2 values for facies 1, 2, and 3 being 0.982, 0.691, and 0.955, respectively. Overall, when assessed against MPS, which is believed to get the most accurate probabilities, SOCK is the kriging algorithm with estimates most closely aligning with those of MPS.

4.4 Grid study

Observing the strong performance of the SOCK method in comparison with MPS, the study was extended to analyze performances across an entire grid. Due to their less effective results, SK and SCK were excluded from this further analysis, focusing only on OK and SOCK. The study will compare each method's estimates against the MPS probabilities. This includes a scenario where the covariances and cross covariances were fitted using the very large linear model of coregionalization (LMC) through the `vl_lmc.exe` program from `GSLIB`. To determine which method performs better, the mean squared error (MSE), R^2, and correlation (ρ) values will be used. Subsequently, the OK and SOCK estimates will be compared to the actual Training Image (TI) to assess their performance. When compared to the TI, the indicator cross validation is used to analyze the performance, and the B value will be the metric used. B represents the difference between the average predicted probability when the true value is 1 and the probability when the true is 0 and it is defined in Equation 4.1.

$$B = \frac{1}{\sum_{n=1}^{N} n_{i_k=1}} \sum_{n=1}^{N} \sum_{k=1}^{K} p_{k,i_k=1} - \frac{1}{\sum_{n=1}^{N} n_{i_k=0}} \sum_{n=1}^{N} \sum_{k=1}^{K} p_{k,i_k=0} \tag{4.1}$$

Where K represents the total number of categories, while N denotes the number of locations. The term n_{i_k} refers to the specific locations where the indicator i_k holds a value of either 1 or 0. The probability of category k being present when i_k is 1 or 0 is denoted as p_k, i_k. Higher B values indicate more accurate predictions regarding the presence or absence of a category (C. V. Deutsch, 2010; J. L. Deutsch & Deutsch, 2012).

4.5 Case 1 - Grid Study

Figure 4.7 shows the samples, represented as dots, used for calculating the estimates and the filled portion represents the estimation area. Note that a buffer is left to avoid edge effects.

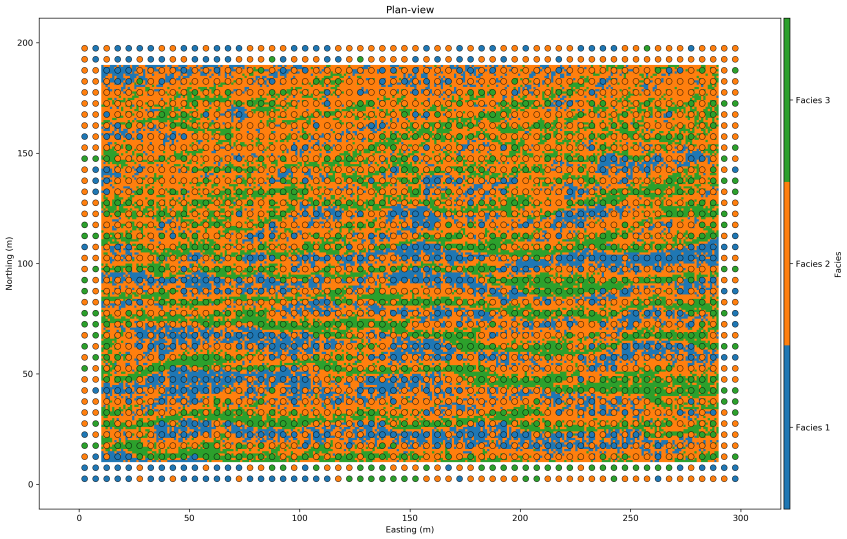
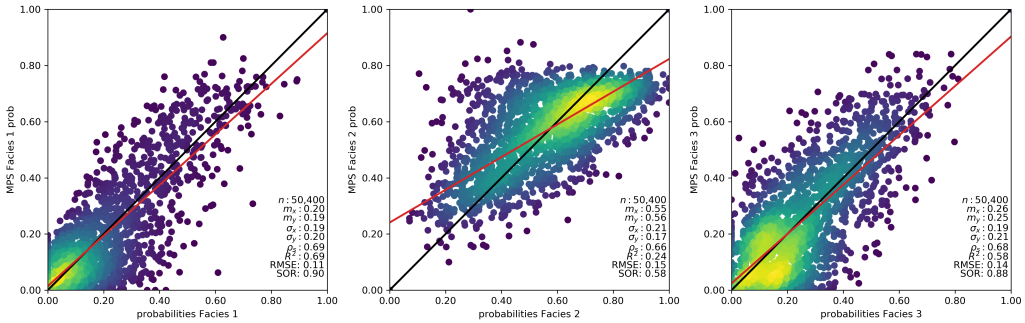
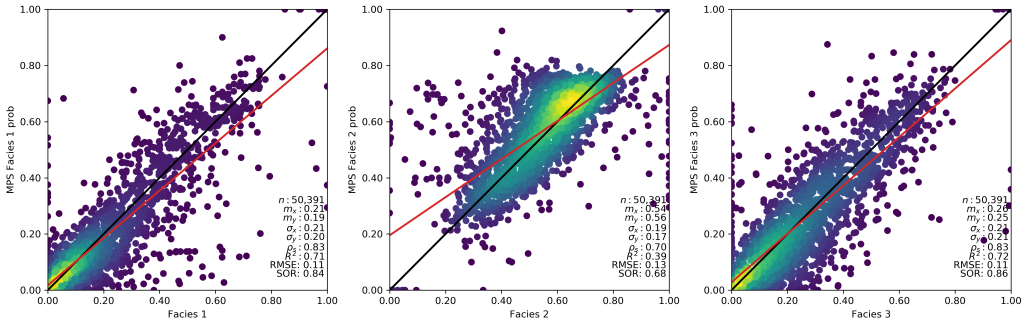


Figure 4.7: Case 1 grid study - Samples represented by dots and estimation grid represented by the filled portion of the TI.

The image is scanned and the MPS probabilities are calculated and saved, the covariances and cross covariances needed for OK and SOCK are also calculated when scanning the image. The results for each of the methods and categories is presented in Figure 4.8. Table 4.4 summarizes the MSE, R^2, and ρ values of the estimates obtained for OK and SOCK compared to the MPS probabilities.



(a) Ordinary Kriging results for Case 1 grid study



(b) Standardized Ordinary Cokriging results for Case 1 grid study

Figure 4.8: Comparison of OK and SOCK methods in Case 1 grid study, showing cross-validation plots versus MPS probabilities across three categories.

Table 4.4: Comparison of MSE, R^2, and ρ for each kriging algorithm and each category.

Facies 1	OK	SOCK
MSE	0.012	0.012
R^2	0.690	0.710
ρ	0.690	0.830
Facies 2	**OK**	**SOCK**
MSE	0.022	0.017
R^2	0.240	0.390
ρ	0.660	0.700
Facies 3	**OK**	**SOCK**
MSE	0.020	0.012
R^2	0.580	0.720
ρ	0.680	0.830

Compared to MPS probabilities, SOCK estimates outperform those of OK. The MSE and R^2 values show that SOCK estimates are closer to the MPS probabilities. A second case is presented.

4.6 Case 2 - Grid Study

Figure 4.9 displays the dots representing the samples used for generating the estimates, with the shaded area indicating the region of estimation for this second case. A boundary buffer is maintained around the area to mitigate edge effects.

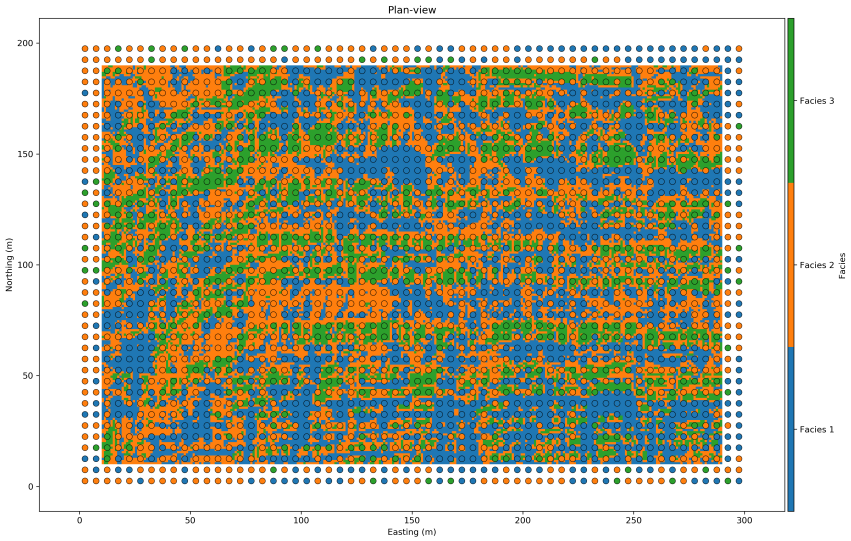
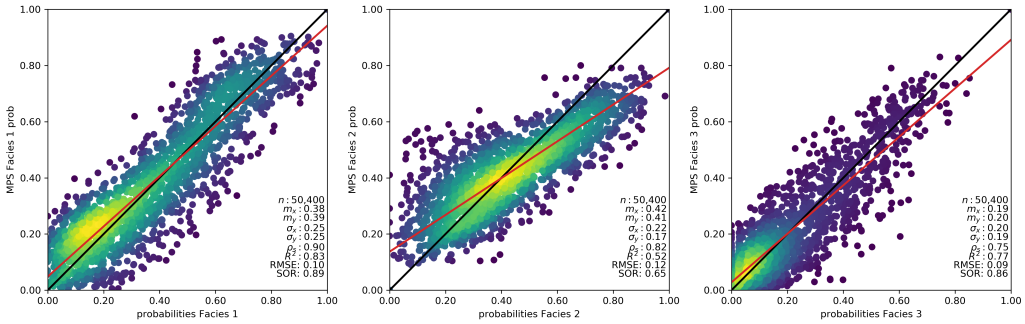
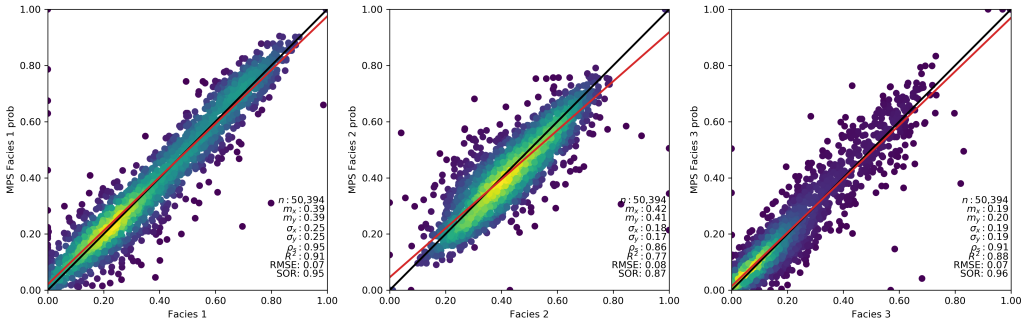


Figure 4.9: Case 2 grid study - Samples represented by dots and estimation grid represented by the filled portion of the TI.

The image for Case 2 is scanned and the MPS probabilities are computed and stored. At the same time, the necessary covariances and cross-covariances for OK and SOCK are derived from the same scanning procedure. The cross-validation plots for each method and category are presented in Figure 4.10. Table 4.4 is a summary of MSE, R^2, and ρ metrics, detailing the performance of OK and SOCK relative to the MPS outcomes for Case 2.



(a) Ordinary Kriging results for Case 2 grid study



(b) Standardized Ordinary Cokriging results for Case 2 grid study

Figure 4.10: Comparison of OK and SOCK methods in Case 2 grid study, showing cross-validation plots versus MPS probabilities across three categories.

Table 4.5: Case 2 - Comparison of MSE, R^2, and ρ for each kriging algorithm and each category

Facies 1	OK	SOCK
MSE	0.010	0.005
R^2	0.830	0.910
ρ	0.900	0.950
Facies 2	**OK**	**SOCK**
MSE	0.014	0.006
R^2	0.520	0.770
ρ	0.820	0.860
Facies 3	**OK**	**SOCK**
MSE	0.008	0.005
R^2	0.770	0.880
ρ	0.750	0.960

In Case 2, in relation to MPS probabilities, SOCK estimates outperform the ones from OK, as indicated by the lower MSE and higher R^2 values, highlighting greater alignment with MPS probabilities. These two cases show that when compared to MPS, SOCK outperforms OK. The next step is to fit the variograms and cross variograms with the very large linear model of corregionalization.

4.7 Very Large Linear Model of Coregionalization (VL-LMC) to fit variograms and cross variograms

Despite having modern approaches for decorrelation or even intrinsic models of coregionalization, the LMC continues to be useful and mathematically tractable. However, the challenge with the LMC lies in modeling K categories, as it requires fitting $K \cdot (K-1)/2$ cross variograms and K variogrmas in a way that ensures joint positive definiteness. The difficulty in fitting is aggravated when only one to four nested structures are used (R. Barnett & Deutsch, 2018). Those reasons motivated the authors R. Barnett and Deutsch (2018) to propose the very large LMC (VL-LMC). This approach starts by independently fitting the K direct variograms, with no constraints on their shape, direction and range. The LMC is then fit using the combination of all the nested structure found in these K variograms models. For instance, if each variogram model are done with three nested structures, the LMC employs $K \cdot 3$ nested structures. This method ensure a reasonable fit for the direct variograms and enhances the fitting of cross variograms, as the extra nested structures offer increased flexibility in the model (R. Barnett & Deutsch, 2018). This method was used to fit the variograms and cross variograms of the first case, represented by Figure 4.1, and the fit is shown in Figure 4.11.

