

Western Australian School of Mines

Development of a Grade Control Technique Optimizing  
Dilution and Ore Loss Trade-off in Lateritic Bauxite  
Deposits

Yasin Dagan


The thesis is presented for the Degree of  
Doctor of Philosophy  
of  
Curtin University

August 2018

## Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature :  .....

Date : 15.08.2018 .....

# Acknowledgements

First, I would like to sincerely thank Oktay Erten for giving me this priceless opportunity to pursue a PhD study at WASM. He has always been very supportive and provided me with great guidance and motivation until the end of my studies. I also thank him for encouraging me to incorporate multiple-point statistics in the thesis. I extend my thanks to Erkan Topal for giving me the opportunity to study at WASM as well. He has never deprived me of his support even when he was far away. I also much appreciate Roger Thompson for taking the role of being my supervisor towards the end of my studies. He has always been very kind to allocate some time for me whenever I needed it and supported me during the final stages of my studies.

I am greatly indebted to Philippe Renard for his kind invitation to the University of Neuchâtel to conduct my multiple-point statistical studies there. Our discussions and his endless assistance throughout my studies have substantially shaped this thesis. Without his help, this thesis would not have been possible. I am also grateful to Julien Straubhaar for helping me with scripting and algorithmic topics. He generously allocated many hours during my stay in Neuchâtel for discussions and assistance on multiple-point statistics.

I do thank Ute Mueller from Edith Cowan University for her comments and ideas on the potential avenues for future research. I also wish to thank Ilnur Minniakhmetov and Ryan Goodfellow from McGill university for providing me the hosc software and their kind assistance.

Lastly, I would like to thank my parents, brother and sister for their unconditional caring, love and motivation.

# Abstract

Mining operations taking place near the footwall contact of lateritic bauxite deposits incur significant costs due to unplanned dilution and ore losses. The major cause of these losses is the short-scale random fluctuations seen in the footwall topography. Economical drill spacing chosen to model the grade distributions is often large and inadequate to capture the spatial variations of the footwall contact. Consequently, ore body models created utilising sparse boreholes become smoother than the reality and incorporate substantial uncertainties. Under these circumstances, it becomes rather challenging to achieve planned mining targets, as accurate volume and tonnage calculations cannot be satisfactorily made.

The primary factors contributing to the uncertainties in the volume calculations can be divided into two. The first one of these is related to the models created for the footwall topography. Ore boundaries of the bauxite unit are mainly defined by the footwall contact representing the geological interface between the bauxite and ferricrete units. Since this contact is modelled using sparse boreholes, the created orebody models involve high uncertainties. The second uncertainty stems from the mining strategies. The extraction of a bauxite deposit is performed using a front-end loader type of mining equipment. Due to the short-scale irregularities of the ore boundaries, the mining equipment cannot accurately track down the ore boundaries during the excavations. Hence, the selectivity of the mining equipment used becomes an additional source for the uncertainties in the mined volumes.

This thesis aims at improving the current resource modelling practices of lateritic bauxite deposits. The issues mentioned above are attempted to be addressed

by (1) better modelling of the ore boundaries, (2) decision making under uncertainty and (3) accounting for the mining equipment selectivity. Potential modelling improvements were investigated through the adoption of a multiple-point statistical simulation framework. The motivation was the opportunity of utilising previously mined out areas to infer rich structural information on the contact variability. The bivariate training image constructed included mined-out topographies as well as the geophysical data (GPR). The multiple-point dependence between the variables of the training image was utilised by the Direct Sampling multiple-point statistical algorithm. As a result, a novel approach to variogram-free modelling of the ore boundaries is presented.

In order to make the modelling process more manageable and less labour intensive, a useful tool to automatically tune the input parameters of multiple-point statistical algorithms was also developed. The appropriate parameters were iteratively determined using the simulated annealing optimisation technique. The main objective was to maximise pattern reproduction of the conditioning data in the simulations. The advantages of the method include more objective and faster tuning of the input parameters.