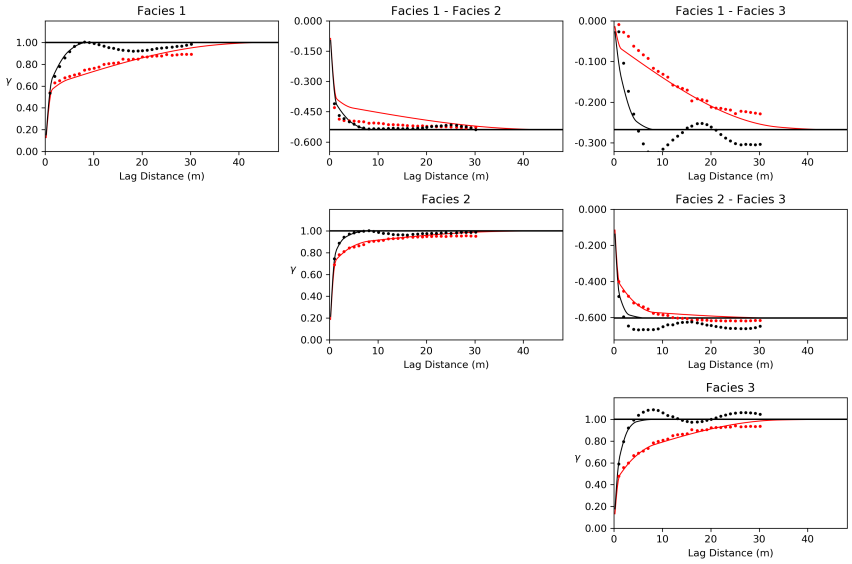
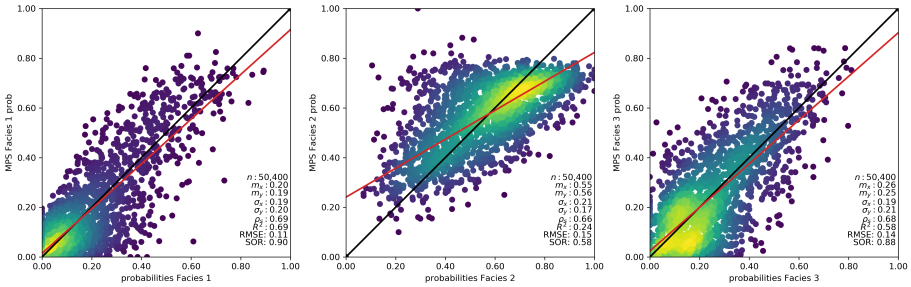


Figure 4.11: Variograms and cross variograms of first case fitted by the VL-LMC method.

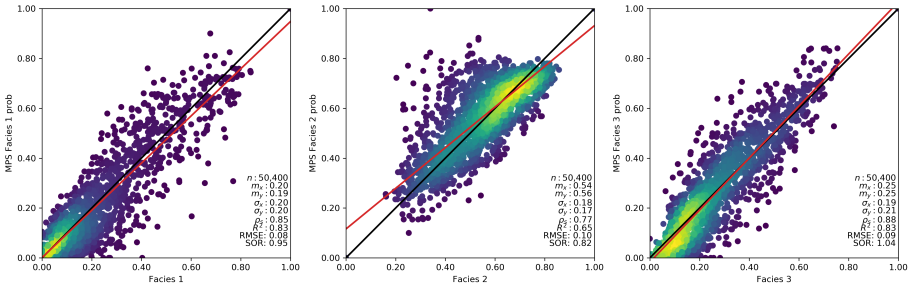
In Figure 4.11 the diagonal plots represent the direct variograms models while the off-diagonal plots are the cross variograms models. Red indicates the major direction while black is the minor direction.

Like the previous cases, OK and SOCK estimates are calculated and compared to the MPS probabilities. The goal here is to evaluate if by fitting an LMC the results of SOCK will still outperform the ones from OK when compared to MPS. This assessment is important, as it is challenging to obtain a representative TI from which to directly extract the necessary covariances

and cross covariances. Figure 4.12 display the cross-validation plots of the methods' estimates compared to MPS probabilities. Table 4.6 summarizes the results of OK and SOCK when fitted with the VL-LMC approach.



(a) Case 1 - OK results.



(b) Case 1 - SOCK results with variograms and cross variograms fitted with the VL-LMC method.

Figure 4.12: Comparison of kriging methods in Case 1, showing Ordinary Kriging and Standardized Ordinary Cokriging results with variogram fittings.

Table 4.6: Case 1 - Comparison of MSE, R^2, and ρ for OK, SOCK (TI based) and for when variograms and cross variograms are fitted with VL-LMC for the SOCK algorithm

Facies 1	OK	SOCK (TI based)	SOCK (Fitted with VL-LMC)
MSE	0.010	0.012	0.006
R^2	0.830	0.710	0.830
ρ	0.900	0.830	0.850
Facies 2	**OK**	**SOCK (TI based)**	**SOCK (Fitted with VL-LMC)**
MSE	0.014	0.017	0.010
R^2	0.520	0.390	0.650
ρ	0.820	0.700	0.770
Facies 3	**OK**	**SOCK (TI based)**	**SOCK (Fitted with VL-LMC)**
MSE	0.008	0.012	0.008
R^2	0.770	0.720	0.830
ρ	0.750	0.830	0.880

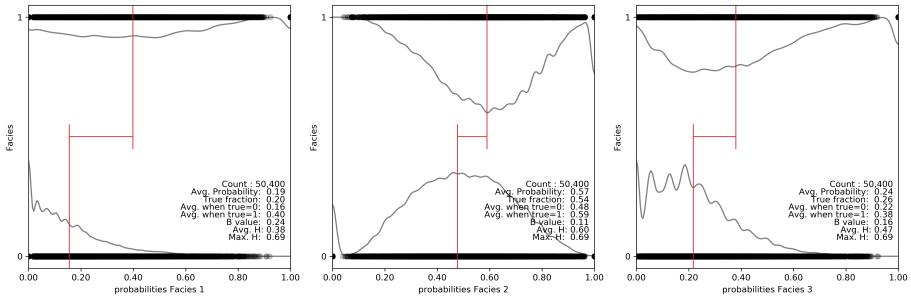
Note that overall the SOCK method still performs better than OK, as demonstrated by the higher R^2 and lower MSE values. Next, the study involves validating the methods by comparing their estimates to the actual training image, which serves as the reference.

4.8 Comparison of OK and SOCK against the reference

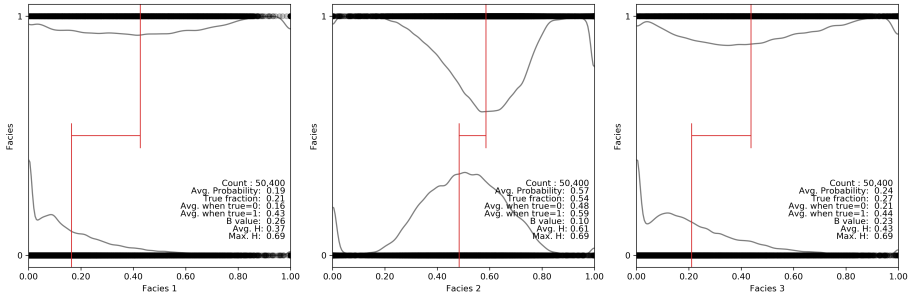
This section will demonstrate how the OK and SOCK estimates compared against the training image itself. Indicator cross validations are plotted and the B metric (Equation 4.1) is analyzed to check what method is performing better. The results are presented.

4.8.1 Case 1

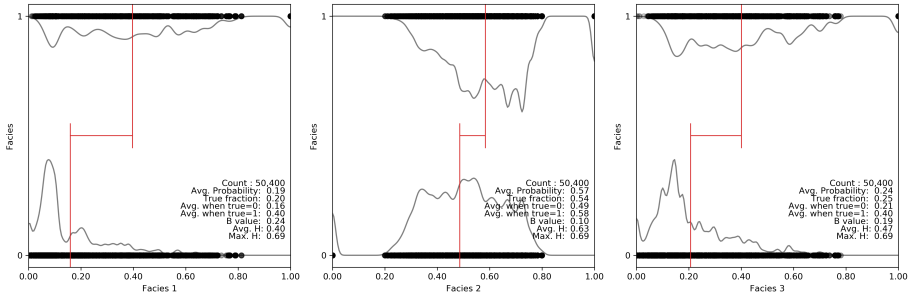
Let us first analyze the results for the first case represented by Figure 4.1. The indicator cross validation plot for the OK and SOCK estimates against the reference is shown in Figure 4.13. Table 4.7 tabulates the B values results for the methods.



(a) Case 1 - OK comparison against reference.



(b) Case 1 - SOCK comparison against reference.



(c) Case 1 - SOCK fitted with VL-LMC comparison against reference.

Figure 4.13: Comparative analysis of advanced kriging methods for Case 1, including OK, SOCK, and SOCK fitted with VL-LMC against the reference dataset.

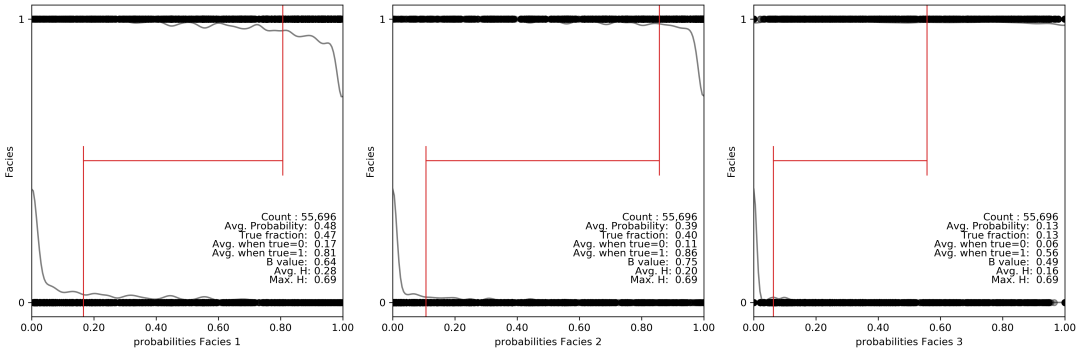
Table 4.7: Case 1 against reference - Comparison of B values for each method and category

Facies 1	OK	SOCK (TI based)	SOCK (Fitted with VL-LMC)
B	0.24	0.26	0.24
Facies 2	OK	SOCK (TI based)	SOCK (Fitted with VL-LMC)
B	0.11	0.10	0.10
Facies 3	OK	SOCK (TI based)	SOCK (Fitted with VL-LMC)
B	0.16	0.23	0.19

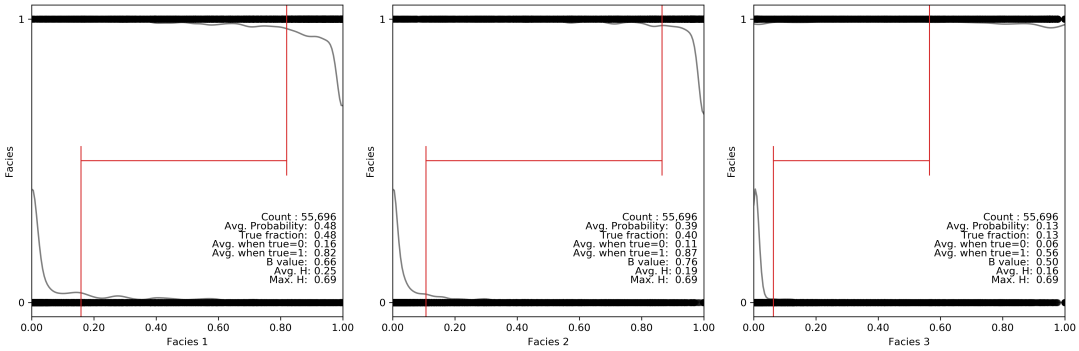
Note that in this case the B value is overall better for the SOCK estimates, indicating that for the first case the SOCK method outperforms OK when compared against the reference. The SOCK method, when fitted with the VL-LMC, also slightly outperforms OK.

4.8.2 Case 2

The second case training image is represented by Figure 4.3. The cross validation plot comparing OK and SOCK estimates with the actual data (refrence) is presented in Figure 4.14. Table 4.8 summarizes the B values for each method and category.



(a) Case 2 - OK comparison against reference.



(b) Case 2 - SOCK comparison against reference.

Figure 4.14: Comparative analysis of kriging methods for Case 2, showing the indicator cross validation plots for OK and SOCK estimates against the reference.

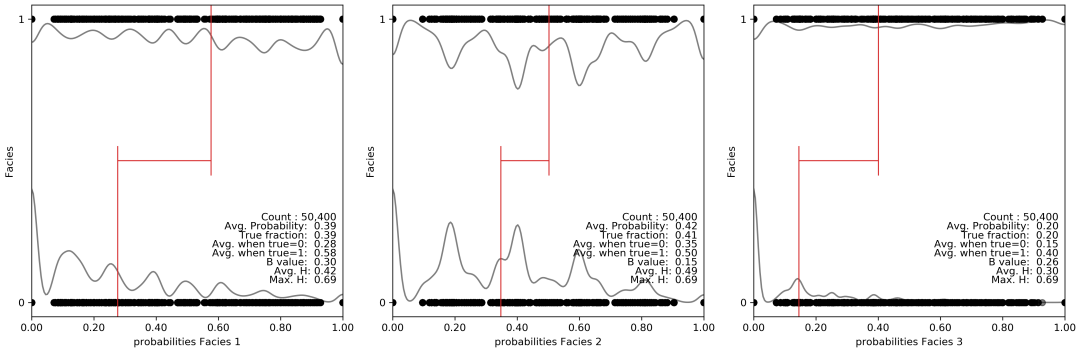
Table 4.8: Case 2 against reference - Comparison of B values for each method and category

Facies 1	OK	SOCK
B	0.64	0.66
Facies 2	OK	SOCK
B	0.75	0.76
Facies 3	OK	SOCK
B	0.49	0.50

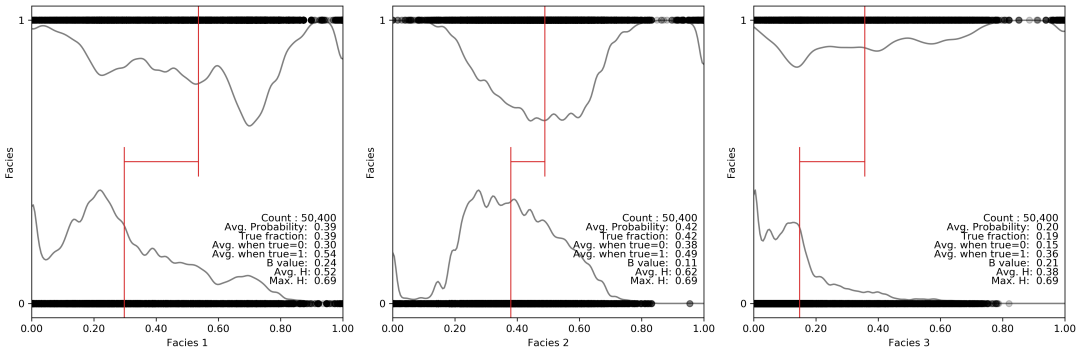
Again, the B value is slightly better for the SOCK estimates, showing that for this case the SOCK method slightly outperforms OK when compared against the reference.

4.8.3 Case 3

For the third case, the image used is represented by Figure 4.5. The cross validation plots showing OK and SOCK estimates against the reference is presented in Figure 4.15. Table 4.9 summarizes the B values for each method and category.



(a) Case 3 - OK comparison against reference.



(b) Case 3 - SOCK comparison against reference.

Figure 4.15: Comparative analysis of kriging methods for Case 3, showing the indicator cross validation plots for OK and SOCK estimates against the reference.

Table 4.9: Case 3 against reference - Comparison of B values for each method and category

Facies 1	OK	SOCK
B	0.30	0.24
Facies 2	OK	SOCK
B	0.15	0.11
Facies 3	OK	SOCK
B	0.26	0.21

Note that in this particular case, the B values overall indicate better performance for the OK estimates compared to SOCK. Consequently, while OK outperforms SOCK in the third case, performance results vary between the two methods across different scenarios. To further explore these variations, this study extends to include an analysis of all the Data Validation project images from `CCG` (Mokdad et al., 2022). The 114 images from the project are transformed into categorical variables (three categories). After conducting OK and SOCK estimations for each image, the differences in root mean squared error (RMSE) are assessed. A summary of the difference in RMSE presented in Figure 4.16

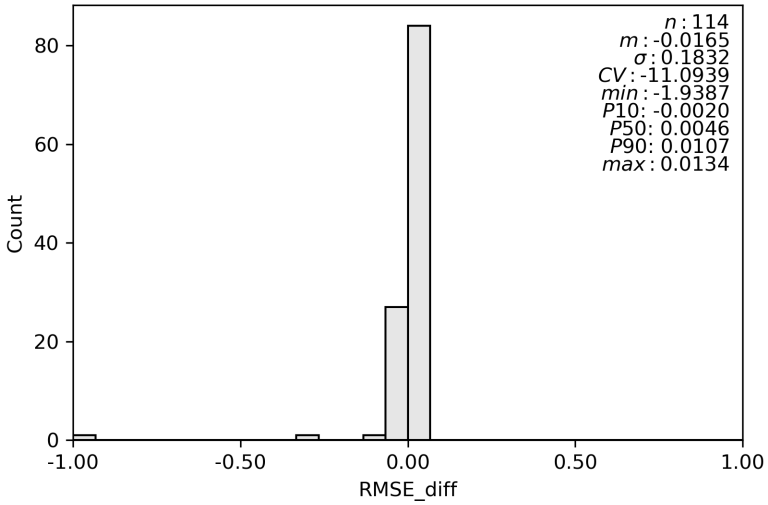


Figure 4.16: Results of the difference in RMSE between OK and SOCK.

Note that for over 80 cases there is a small advantage in using the SOCK method. Although being a small advantage, for those cases SOCK outperforms OK. The four best cases are shown in Figure 4.17 while the four worst cases are presented in Figure 4.18.

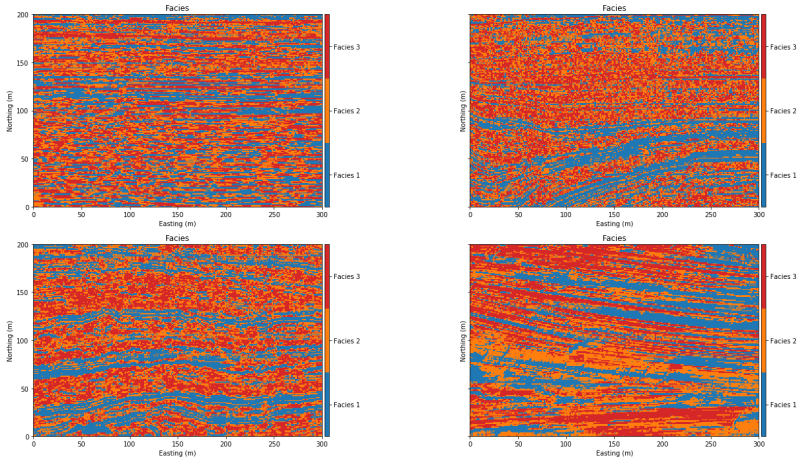


Figure 4.17: The four best cases when using the Data Validation project images.

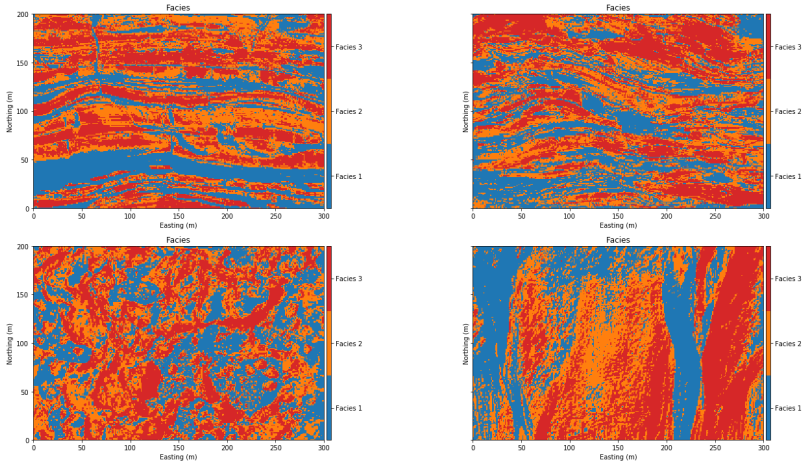


Figure 4.18: The four worst cases when using the Data Validation project images.

Note that SOCK tends to perform better when the short scale variability is high, like the images in Figure 4.17 shows. OK estimates get better results when there is more structure as shown in Figure 4.18.

4.9 Chapter Summary

This chapter explored a study comparing multiple point statistics (MPS) conditional probabilities to various kriging algorithms, including Simple Kriging (SK), simple cokriging (SCK), ordinary kriging (OK), and standardized ordinary cokriging (SOCK). MPS probabilities served as a benchmark, given their presumed accuracy in estimating better results. The performance of each kriging method was evaluated against these MPS results using mean squared error (MSE), R^2, and correlation (ρ) metrics to determine which methodology yielded the most accurate estimates. The chapter described the methodology and presented findings across three distinct cases (Sections 4.1, 4.2, and 4.3), followed by an analysis of the entire grid (Section 4.4), including one case using the very large

linear model of coregionalization method (VL-LMC). In Section 4.8, a critical assessment of OK and SOCK methods against the actual training images was undertaken, extending further to include all images from the Data Validation Project by Mokdad et al. (2022). A summary of the chapter results is presented:

- SK and SCK do not show good results due to issues with order relations.
- SOCK shows good result when applied to small examples (Sections 4.1, 4.2, 4.3).
- SOCK with VL-LMC performs almost as well as image-based methods, as shown in Section 4.8.1.
- When analyzing images from the Data Validation Project, OK performs about as well as SOCK. The difference in RMSE gives a very small advantage to SOCK, but it is not significant.

Chapter 5

A Regularization Technique to Mitigate Extreme Weights

This chapter presents a method to stabilize weights from unstable linear systems of kriging equations. In geostatistics, linear systems of equations emerge when applying kriging and its variants. These systems arise during simulation or when estimating geological variables at unknown grid locations. As a result, these equations can appear thousands to millions of times, depending on the number of grid nodes (Manchuk & Deutsch, 2008). A key requirement for these systems is positive definiteness, which is not always guaranteed as the systems can be indefinite or ill-conditioned. The presence of an indefinite system, indicated by any negative eigenvalues, leads to unacceptable results and extreme weight values (Manchuk & Deutsch, 2008). Chilès and Delfiner (2012) note that weights exceeding one can produce distorted estimates. In this chapter, we analyze the impact of system regularization in standardized ordinary cokriging (SOCK) for categorical variables to mitigate the risk of extreme weights. This regularization involves inflating the left-hand side (LHS) matrix diagonal to ensure it is positive definite and to correct weight anomalies. When performing cokriging, matrix sizes can increase rapidly; for instance, in the presence of three different categories and five samples to estimate at an unsampled location, the LHS matrix will have a size of 15x15. Section 5.1 explains the weights regularization technique. Section 5.2 shows a case in which the regularization technique is applied. Section 5.3 assesses the impact this method has on an entire grid estimation.

5.1 Weights Regularization in SOCK

The regularization technique consists of inflating the diagonal elements of the LHS covariance matrix of the system of equations. In matrix format, the SOCK system of equations are expressed as Equation 5.1.

$$\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} & \mathbf{1} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \cdots & \mathbf{C}_{2n} & \mathbf{1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}_{nn} & \mathbf{1} \\ \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} \\ \boldsymbol{\lambda}_2 \\ \vdots \\ \boldsymbol{\lambda}_n \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \\ 1 \end{bmatrix} \tag{5.1}$$

Where $\mathbf{C}_{ij}$ is the covariance matrix between the known data points; λ_i is a vector of weights to be solved for; $\mathbf{c}_i$ is the vector containing covariances between the known data points and the unknown

location, and μ is the Lagrange multiplier.

To implement the regularization, each diagonal element ($\mathbf{C}_{ii}$) of the LHS covariance matrix is increased by ϵ, a small positive constant. This is represented in Equation 5.2.

$$\mathbf{C}'_{ii} = \mathbf{C}_{ii} + \epsilon\mathbf{I} \tag{5.2}$$

Where $\mathbf{C}'_{ii}$ is the regularized diagonal elements of the LHS covariance matrix; ϵ is a small positive constant, and $\mathbf{I}$ is the identity matrix of the same dimension as $\mathbf{C}$. Thus, the regularized SOCK system is expressed in matrix format as shown in Equation 5.3.

$$\begin{bmatrix} \mathbf{C}'_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1n} & \mathbf{1} \\ \mathbf{C}_{21} & \mathbf{C}'_{22} & \cdots & \mathbf{C}_{2n} & \mathbf{1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}_{n1} & \mathbf{C}_{n2} & \cdots & \mathbf{C}'_{nn} & \mathbf{1} \\ \mathbf{1}^T & \mathbf{1}^T & \cdots & \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda_2} \\ \vdots \\ \boldsymbol{\lambda}_n \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \vdots \\ \mathbf{c}_n \\ 1 \end{bmatrix} \tag{5.3}$$

In the standardized ordinary cokriging (SOCK) model, it is required that all primary and secondary data weights sum to one. Mathematically, this is expressed as:

$$\sum_{k=1}^{K}\sum_{\alpha=1}^{n} \lambda_\alpha^{\text{SOCK}}\left(\mathbf{u}_0; k_0; k\right) = 1 \tag{5.4}$$

where k_0 is the category being estimated, k represents the indicator of the data, and $\mathbf{u}_0$ is the unknown location.

As part of regularization, these weights are adjusted such that as the regularization parameter ϵ increases, each weight λ_i approaches an equal share, represented by $\frac{1}{n}$ where n is the total number of weights.

To demonstrate convergence to $\frac{1}{n}$, consider the function $h(\lambda_1, \ldots, \lambda_n, \mu)$ which is formulated to include a Lagrange multiplier μ to enforce the constraint that the sum of the weights equals one:

$$h\left(\lambda_1, \ldots, \lambda_n, \mu\right) = \sum_{i=1}^{n} \lambda_i^2 + \mu\left(1 - \sum_{i=1}^{n} \lambda_i\right) \tag{5.5}$$

The term $\sum_{i=1}^{n} \lambda_i^2$ is the objective function being minimized. Minimizing this term tends to push the weights λ_i towards zero, as the square of a smaller numbers is even smaller, promoting a distribution of weights that avoids extreme values. The term $\mu\left(1 - \sum_{i=1}^{n} \lambda_i\right)$ is where the Lagrange multiplier μ appears. Here, μ is a scalar that adjusts to satisfy the constraint $\sum_{i=1}^{n} \lambda_i = 1$. The expression $1 - \sum_{i=1}^{n} \lambda_i$ represents the deviation from the constraint (it should be zero if the constraint is met).