The simulated orebody boundaries were used to serve as a model of uncertainty and incorporated into the decision-making process of creating optimum mining boundaries. To achieve this, a novel approach inspired by a model calibration technique called *Pilot Points* was utilised. The objective was to generate mining boundaries which minimise the economic losses due to dilution and ore losses in each of the realisations. In addition, the proposed method allowed adjusting the degree of smoothness of the created mining boundaries to account for the equipment flexibility. The comparison made with the mined-out surface demonstrated a significant reduction in the possible economical losses.

# List of Publications Included as Part of the Thesis

This thesis incorporates a number of papers that were published, submitted or in preparation at the time of thesis writing. Authorship statements regarding the papers are given in Appendix H and the published paper 1 is re-printed in its published form in Appendix J.

## Journal Papers

1. **Dagasan, Y.**, Erten, O., Topal, E. (2018). Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study. *The Journal of South African Institute of Mining and Metallurgy*, 118 (2), 173-184. DOI: 10.17159/2411-9717/2018/v118n12a11
2. **Dagasan, Y.**, Renard, P., Straubhaar, J., Erten, O., Topal, E. (2018) Pilot-point optimisation of mining boundaries for lateritic bauxite deposits: finding the trade-off between dilution and ore loss. *Natural Resources Research*, DOI: 10.1007/s11053-018-9380-9
3. **Dagasan, Y.**, Renard, P., Straubhaar, J., Erten, O., Topal, E. (2018) Automatic Parameter Tuning of Multiple-Point Statistical Simulations of Lateritic Bauxite deposits. *Minerals*, DOI: 10.3390/min8050220
4. **Dagasan, Y.**, Erten, O., Renard, P., Straubhaar, J., Topal, E. (2018) Multiple-point statistical simulation of the ore boundaries for the lateritic bauxite deposits (In Preparation)

# Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>List of Publications Included as Part of the Thesis</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.2 Motivation . . . . .	7
1.3 Research Questions, Aims and Objectives . . . . .	8
1.4 Thesis Outline . . . . .	10
References . . . . .	14
<b>2 Theoretical Foundations</b>	<b>17</b>
2.1 Introduction . . . . .	18
2.2 Geostatistics in Resource Estimation . . . . .	18
2.2.1 Regionalised Variables and Random Functions . . . . .	19
2.2.2 Stationarity . . . . .	19
2.2.3 Spatial Correlation and the Variogram . . . . .	21
2.2.4 Kriging . . . . .	23
2.2.5 Geostatistical Simulations . . . . .	26
2.3 Multiple-Point Statistics . . . . .	28
2.3.1 Higher-Order Spatial Statistics with Cumulants . . . . .	30
2.4 Summary and Discussion . . . . .	33

References . . . . .	34
<b>3 Weipa Bauxite Mine and Data Description</b>	<b>37</b>
3.1 Introduction . . . . .	38
3.2 Weipa Bauxite Mine . . . . .	38
3.2.1 Geology of the Mine Site . . . . .	38
3.2.2 Bauxite Mining Practices at Weipa . . . . .	42
3.3 Oak and Kumbur Mine Areas . . . . .	43
3.3.1 Data . . . . .	43
3.4 Exploratory Data Analysis . . . . .	46
3.5 Summary and Discussion . . . . .	50
References . . . . .	52
<b>4 Multiple-point statistical simulation of the ore boundaries for lateritic metal deposits</b>	<b>54</b>
4.1 Introduction . . . . .	55
4.2 Background . . . . .	55
4.3 Structural Analysis and Pre-Processing of the Data . . . . .	58
4.3.1 Structural Analysis and Trend Detection . . . . .	58
4.3.2 Migrating the Data Points to the Grid Nodes . . . . .	60
4.4 Setting up the MPS Simulations . . . . .	64
4.4.1 Choosing the Simulation Variables . . . . .	65
4.4.2 Algorithmic Input Parameter Selection . . . . .	66
4.5 Results and Comparison . . . . .	67
4.6 Summary and Discussion . . . . .	76
References . . . . .	78
<b>5 Automatic Parameter Tuning of Multiple-Point Statistical Simulations for Lateritic Bauxite Deposits</b>	<b>82</b>
5.1 Introduction . . . . .	83
5.2 Overview of Underlying Methods . . . . .	86
5.2.1 Direct Sampling Algorithm . . . . .	86