When optimizing h, values for λ_i and μ are found such that h is minimized. To find the minimum, the partial derivatives of h with respect to each weight λ_i and the Lagrange multiplier μ

are computed. The partial derivative with respect to each λ_i:

$$\begin{aligned} \frac{\partial h}{\partial \lambda_i} &= 2\lambda_i - \mu = 0 \quad \text{for } i = 1, \ldots, n \\ \lambda_i &= \frac{\mu}{2} \quad \forall i \end{aligned} \tag{5.6}$$

The partial derivative with respect to the Lagrange multiplier μ:

$$\frac{\partial h}{\partial \mu} = 1 - \sum_{i=1}^{n} \lambda_i = 0 \tag{5.7}$$

Since all λ_i are equal to $\frac{\mu}{2}$:

$$1 - \sum_{i=1}^{n} \frac{\mu}{2} = 0 \tag{5.8}$$

Now, summing up $\lambda_i = \frac{\mu}{2}$ for all i:

$$\begin{aligned} \sum_{i=1}^{n} \frac{\mu}{2} &= 1 \\ \frac{n\mu}{2} &= 1 \\ \mu &= \frac{2}{n} \end{aligned} \tag{5.9}$$

Substituting back, we find that each weight:

$$\lambda_i = \frac{\mu}{2} = \frac{1}{n} \quad \forall i \tag{5.10}$$

The demonstration presented above does not incorporate the regularization constant ϵ. The regularization parameter can be included for the quadratic term $\sum_{i=1}^{n} \lambda_i^2$ as follows:

$$h\left(\lambda_1, \ldots, \lambda_n, \mu\right) = \epsilon \sum_{i=1}^{n} \lambda_i^2 + \mu \left(1 - \sum_{i=1}^{n} \lambda_i\right) \tag{5.11}$$

With the incorporation of ϵ, a similar derivation process with respect to λ_i follows:

$$\begin{aligned} \frac{\partial h}{\partial \lambda_i} &= 2\epsilon\lambda_i - \mu = 0 \quad \text{for } i = 1, \ldots, n \\ \lambda_i &= \frac{\mu}{2\epsilon} \quad \forall i \end{aligned} \tag{5.12}$$

And with respect to μ:

$$\frac{\partial h}{\partial \mu} = 1 - \sum_{i=1}^{n} \lambda_i = 0 \tag{5.13}$$

Since now all λ_i are equal to $\frac{\mu}{2\epsilon}$:

$$1 - \sum_{i=1}^{n} \frac{\mu}{2\epsilon} = 0 \tag{5.14}$$

Now, summing up $\lambda_i = \frac{\mu}{2\epsilon}$ for all i:

$$\begin{aligned} \sum_{i=1}^{n} \frac{\mu}{2\epsilon} &= 1 \\ \frac{n\mu}{2\epsilon} &= 1 \\ \mu &= \frac{2\epsilon}{n} \end{aligned} \tag{5.15}$$

Finally, substituting the value of μ back into the equation for λ_i:

$$\lambda_i = \frac{\mu}{2\epsilon} = \frac{\frac{2\epsilon}{n}}{2\epsilon} = \frac{1}{n} \tag{5.16}$$

This shows that each weight λ_i converges to $\frac{1}{n}$, uniformly distributing the weights among all n data points. The parameter ϵ controls the regularization and an analysis is necessary to find an optimal ϵ value. This regularization helps stabilize the solution by avoiding extreme values in weights. This approach is validated through a practical example, confirming that the regularization effectively achieves the mitigation of extreme weights under the constraints of the SOCK method. A case is presented to demonstrate the regularization technique.

5.2 Case - Regularization

Figure 5.1 displays the TI used for this case where the regularization method is applied.

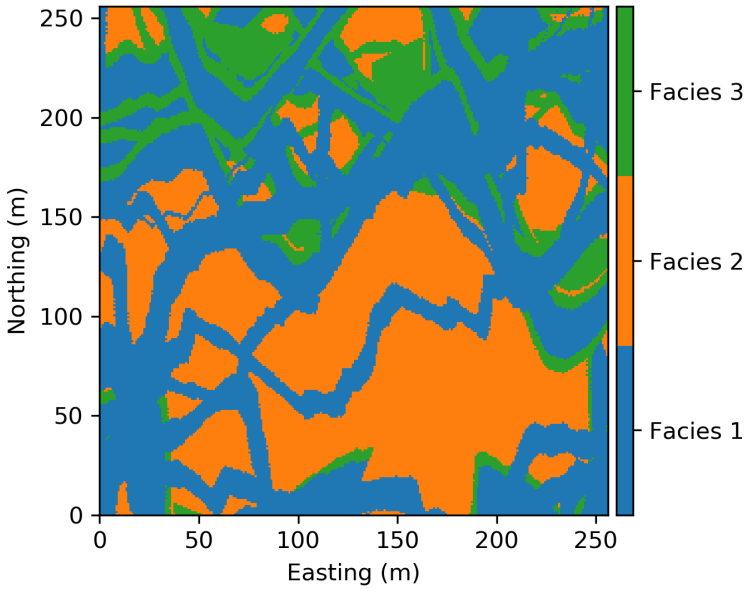


Figure 5.1: TI used for the regularization case.

Regular samples are taken from the TI as shown in Figure 5.2.

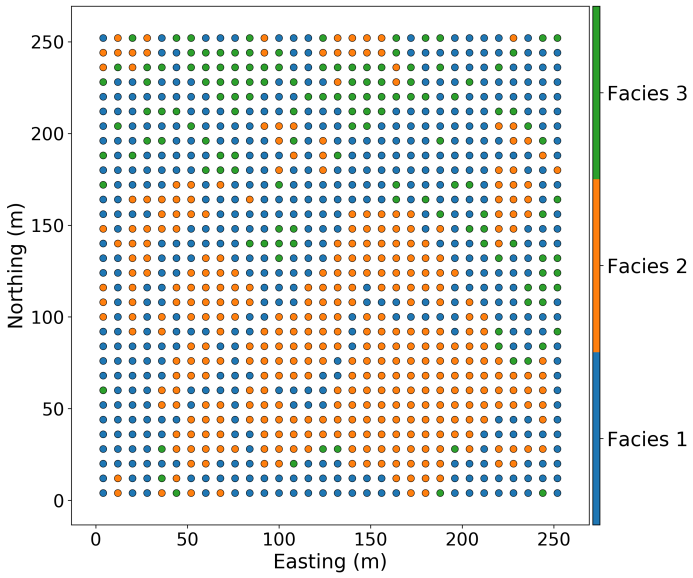


Figure 5.2: Samples taken form TI used in the regularization case.

The SOCK algorithm is executed, and when a system of equations proves to be unstable because the left-hand side (LHS) covariance matrix is not positive definite, the matrix is extracted and analyzed to observe how the weights decrease when a regularization technique is applied (see Figure 5.3).

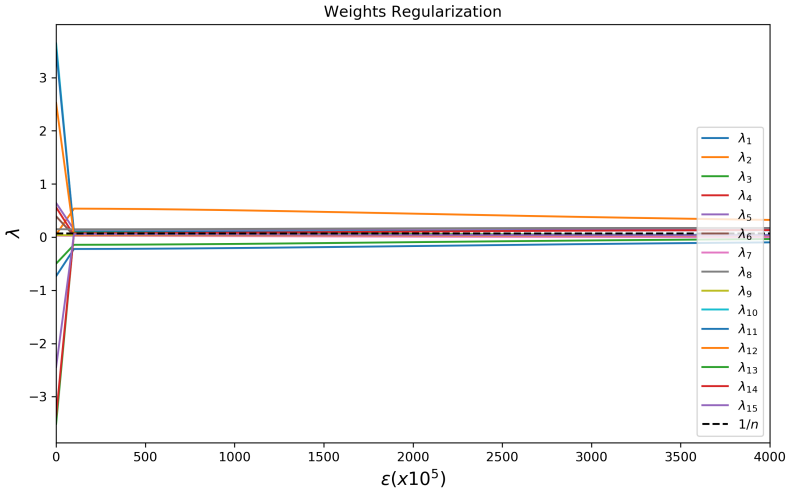


Figure 5.3: Weights behaviour as ϵ increases.

Note that as ϵ increases, the weights converge to $\frac{1}{n}$, as expected (see Equation 5.10). SOCK is executed for a hundred different locations on the grid with various ϵ values to assess the impact on the RMSE, given that the reference is known. The result is displayed in Figure 5.4.

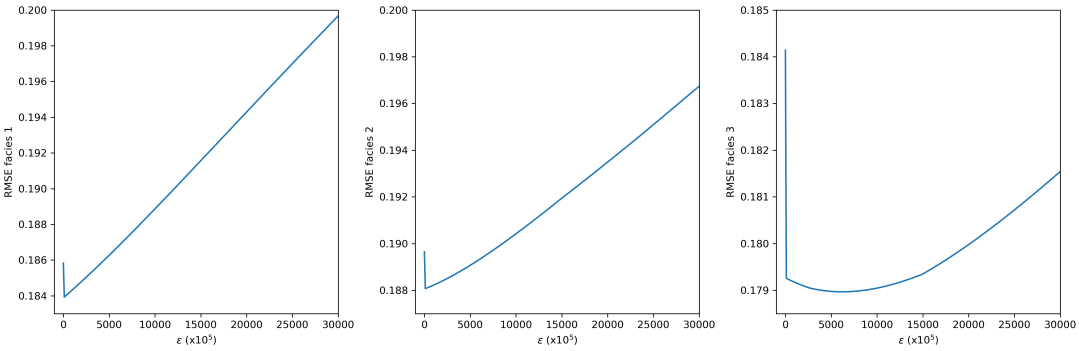


Figure 5.4: RMSE behaviour as ϵ increases for all categories.

Note that although it occurs on a small scale, the RMSE initially decreases, thus improving the estimation. As the weights approach $\frac{1}{n}$, the RMSE increases, which degrades the estimation; this behavior is expected because all the weights become the same. Ideally, a calibrated value for ϵ should be chosen based on the behavior plots of both the RMSE (Figure 5.4) and the weights (Figure 5.3). It is observed that this optimal point occurs for a small ϵ value. For instance, SOCK is executed using $\epsilon = 0$ (without regularization) and then it is executed with $\epsilon = 0.01$, see Table 5.1.

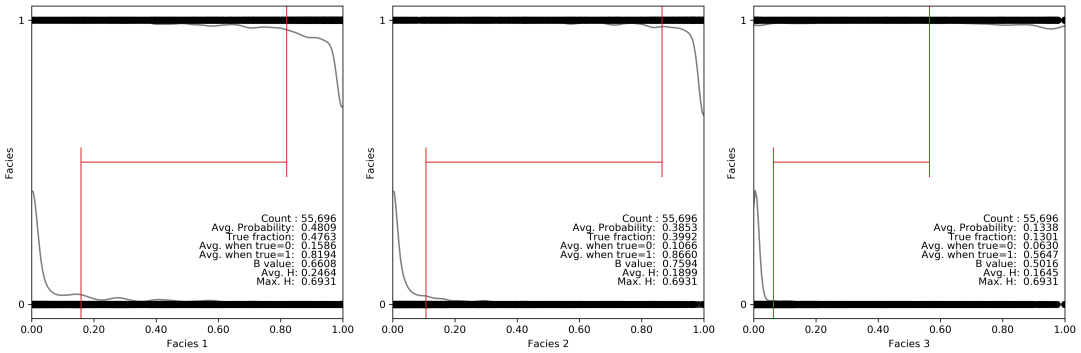
Table 5.1: Example of extreme weight being mitigated after applying $\epsilon = 0.01$.

	$\epsilon = 0$	$\epsilon = 0.01$
Max. weight (λ)	12.708	0.830

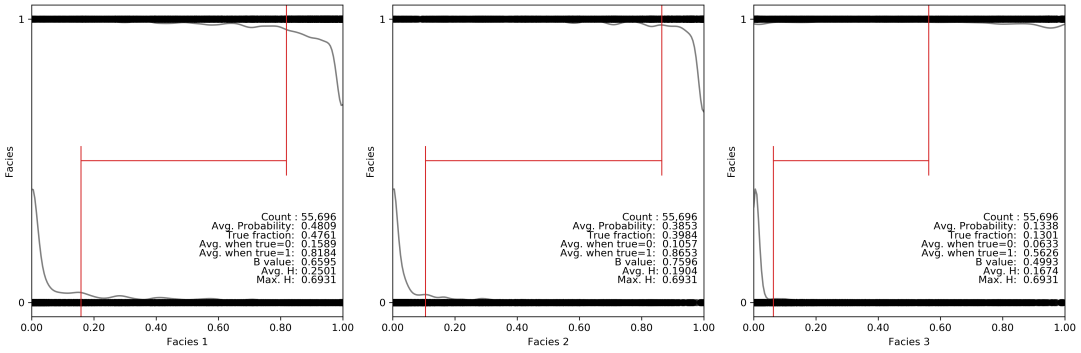
The table shows that the regularization mitigates the extreme weight bringing it down to 0.830 from 12.708 when $\epsilon = 0.01$ is applied to inflate the diagonal elements of the LHS covariance matrix.

5.3 Impact When Estimating on a Grid

The regularization is able to mitigate extreme weights, this section analyzes whether or not this mitigation has an impact when estimating over the entire grid. SOCK is executed for the entire grid without regularization and then with regularization and an $\epsilon = 0.01$.



(a) Indicator cross validation for $\epsilon = 0$.



(b) Indicator cross validation for $\epsilon = 0.01$.

Figure 5.5: Comparison of indicator cross validation for (**a**) $\epsilon = 0$ and (**b**) $\epsilon = 0.01$.

Figure 5.5 shows that after applying the majority rule, the final estimate does not change. This occurs because the RMSE difference shown in Figure 5.4 is not enough to significantly change the final estimation. It also demonstrates that, even in the presence of extreme weights, SOCK still performs well in estimation.

Table 5.2: Table showing the difference in mean and maximum values between estimation with $\epsilon = 0.01$ and $\epsilon = 0$.

	Facies 1	Facies 2	Facies 3
mean	0.000016	-0.000042	0.000022
max	0.351991	0.255687	0.109902

Table 5.2 shows the difference in mean and maximum values between estimations with $\epsilon = 0.01$ and $\epsilon = 0$. Note that the mean difference is insignificant for all three categories, indicating that the final estimation remains unchanged after applying the majority rule. However, examining the maximum difference value reveals that, while regularization impacts the estimate locally, it is not substantial enough to change the final grid estimate.

Table 5.3 shows the weights for a location in the grid applying regularization ($\epsilon = 0.01$) and without applying it when estimating Facies 2.

Table 5.3: Table showing the difference in weights for one location in the grid with ($\epsilon = 0.01$) and without regularization for Facies 2 estimation. The yellow highlights represent the biggest differences in weight.