5.2.2	Comparing Patterns with Smooth Histograms . . . . .	88
5.2.3	Generalised Simulated Annealing . . . . .	90
5.3	Problem Setup and Methodology . . . . .	91
5.3.1	Methodology . . . . .	93
5.4	Results and Discussion . . . . .	96
5.4.1	Implementation and Analysis of the Tuned Parameters on the Simulations . . . . .	96
5.4.2	Effect of the Optimisation Method and the Initial Parame- ters on the Final Results . . . . .	101
5.4.3	Effect of the Chosen Parameters on the Pattern Reproduc- tion . . . . .	103
5.4.4	Analysis of the Automatically Tuned Parameters . . . . .	107
5.5	Summary and Conclusions . . . . .	108
	References . . . . .	111

**6 Pilot-point optimisation of mining boundaries for lateritic metal deposits: finding the trade-off between dilution and ore loss 116**

6.1	Introduction . . . . .	117
6.2	Review of underlying methods . . . . .	120
6.2.1	The Direct Sampling MPS Algorithm . . . . .	121
6.2.2	Pilot Point Method (PPM) . . . . .	122
6.2.3	Multilevel B-Splines . . . . .	122
6.2.4	Simulated Annealing (SA) Algorithm . . . . .	123
6.3	Methodology . . . . .	125
6.3.1	Simulations of the bauxite/ferricrete contact . . . . .	125
6.3.2	Locating pilot points . . . . .	126
6.3.3	Smooth excavation surface design . . . . .	128
6.3.4	Loss calculation-Objective Function . . . . .	129
6.3.5	Determination of optimum values at the pilot points . . . . .	131
6.4	Results and Discussion . . . . .	132
6.5	Summary and conclusions . . . . .	143
	References . . . . .	146

<b>7</b>	<b>Conclusions and Future Work</b>	<b>152</b>
7.1	Conclusions . . . . .	153
7.1.1	Contributions to the Current State of Knowledge . . . . .	154
7.2	Future Work . . . . .	155
<b>Appendix A</b>	<b>Simulation and TI Variables</b>	<b>160</b>
<b>Appendix B</b>	<b>Trend Surfaces and the Residuals</b>	<b>163</b>
<b>Appendix C</b>	<b>Variogram Models Fitted in Trend Free Directions</b>	<b>167</b>
<b>Appendix D</b>	<b>Residual Variograms and Models</b>	<b>171</b>
<b>Appendix E</b>	<b>Kumbur Simulation Results</b>	<b>175</b>
<b>Appendix F</b>	<b>Oak Simulation Results</b>	<b>182</b>
<b>Appendix G</b>	<b>Statistical and Structural Analyses Results</b>	<b>189</b>
<b>Appendix H</b>	<b>Declaration of Authorship</b>	<b>202</b>
<b>Appendix I</b>	<b>Copyright Licenses for typeset Chapters 5 and 6</b>	<b>204</b>
<b>Appendix J</b>	<b>Paper 1</b>	<b>208</b>

# List of Tables

3.1	Descriptive Statistics of the Variables for Oak and Kumbur Mine Areas . . . . .	49
4.1	Drift Coefficients Determined for the Oak and Kumbur Mine Variables	60
4.2	Optimised DS parameters used to simulate the position of the geological contact in Kumbur mine area . . . . .	67
4.3	DS parameters used to simulate the position of the geological contact in Oak and Kumbur mine areas . . . . .	68
4.4	Summary of the simulation results . . . . .	75
5.1	Initial and optimum parameters together with the associated JS divergence values. . . . .	99
6.1	Statistics of the proposed and mined out surfaces . . . . .	142
C.1	Variogram models used to migrate the GPR and floor surveys into the grid nodes . . . . .	170