	Weights - Regularization	Weights - No Regularization	Difference
Facies 2 wt 0	-0.095	-0.280	-0.185
Facies 2 wt 1	0.325	0.152	-0.173
Facies 2 wt 2	-0.037	-0.165	-0.129
Facies 2 wt 3	-0.053	0.096	0.149
Facies 2 wt 4	0.270	0.414	0.144
Facies 2 wt 5	-0.022	0.081	0.103
Facies 2 wt 6	0.003	4.631	4.628
Facies 2 wt 7	0.188	4.615	4.427
Facies 2 wt 8	0.011	3.224	3.213
Facies 2 wt 9	0.029	-1.146	-1.174
Facies 2 wt 10	0.154	-0.971	-1.124
Facies 2 wt 11	0.021	-0.795	-0.815
Facies 2 wt 12	0.091	-3.327	-3.419
Facies 2 wt 13	0.070	-3.202	-3.272
Facies 2 wt 14	0.046	-2.328	-2.373

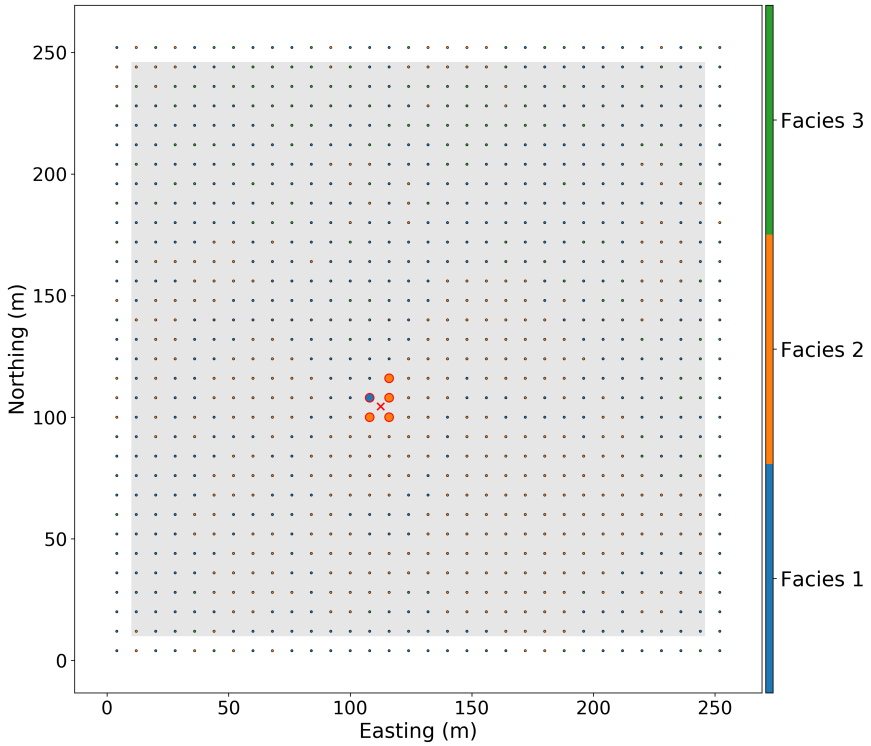


Figure 5.6: Samples configuration for the example shown in Table 5.3

Note that with regularization, extreme weights are mitigated, and the constraint of summing to one is maintained. The highlighted lines in Table 5.3 show the largest differences in weights; for instance, there are cases where the difference is greater than four. Figure 5.6 shows the configuration for the example presented in Table 5.3, where the red cross represents the estimated location and the larger dots represent the five closest samples.

5.4 Chapter Summary

This chapter implemented a method for regularizing weights in linear systems of equations, which are frequently encountered in geostatistics, particularly in applications involving kriging and its variants. These systems are crucial for simulations and for estimating geological variables at unknown grid locations, appearing numerous times based on the number of grid nodes. The chapter discussed the importance of ensuring positive definiteness in these systems to avoid the consequences of indefinite or ill-conditioned matrices, such as extreme weight values. Section 5.1 explained the technique for regularizing weights. Section 5.2 presented a case study where the regularization technique was applied. Section 5.3 evaluated the impact of this method on the estimation of an entire grid. To summarize, this chapter demonstrated that inflating the diagonal of the LHS covariance matrix served as a regularization technique to mitigate extreme weights. Additionally, it showed that as the regularization constant (ϵ) increased, all weights converged to $\frac{1}{n}$. An analysis to determine the optimal ϵ value was conducted using both the RMSE and weights behavior charts as ϵ increased. The application of this technique revealed that while extreme weights were mitigated, it did not alter the overall estimation across an entire grid significantly. This illustrated that kriging performed well even in cases where extreme weights were present.

Chapter 6

Principal Component Analysis for Categorical Indicators

This chapter introduces the Principal Component Analysis (PCA) technique to decorrelate categorical indicators variables and then model each component independently. Section 6.1 introduces the decorrelation tecnique that is going to be applied in a case. Section 6.2 presents a case where the indicators are decorrelated through principal component analysis and independently estimated. Section 6.3 covers a second case and a brief summary of what is covered in this chapter and the results obtained.

6.1 Decorrelation - principal component analysis (PCA)

The PCA transform is used in geostatitics to decorrelate multiple variables. Subsequently, the principal components are modeled independently and the back-transformation reestablishes the original correlation to the modeled variables. Suro (1988) used PCA to decorrelate indicators of continuous variables, this study presents cases where PCA is used to decorrelate indicators of categorical variables. Standardization of the indicators ($\mathbf{Y}$) is done prior to the PCA transformation (Equation 6.1).

$$\mathbf{Y} : y_{\alpha,k} = \frac{\left(i_{\alpha,k} - p_k\right)}{\sigma_k}, \text{ for } \alpha = 1, \ldots, n, k =, 1, \ldots, K \tag{6.1}$$

PCA focuses on the covariance matrix $\mathbf{\Sigma}$ of the standardized data (Equation 6.2). The elements of the covariance matrix describe the linear dependence within the multivariate system of the standardized data.

$$\mathbf{\Sigma} : C_{k,k'} = \frac{1}{n}\sum_{\alpha=1}^{n} y_{\alpha,k} \cdot y_{\alpha,k'}, \text{ for } k, k' = 1, \ldots, K \tag{6.2}$$

The PCA transform begins with the spectral decomposition of the covariance matrix $\mathbf{\Sigma}$, which results in the eigenvector matrix $\mathbf{V}$ and a diagonal eigenvalue matrix $\mathbf{D}$ (Equation 6.3).

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{6.3}$$

The PCA transformation is executed by multiplying the standardized data matrix $\mathbf{Y}$ with the eigenvector matrix $\mathbf{V}$ which rotates the multivariate data and decorrelates the principal components in $\mathbf{P}$ (Equation **??**).

$$\mathbf{P} = \mathbf{Y}\mathbf{V} \tag{6.4}$$

To back transform the data to the standardized form, $\mathbf{P}$ is multiplied by the transpose of $\mathbf{V}$ (Equation

6.5).

$$\mathbf{Y} = \mathbf{P}\mathbf{V}^T \tag{6.5}$$

Each principal component is a linear combination of the original variables. Two cases are presented where PCA is executed to decorrelate categorical indicator variables and model the independent components subsequently.

6.2 Case 1 - PCA for Categorical Indicators

For this case, the TI represented by Figure 6.1 is used. Samples are obtained and Figure 6.2 displays them.

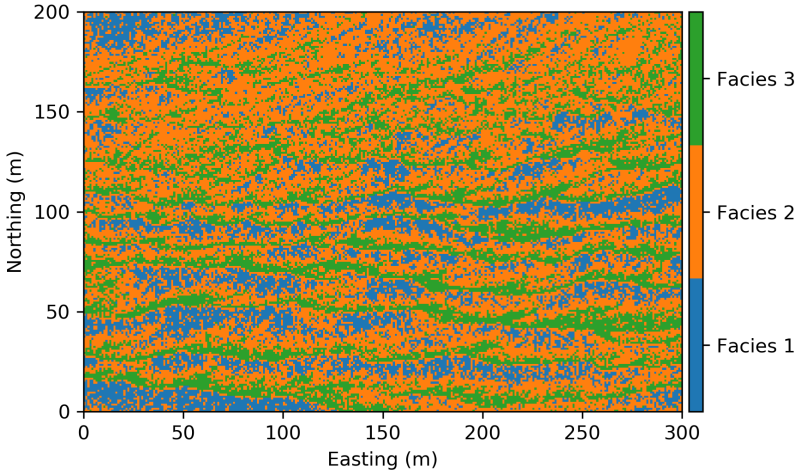


Figure 6.1: TI used for PCA Case 1.

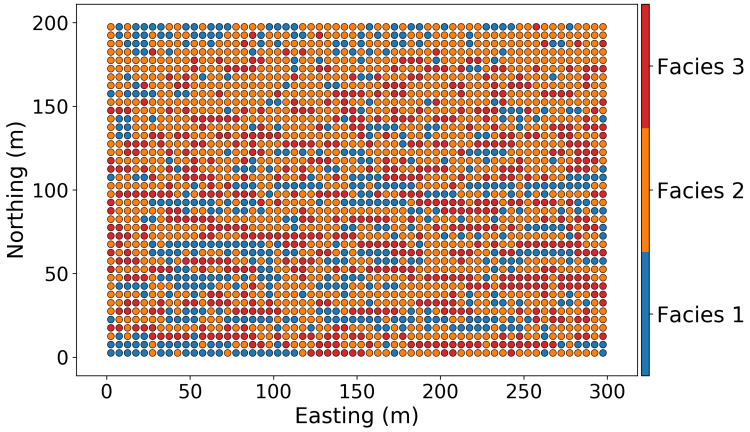


Figure 6.2: Samples obtained from the TI used for the PCA Case 1.

The first step is to standardize the indicators. Table 6.1 shows the statistics of the standardized indicator values.

Table 6.1: Case 1 - Statistics of standardized indicators values

	Facies 1	Facies 2	Facies 3
count	2400.000	2400.000	2400.000
mean	0.000	0.000	0.000
std	1.000	1.000	1.000
min	-0.517	-1.083	-0.576
25%	-0.517	-1.083	-0.576
50%	-0.517	0.923	-0.576
75%	-0.517	0.923	-0.576
max	1.935	0.923	1.736

Note that since indicators are binary (0 or 1) variables, the standardized values of each category indicator remain binary, represented by the maximum and minimum values in Table 6.1. When standardized, indicators remain binary but are dispersed depending on the proportion of each category ($y_{\alpha,k} = \frac{(i_{\alpha,k} - p_k)}{\sqrt{p_k(1-p_k)}}$) because the standard deviation is proportion-dependent. The lower the proportion, the more dispersed these values are. For instance, Facies 1 has the lowest proportion, so its binary values are more dispersed.

Table 6.2: Case 1 - Correlation matrix of the standardized indicators.

	Facies 1	Facies 2	Facies 3
Facies 1	1.000	-0.560	-0.298
Facies 2	-0.560	1.000	-0.624
Facies 3	-0.298	-0.624	1.000

As the standardized covariance equals correlation, the correlation matrix is calculated and shown in Table 6.2. The PCA transformation focuses on this matrix, and the next step involves performing its spectral decomposition. The eigenvector matrix obtained from the decomposition is displayed in Table 6.3.

Table 6.3: Case 1 - Eigenvector matrix.

PC1	PC2	PC3
0.3914	-0.7552	0.5258
-0.7646	0.0510	0.6424
0.5120	0.6535	0.5575

The eigenvector matrix serves as the rotation matrix, as its coefficients rotate the variables to the basis of the standardized matrix. The diagonal eigenvalue matrix is presented in Table 6.4.

Table 6.4: Case 1 - Diagonal eigenvalue matrix.

PC1	PC2	PC3
1.7046	0.0000	0.0000
0.0000	1.2954	0.0000
0.0000	0.0000	0.0000

The diagonal eigenvalue matrix explains the relative variability that each variable contributes to the multivariate system. Note that the third principal component does not contribute to the system's variability; therefore, the multivariate system is explained only by the first and second components. The lack of variability in the third component is due to the fact that the indicators are linearly related. For instance, category 3 can be explained by a combination of categories 1 and 2: $I(u;3) = 1 - \sum_{k=1}^{2} i(u;k)$, since $\sum_{k=1}^{K} i(u;k) = 1$, where K is the number of categories. Since there is no variability in the third component, the estimation can proceed with one fewer variable, resulting in dimension reduction when using the PCA transformation for categorical indicators. Figure 6.3 shows the uncorrelated principal components.

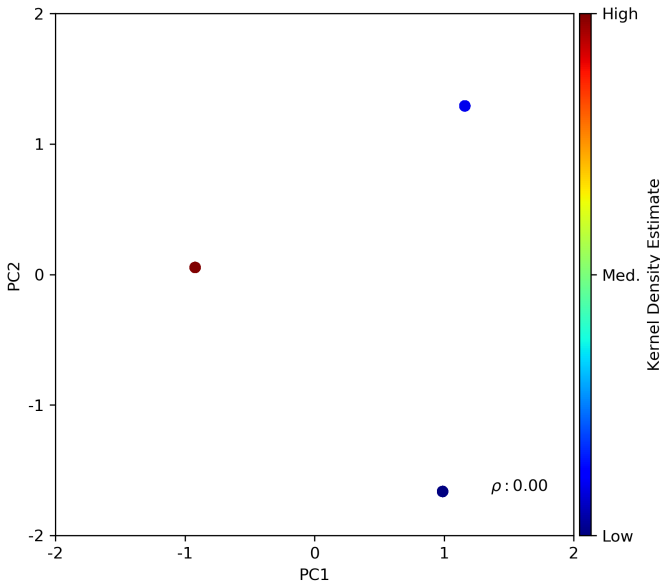


Figure 6.3: Case 1 - Uncorrelated principal components.

Figure 6.4 displays the principal components in three dimensions, more clearly illustrating the lack of variability in the third component.

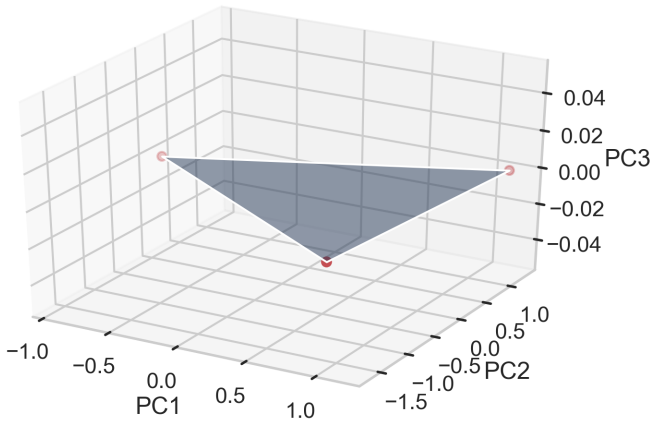
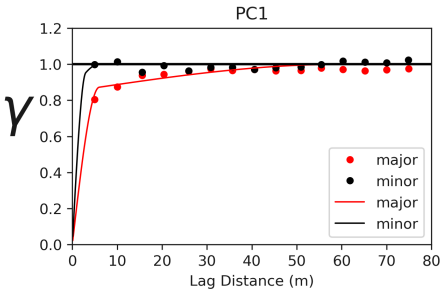
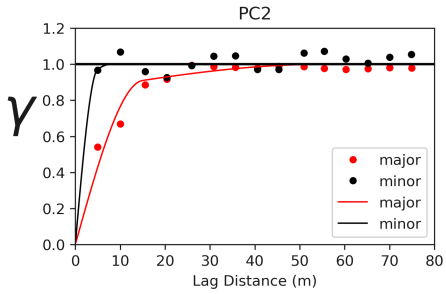


Figure 6.4: Case 1 - Principal components in three dimension, illustrating the lack of variability of the third component.

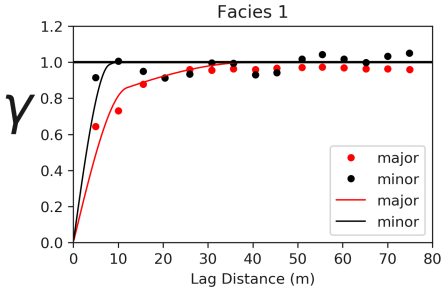
The triangular plane in Figure 6.4 represents the possible outcomes for the estimation. Enforcing that the estimates fall inside the triangular plane will mitigate the order relation problem that kriging might cause. As the dimension is reduced, variogram modeling and estimation are performed for the first and second principal components.



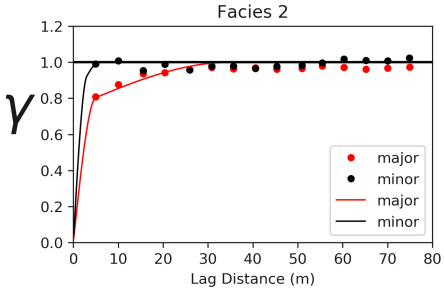
(a) Case 1 - Principal component 1 (PC1) experimental and modeled variograms.



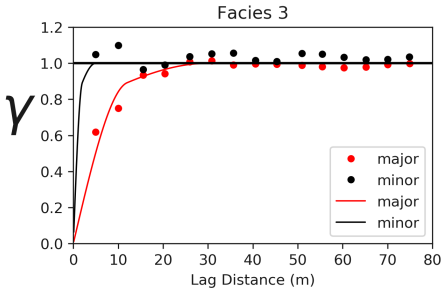
(b) Case 1 - Principal component 2 (PC2) experimental and modeled variograms.



(c) Case 1 - Facies 1 experimental and modeled variograms.



(d) Case 1 - Facies 2 experimental and modeled variograms.



(e) Case 1 -Facies 3 experimental and modeled variograms.

Figure 6.5: Case 1 - Experimental and modeled variograms of principal components 1 (**a**) and 2 (**b**); and experimental and modeled variograms of indicators of Facies 1 (**c**), 2 (**d**) and 3 (**e**).

Figure 6.5 displays the experimental and modeled variograms for principal components 1 and 2, as well as for each facies indicator. The dots represent the experimental values, and the lines represent the fitted models. The color red indicates the major direction, while black denotes the minor direction. Observe how the major direction of the variogram of the principal components has a longer range than the indicator variograms, which implies that the spatial structure is mixed after the PCA transformation. Additionally, note that the variogram points are less correlated for principal component 1 (PC1) compared to principal component 2 (PC2) and the indicator variograms. The estimations for principal components 1 and 2 are performed independently using ordinary kriging.

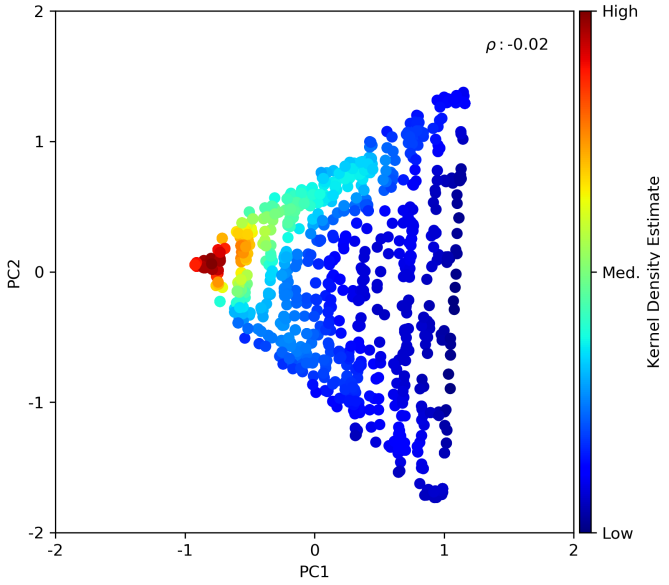
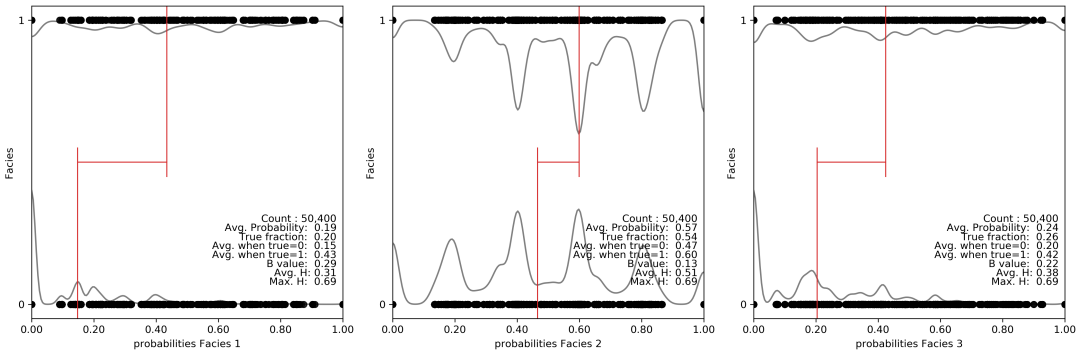
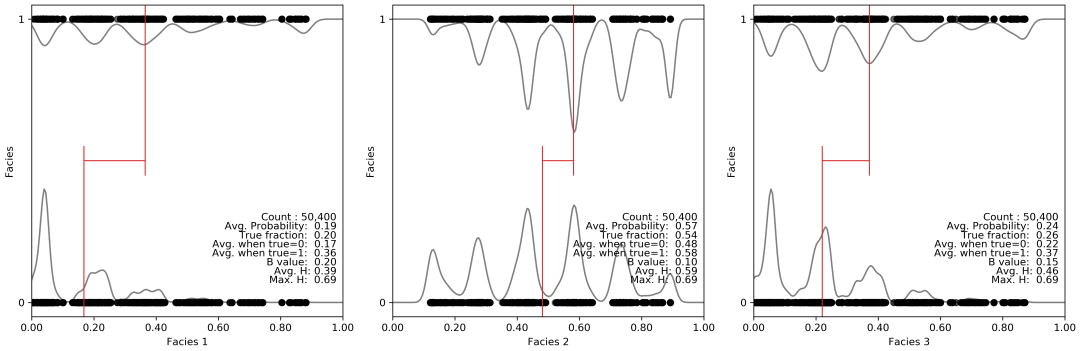


Figure 6.6: Case 1 - Estimates of principal components 1 and 2.

Figure 6.6 presents the estimates for principal components 1 and 2. Note that not all estimates fall within the triangular plane. Those outside the triangular plane are adjusted to the nearest side of the triangle to mitigate order relation issues. The back transformation is then executed by multiplying the principal component estimates by the transpose of the eigenvector matrix. Since this transformation converts the values back to a standardized form, they must be reverted to their original form.



(a) Case 1 - Cross validation against the reference when estimating with ordinary kriging .



(b) Case 1 - Cross validation against the reference when using PCA and estimating the components independently with ordinary kriging.

Figure 6.7: Case 1 - Cross validations of (a) estimating with ordinary kriging; and (b) decorrelating through PCA and independently estimating the components.

Table 6.5: Case 1 - Table comparing B values of OK estimates and the estimates from the PCA workflow.

Facies 1	OK	PCA workflow
B	0.29	0.20
Facies 2	OK	PCA workflow
B	0.13	0.10
Facies 3	OK	PCA workflow
B	0.22	0.15

Figure 6.7 shows the indicator cross-validation plots for OK estimates as well as for the estimates using the PCA workflow. Table 6.5 summarizes the B values for each method. The B value indicates that the PCA workflow does not outperform the ordinary kriging algorithm. Therefore, although it is useful for dimension reduction and variable decorrelation, the estimation accuracy of the PCA workflow is poor. PC1 represents most of the variability of the multivariate system, but its variogram points are less correlated compared to PC2 and the indicator variograms. Additionally, the variograms' ranges for PC1 and PC2 account for the ranges of the three indicator variograms, implying that the spatial structure mixes during the PCA transformation. These factors could

explain why the PCA workflow is less accurate than ordinary kriging. A second case is also presented.

6.3 Case 2 - PCA

The TI used for Case 2 is displayed in Figure 6.8. Samples are taken, and they are shown in Figure 6.9.

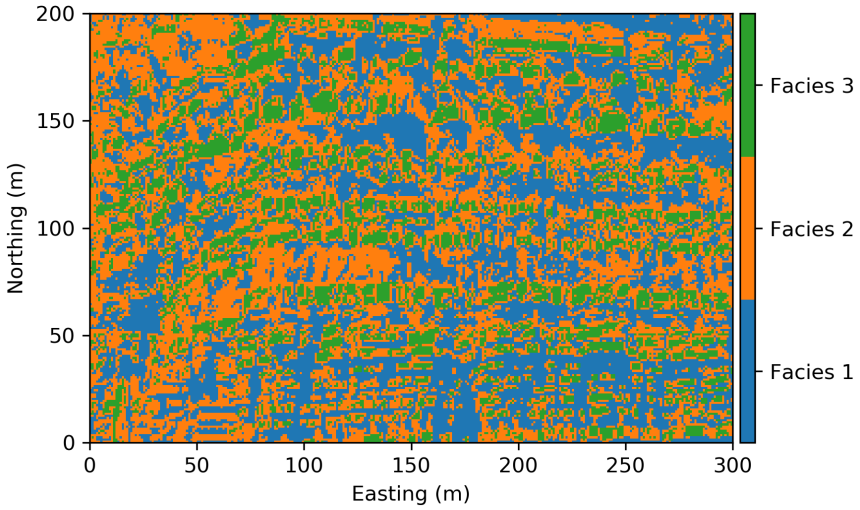


Figure 6.8: TI used in Case 2.

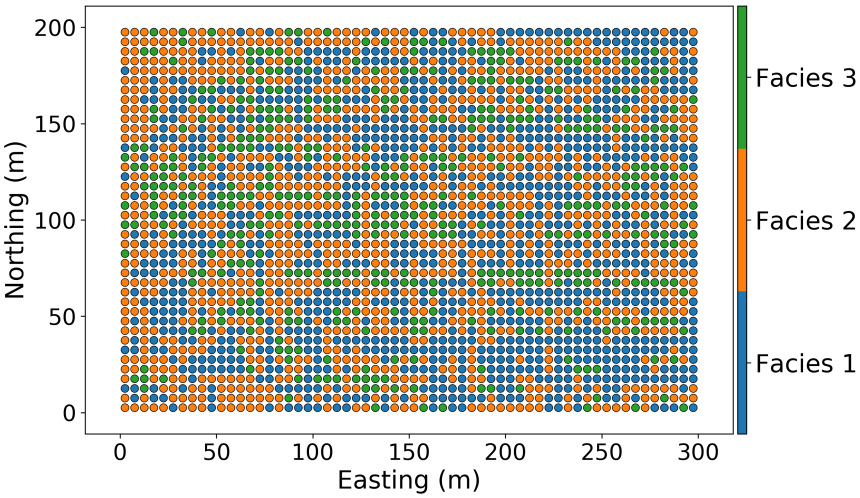


Figure 6.9: Samples taken from TI used in Case 2.

Initially, the indicators are standardized to facilitate the interpretation of the PCA results. Table 6.6 summarizes the statistics of the standardized indicators.

Table 6.6: Case 2 - Statistics of standardized indicators values

	Facies 1	Facies 2	Facies 3
count	2400.000	2400.000	2400.000
mean	0.000	0.000	0.000
std	1.000	1.000	1.000
min	-0.791	-0.875	-0.471
25%	-0.791	-0.875	-0.471
50%	-0.791	-0.875	-0.471
75%	1.264	1.143	-0.471
max	1.264	1.143	2.125

Because indicators are binary variables, the standardized variables are also binary and are represented by the minimum and maximum values in Table 6.6.

Table 6.7: Case 2 - Correlation matrix of the standardized indicators.

	Facies 1	Facies 2	Facies 3
Facies 1	1.000	-0.692	-0.372
Facies 2	-0.692	1.000	-0.412
Facies 3	-0.372	-0.412	1.000

The correlation matrix is calculated and represented in Table 6.7. The spectral decomposition of this matrix is performed, resulting in the eigenvector matrix (Table 6.8) and the diagonal eigenvalue matrix (Table 6.9).

Table 6.8: Case 2 - Eigenvector matrix.

PC1	PC2	PC3
-0.6890	-0.3872	0.6127
0.7224	-0.2980	0.6240
-0.0590	0.8725	0.4850

Table 6.9: Case 2 - Diagonal eigenvalue matrix.

PC1	PC2	PC3
1.6941	0.0000	0.0000
0.0000	1.3059	0.0000
0.0000	0.0000	0.0000

The eigenvector serves as a rotation matrix, as its coefficients rotate the variables to the basis of the standardized matrix. The diagonal eigenvalue matrix explains the contribution of each component to the total variability of the system. Note that the third principal component contributes nothing to the system's variability because the third variable can be explained through a linear combination of the other two variables. The multivariate system is explained solely by the first and second principal components. The estimation uses just these first two components, demonstrating

that dimension reduction occurs when using PCA for categorical variables. The variables are now uncorrelated, as displayed in Figure 6.10.

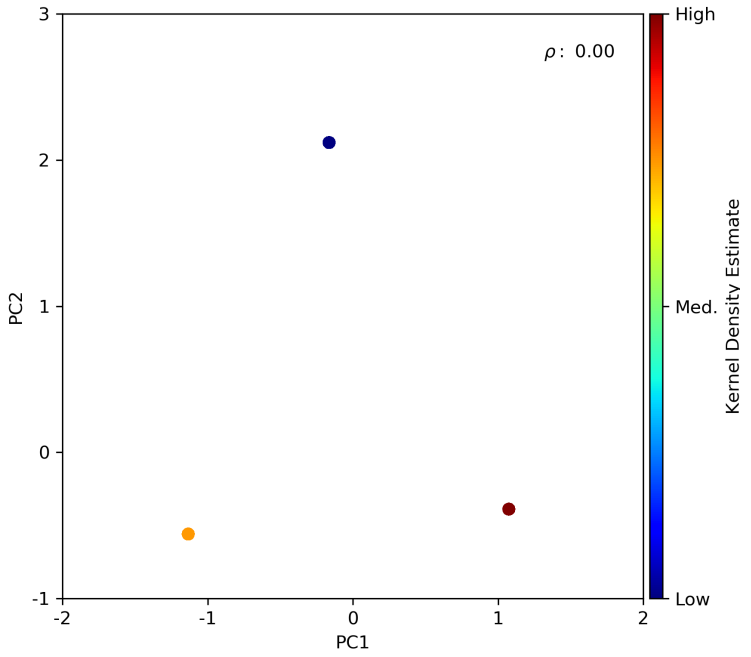


Figure 6.10: Case 2 - Uncorrelated principal components.

Figure 6.11 displays the principal components in three dimension, showing the lack of variability on the third component.

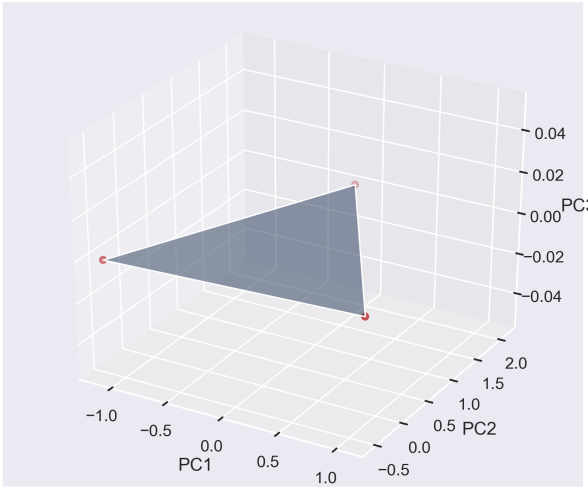
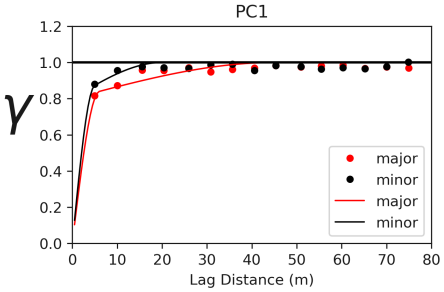
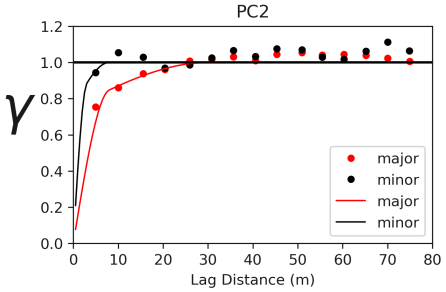


Figure 6.11: Case 2 - Principal components in three dimension, illustrating the lack of variability of the third principal component.

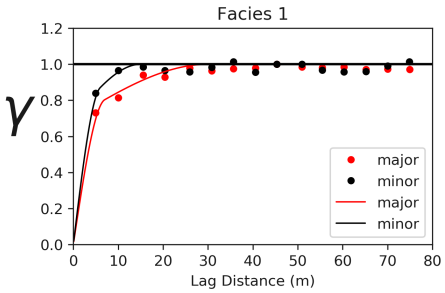
The shaded triangular plane in Figure 6.11 represents the possible outcomes for the estimates. Forcing the estimates to fall within this plane ensures mitigation of the order relation issues that kriging might cause. Due to the dimension reduction, variograms are calculated and modeled only for the first and second principal components.



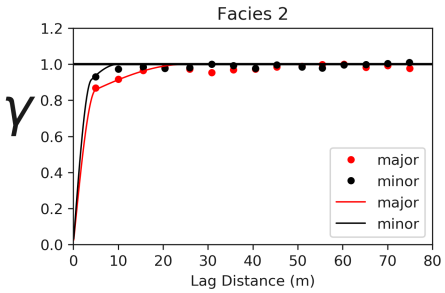
(a) Case 2 - Principal component 1 (PC1) experimental and modeled variograms.



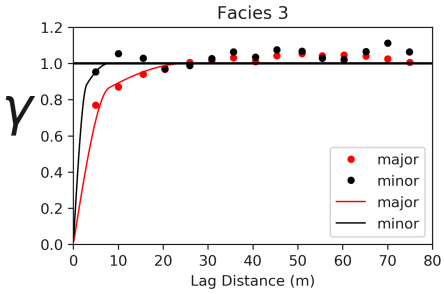
(b) Case 2 - Principal component 2 (PC2) experimental and modeled variograms.



(c) Case 2 - Facies 1 experimental and modeled variograms.



(d) Case 2 - Facies 2 experimental and modeled variograms.



(e) Case 2 -Facies 3 experimental and modeled variograms.

Figure 6.12: Case 2 - Experimental and modeled variograms of principal components 1 (**a**) and 2 (**b**); and experimental and modeled variograms of indicators of Facies 1 (**c**), 2 (**d**) and 3 (**e**).

Figure 6.12 shows the experimental and modeled variograms for principal components 1 and 2, along with the variograms for each facies indicator. The dots represent the experimental values, and the lines represent the fitted models. The color red indicates the major direction, while black denotes the minor direction. Again, notice that the variograms of the principal components have a longer range than the indicator variograms, and that PC1 variogram points are less correlated

than the other variograms. This implies that the spatial structure becomes mixed after the PCA transformation. The estimations for principal components 1 and 2 are done independently using ordinary kriging.

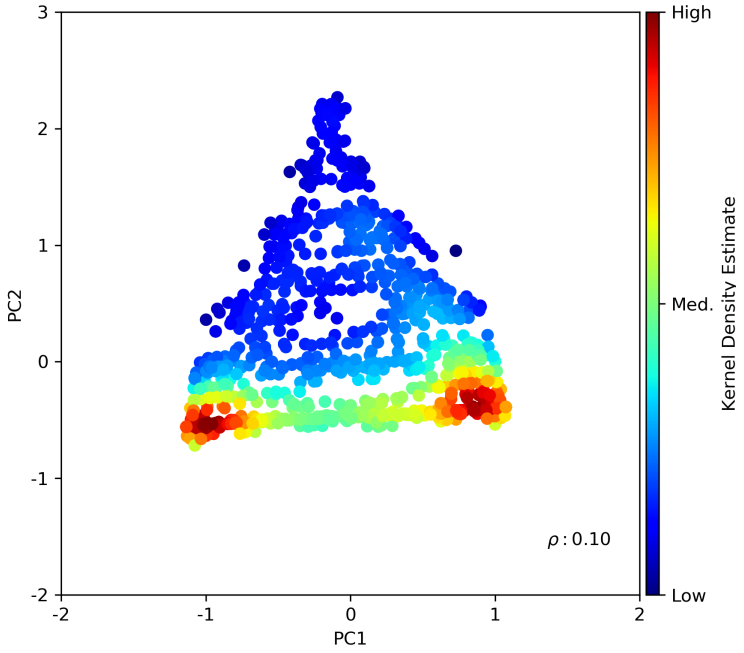
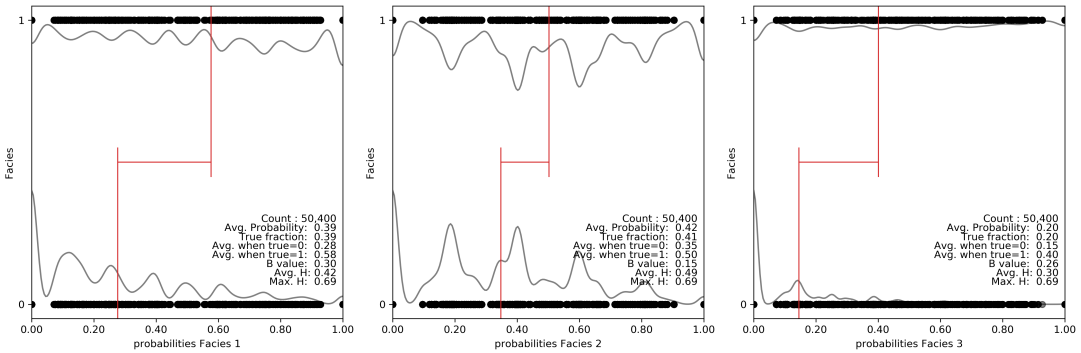
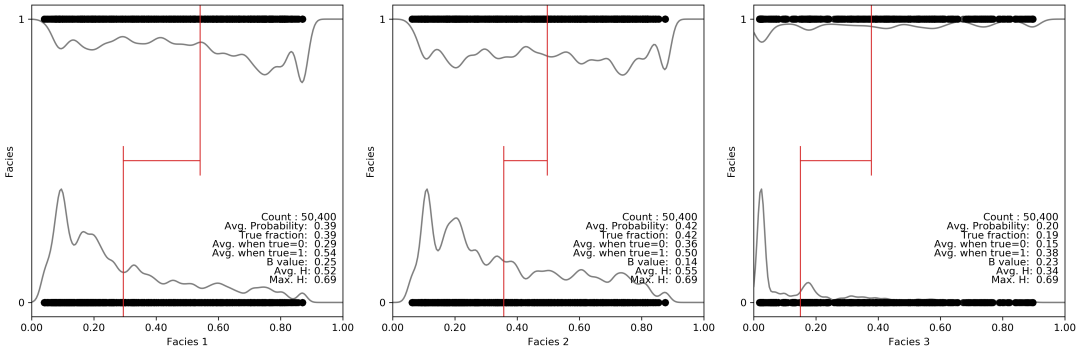


Figure 6.13: Case 2 - Estimates of principal components 1 and 2.

Figure 6.13 shows the scatter plots of the first and second principal components. Note that not all estimates fall inside the triangular plane. Estimates outside the plane have been adjusted to the nearest side of the triangle to mitigate order relation issues. The back transformation to standardized values is executed by multiplying the estimates by the transpose of the eigenvector matrix. Since this back transformation converts the values to a standardized format, they must be converted back to their original values.



(a) Case 2 - Cross validation against the reference when estimating with ordinary kriging .



(b) Cross validation against the reference when using PCA and estimating the components independently with ordinary kriging.

Figure 6.14: Case 2 - Cross validations of (a) estimating with ordinary kriging; and (b) decorrelating through PCA and independently estimating the components.

Table 6.10: Case 2 - Comparison of B values of OK estimates and the estimates from the PCA workflow.

Facies 1	OK	PCA workflow
B	0.30	0.25
Facies 2	OK	PCA workflow
B	0.15	0.14
Facies 3	OK	PCA workflow
B	0.26	0.23

Figure 6.14 displays the indicator cross-validation plots for OK estimates and for those calculated using the PCA workflow for Case 2. Table 6.10 details the B values for each method. Similar to Case 1, the B value indicates that the PCA workflow does not outperform the ordinary kriging method due to the spatial structure mixing when doing the PCA transformation. Thus, while the PCA workflow is useful for reducing dimensions and decorrelating variables, its estimation accuracy remains low.

6.4 Chapter Summary

This chapter introduced the PCA technique for decorrelating indicator variables and independently estimating each decorrelated component. The technique is interesting because it reduces the dimensions by $K-1$, where K represents the number of categorical variables involved. By ensuring that the estimates remain within the triangular plane, this approach successfully addresses the order relation issues often associated with kriging. However, despite these benefits, the estimation of the independent components and their subsequent transformation back to the original values proved to be less effective than ordinary kriging. This was demonstrated by the B values in the presented cases, showing that although PCA reduces complexity in data relationships by decorrelating the variables, it may not improve estimation accuracy. When the original variables are transformed using PCA, there is mixing in the spatial structure, as seen in Figures 6.5 and 6.12, where the variogram ranges of the two principal components account for the variogram ranges of the three category indicators. Additionally, PC1, which represents most of the variability of the multivariate system, has its variogram points less correlated than the other variograms. This can lead to poor estimates, as seen in the PCA workflow.

Chapter 7

Conclusion

This chapter reviews the research motivation (Section 7.1), the contributions of this thesis (Section 7.2), the limitations of the thesis (Section 7.3), and suggests future work for advancing the modeling of categorical variables.

7.1 Motivation

Categorical variables are modeled first as they establish volumes in which the continuous variables are considered stationary. Therefore, categorical modeling is crucial in geostatistical workflows and can significantly influence subsequent stages of resource estimation. This thesis explores an indicator-based approach to assess the capability of estimating categorical variables using covariances and cross-covariances through a linear model of coregionalization. The characteristics of indicator variograms and covariances are also examined, including an explanation of why the nugget effect for a categorical indicator variogram must be zero. Additionally, the thesis implements a regularization technique to mitigate extreme weights arising from unstable linear systems in kriging equations. Moreover, principal component analysis is used to decorrelate the indicators and estimate the components independently. The thesis aims to enhance understanding of indicators for categorical variables, as this would improve the modeling of categorical variables and, consequently, resource estimation.

7.2 Contributions summary

Throughout the development of this thesis, the characteristics of categorical indicator variograms and covariances are studied. The study demonstrates that the nugget effect for categorical indicator variograms must be zero and includes a proof of this assertion. A non-stationary relationship between the variogram and covariance is derived as part of this study. The robustness of models for different scenarios, including clusters of data, is also analyzed. The analysis shows that the correlogram is a robust alternative compared to the variogram for indicators of categories.

An extensive comparative study of multiple-point statistics (MPS) based conditional probabilities against a range of kriging algorithms, including simple kriging (SK), simple cokriging (SCK), ordinary kriging (OK), and standardized ordinary cokriging (SOCK) is conducted. The MPS probabilities are used as a benchmark as they consider the most information. The study considers small example, and the performance of each kriging method is assessed using metrics such as mean squared error (MSE), coefficient of determination (R^2), and correlation (ρ). Subsequently, the study was

applied to entire grids and compared against the reference, including a case where a very large linear model of coregionalization was used. Moreover, the study was extended to include images from the Data Validation Project by Mokdad et al. (2022). The results and contributions from this study are summarized:

- SK and SCK had issues with order relations and did not perform well.
- SOCK provided positive results when applied to small examples.
- SOCK with VL-LMC performs almost as well as MPS methods.
- When analyzing images from the Data Validation Project, OK performs about as well as SOCK. The difference in RMSE gives a small advantage to SOCK.

A technique to regularize weights from unstable linear systems of kriging equations by inflating the diagonal of the left-hand side covariance matrix with a constant ϵ is implemented. Because the weights in standardized ordinary cokriging are constrained to sum to one, as the constant ϵ increases, the weights converge to $\frac{1}{n}$, where n represents the number of data points. The application of the technique shows that while extreme weights are mitigated, it does not change the overall estimation of an entire grid. This demonstrates that kriging performs well even in the presence of extreme weights.

Lastly, this thesis decorrelates the categorical indicator variables through principal component analysis and estimates the components independently. An interesting observation noted during the decorrelation of the indicator variables is the reduction in dimension to $K - 1$, where K is the number of categories. This dimension reduction occurs because the last principal component has no variance. For cases with three categories presented in this thesis, the decorrelated principal components form a triangular 2-D plane. Enforcing the estimates to fall within the valid region of this plane ensures there are no issues with order relations. However, despite the dimension reduction and the addressed order relation issues, the estimation of the independent components and their subsequent back transformation to original values proved to be less effective than ordinary kriging. This could be explained by the mixing of spatial structure when the indicator variables are transformed into principal components, which consequently results in poorer estimates through the PCA workflow.

7.3 Limitations and future work

The work presented in this thesis is primarily image-based, and the methods have not been tested on real-case datasets. Therefore, the study could be extended to real-case scenarios. Additionally, the cases shown in this thesis include only three categories. Thus, the study could be expanded to encompass more categories. An interesting extension of this study is the possibility of applying it

to multiple categories; for example, it could include lithology, mineralization, and alteration. This would allow for the exploration of cross-relationships between different categories.

Regarding the regularization technique, the analysis shows an impact on the estimation locally, but not sufficient to change the final grid estimates. Therefore, a study could be done to analyze the importance of correcting estimates locally as it might be valuable to know the actual value of locally correcting estimates.

There is a need to mention the importance of other categorical modeling methods. Until the hierarchical truncated pluri-Gaussian method was developed by D. Silva and Deutsch (2019), the approach for numerical modeling of categorical variables was to either combine the variables or to model them independently. The collocated joint relationships between categories are not reproduced when modeling them independently. The categorical multivariate approach through the hierarchical truncated pluri-Gaussian method improves the reproduction of multivariate relationships between the variables (D. Silva & Deutsch, 2019). D. Silva and Deutsch (2019) did not account for the spatial cross-correlation; therefore, the inclusion of cross-correlation is another topic of future research for the simultaneous multivariate categorical modeling as it contains information about the spatial configuration between categories.

The MPS method was developed to simulate complex geological models based on a training image while having the flexibility to honor conditioning data. It works by taking patterns from a single image and integrating them with actual data from specific locations, like wells. However, one of the challenges with MPS is that it has difficulty recreating complex, nonlinear patterns in geological models. This is partly due to the restriction to small training patterns and computational limitations, which compromise the representation of larger-scale features. Additionally, MPS does not fully capture the variability and uncertainty of geological inference because it often assumes stationarity, relies on repetitive patterns from training images, and treats patterns as independent and with the same statistical properties. These simplified assumptions can lead to incomplete or biased representations of geological variability and uncertainty. Recent advances in deep machine learning techniques show that the generative adversarial networks (GANs) method outperforms MPS in generating more geologically realistic models constrained by well data. Because of its potential and the ability to be conditioned by numerous well measurements, researching GANs for modeling categorical variables is valuable.

References

Agresti, A. (2002). Categorical data analysis. *Wiley Series in Probability and Statistics*, 721.

Alabert, F. G. (1987). *Stochastic imaging of spatial distributions using hard and soft information* (Unpublished doctoral dissertation). Stanford University Press.

Armstrong, M. (1992). Positive definiteness is not enough. *Mathematical geology*, *24*, 135–143.

Armstrong, M., Galli, A., Beucher, H., Loc'h, G., Renard, D., Doligez, B., … Geffroy, F. (2011). *Plurigaussian simulations in geosciences*. Springer Science & Business Media.

Bai, J., & Deutsch, C. V. (2020). The pairwise relative variogram. *Geostatistics Lessons, 2020a. URL https://geostatisticslessons. com/lessons/pairwiserelative*.

Barnett, R., & Deutsch, C. (2018). Straightforward modeling of a linear model of coregionalization: the very large lmc alternative. *CCG Paper 2018-121*.

Barnett, R. M. (2015). *Managing complex multivariate relations in the presence of incomplete spatial data.* (Doctoral dissertation, University of Alberta). Retrieved from `https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat03710a&AN=alb.7083693&site=eds-live&scope=site`

Barnett, R. M. (2017). Principal component analysis. *Geostatistics Lessons; Deutsch, JL.* (Retrieved from `https://geostatisticslessons.com/lessons/principalcomponentanalysis`)

Barnett, R. M., Manchuk, J. G., & Deutsch, C. V. (2014). Projection pursuit multivariate transform. *Mathematical Geosciences*, *46*, 337–359.

Bezier, P. (1974). Mathematical and practical possibilities of unisurf. In *Computer aided geometric design* (pp. 127–152). Elsevier.

Boisvert, J. B., Pyrcz, M. J., & Deutsch, C. V. (2007). Multiple-point statistics for training image selection. *Natural Resources Research*, *16*, 313–321.

Boucher, A. (2009). Sub-pixel mapping of coarse satellite remote sensing images with stochastic simulations from training images. *Mathematical geosciences*, *41*, 265–290.

Chilès, J.-P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (Vol. 713). John Wiley & Sons.

Christakos, G. (1984). On the problem of permissible covariance and variogram models. *Water Resources Research*, *20*(2), 251–265.

Cowan, J., Beatson, R., Ross, H., Fright, W., McLennan, T., Evans, T., … Titley, M. (2003, 11). Practical implicit geological modelling..

Davis, M. W. (1987). Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical geology*, *19*, 91–98.

Desbarats, A., & Dimitrakopoulos, R. (2000). Geostatistical simulation of regionalized pore-size

distributions using min/max autocorrelation factors. *Mathematical Geology*, *32*, 919–942.

Deutsch, C. V. (1992). *Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data*. stanford university.

Deutsch, C. V. (1998). Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers & Geosciences*, *24*(6), 551–562.

Deutsch, C. V. (2006). A sequential indicator simulation program for categorical variables with point and block data: Blocksis. *Computers & Geosciences*, *32*(10), 1669–1681.

Deutsch, C. V. (2010). Display of cross validation/jackknife results. *Centre for Computational Geostatistics Annual Report*, *12*(406), 1–4.

Deutsch, C. V., & Journel, A. G. (1997). *GSLIB Geostatistical software library and user's guide* (second ed., Vol. 369). New York: Oxford University Press.

Deutsch, J. L., & Deutsch, C. V. (2012). Accuracy plots for categorical variables. *Centre for Computational Geostatistics Report*, *14*, 404.

Emery, X., & Cornejo, J. (2010). Truncated gaussian simulation of discrete-valued, ordinal coregionalized variables. *Computers & geosciences*, *36*(10), 1325–1338.

Friedman, J. H. (1987). Exploratory projection pursuit. *Journal of the American statistical association*, *82*(397), 249–266.

Galli, A., et al. (1994). The pros and cons of the truncated gaussian method. *Geostatistical Simulations: Proceedings of the Geostatistical Simulation Workshop, Fontainebleau, France, 27–28 May 1993*, 217–233.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Applied Geostatistics.

Guardiano, F. B., & Srivastava, R. M. (1993). Multivariate geostatistics: beyond bivariate moments. In *Geostatistics tróia'92: Volume 1* (pp. 133–144). Springer.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Hu, L., & Chugunova, T. (2008). Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. *Water Resources Research*, *44*(11).

Isaaks, E., & Srivastava, R. (1989). *An introduction to applied geostatistics*. Oxford University Press.

Journel, A., & Alabert, F. (1989). Focusing on spatial connectivity of extreme-valued attributes: Stochastic indicator models of reservoir heterogeneities. *AAPG Bull.;(United States)*, *73*(CONF-890404-).

Journel, A., & Isaaks, E. (1984). Conditional indicator simulation: application to a saskatchewan uranium deposit. *Journal of the International Association for Mathematical Geology*, *16*, 685–718.

Journel, A., & Zhang, T. (2006). The necessity of a multiple-point prior model. *Mathematical geology*, *38*, 591–610.

Journel, A. G. (1983). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, *15*(3), 445–468.

Journel, A. G. (2004). Beyond covariance: the advent of multiple-point geostatistics. In *Geostatistics banff 2004* (pp. 225–233). Springer.

Journel, A. G., & Alabert, F. G. (1990). New method for reservoir mapping. *Journal of Petroleum technology*, *42*(02), 212–218.

Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, *52*(6), 119–139.

Lantuéjoul, C. (2002). Object based models. *Geostatistical Simulation: Models and Algorithms*, 167–182.

Lyster, S. J. (2009). *Simulation of geologic phenomena using multiple-point statistics in a gibbs sampler algorithm.* Retrieved from `https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat03710a&AN=alb.4492606&site=eds-live&scope=site`

Manchuk, J. G., & Deutsch, C. (2008). Robust solution of systems of equations in geostatistics. *Centre for Computational Geostatistics Report*, *10*.

Mariethoz, G., Renard, P., Cornaton, F., & Jaquet, O. (2009). Truncated plurigaussian simulations to characterize aquifer heterogeneity. *Groundwater*, *47*(1), 13–24.

Matheron, G. (1962). *Traité de géostatistique appliquée* (No. v. 1). Éditions Technip. Retrieved from `https://books.google.ca/books?id=88YKAQAAMAAJ`

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246–1266.

Matheron, G. (1971). *The theory of regionalized variables and its applications.* École Nationale Supérieure des Mines. Retrieved from `https://books.google.ca/books?id=5PcpnQEACAAJ`

Matheron, G. (1989). The internal consistency of models in geostatistics. *Geostatistics: Proceedings of the Third International Geostatistics Congress September 5–9, 1988, Avignon, France*, 21–38.

Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guérillot, D., & Ravenne, C. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs. *SPE Annual Technical Conference and Exhibition?*, SPE–16753.

McLennan, J. A., & Deutsch, C. V. (2006). Implicit boundary modeling (boundsim). *Edmonton: Centre for Computational Geostatistics.*

Mizuno, T., & Deutsch, C. (2022). Sequential indicator simulation (sis). *Geostatistics Lessons.* (Retrieved from `http://www.geostatisticslessons.com/lessons/sequentialindicatorsim`)

Mokdad, A., Binakaj, D., & Boisvert, J. (2022). Data validation project: Validation of 114 spatial 2d datasets (non synthetic data). *Twenty-fourth Annual Report of the Centre for Computational Geostatistics, University of Alberta, Edmonton*, 1–8.

Ortiz, J. M., & Deutsch, C. V. (2004). Indicator simulation accounting for multiple-point statistics.

Mathematical Geology, *36*, 545–565.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, *2*(11), 559–572.

Pyrcz, M. J., & Deutsch, C. V. (2006). Semivariogram models based on geometric offsets. *Mathematical geology*, *38*(4), 475–488.

Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford University Press.

Rossi, M. E., & Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.

Serra, J. (1967). Un critère nouveau de découverte de structures: le variogramme. *Sciences de la Terre*, *12*(4), 275–299.

Silva, D. (2018). *Enhanced geologic modeling of multiple categorical variables.* Retrieved from `https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat03710a&AN=alb.8965834&site=eds-live&scope=site`

Silva, D., & Deutsch, C. (2019). Multivariate categorical modeling with hierarchical truncated pluri-gaussian simulation. *Mathematical Geosciences*, *51*(5), 527–552.

Silva, D. A. (2015). *Enhanced geologic modeling with data-driven training images for improved resources and recoverable reserves* (Doctoral dissertation, University of Alberta). Retrieved from `https://login.ezproxy.library.ualberta.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=cat03710a&AN=alb.7575009&site=eds-live&scope=site`

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical geology*, *34*, 1–21.

Strebelle, S. B. (2000). *Sequential simulation drawing structures from training images*. Stanford University.

Suro, V. (1988). *Indicator kriging based on principal component analysis* (Unpublished doctoral dissertation). Stanford University.