# List of Figures

- 1.1 An example geological interface between the bauxite and ferricrete units . . . . . 3
- 1.2 Drill hole spacing used in resource estimation of laterites . . . . . 4
- 1.3 Flowchart of the thesis . . . . . 11
  
- 2.1 Examples of commonly used semi-variogram models . . . . . 22
  
- 3.1 Location of the Weipa bauxite mine. After Taylor et al. (2008) . . . 39
- 3.2 Geological units present at the Weipa bauxite mine. After Francke (2012) . . . . . 41
- 3.3 The test excavation conducted to reveal the characteristics of the lateral variability inherent in the geological interface (Erten, 2012) . 41
- 3.4 Front-end loader type of mining equipment used to extract the bauxite ore at the Weipa bauxite mine (Erten, 2012) . . . . . 43
- 3.5 Cat bottom-dump type truck used to haul the extracted bauxite material. The capacity is 150 tonne (Erten, 2012) . . . . . 44
- 3.6 Location maps for borehole data of the Oak and Kumbur mine areas. Circles represent the exploration borehole locations whereas ”+” signs represent the saturation drill hole locations . . . . . 45
- 3.7 Location maps of the GPR pick-points collected from the Oak and Kumbur mine areas . . . . . 46
- 3.8 Location maps of the survey pick-points collected after the extraction of bauxite from Oak and Kumbur mine areas (mined-out surfaces) 47
- 3.9 Histogram of the elevation data collected from Oak mine area. . . . 48

3.10	Histogram of the elevation data collected from Kumbur mine area. . . . .	48
3.11	Empirical cumulative distribution function for the elevation variables collected from Oak (a) and Kumbur (b) mine areas . . . . .	49
3.12	Q-Q plots of the Oak mine area variables. . . . .	51
3.13	Q-Q plots of the Kumbur mine area variables. . . . .	51
4.1	Omnidirectional variograms computed for the Oak and Kumbur mine areas . . . . .	59
4.2	Process used to both convert the point data into the gridded type and reconstruct the images . . . . .	62
4.3	(a) Experimental variogram calculated along the trend-free direction (N130) and (b) the variogram model fitted for the Kumbur GPR variable . . . . .	63
4.4	Re-construction of the GPR and floor surveys of the Oak and Kumbur mine areas . . . . .	64
4.5	Inputs and outputs of the bivariate DS simulations . . . . .	65
4.6	Bivariate TI, simulation variables and the omnidirectional residual variograms for the Kumbur mine area . . . . .	69
4.7	Experimental variogram of the PCD residuals along with the fitted model . . . . .	70
4.8	First three realisations of the DS and TB simulations . . . . .	70
4.9	Average of the TB and DS realisations together with the IQR distributions and the variograms of the realisations . . . . .	71
4.10	Cumulant maps of the Kumbur mined-out floor and the first realisations of the DS and TB Kumbur simulations . . . . .	73
4.11	Histograms of the mean cumulant errors for each realisation . . . . .	74
4.12	Summary of the performance indicators used to compare the TB and DS techniques for the Oak and Kumbur mine areas . . . . .	75

5.1	Simulation and the TI variables: <b>(a)</b> GPR variable used as the secondary simulation variable; <b>(b)</b> borehole data used to condition the simulations (primary variable-black dots overlying the GPR survey); <b>(c)</b> floor survey of a previously extracted mine area (primary TI variable) and <b>(d)</b> GPR data of the previously mined area collected prior to the extraction (Secondary TI variable). . . . .	92
5.2	Construction of the smooth histogram categories of the of borehole data. . . . .	94
5.3	Flowchart of the methodology. . . . .	97
5.4	Realisations produced with <b>(a)</b> the initial parameters and <b>(b)</b> final parameters. . . . .	99
5.5	<b>(a)</b> Initial average of the realisations; <b>(b)</b> final average of the realisations. Interquartile range maps: <b>(c)</b> before the parameter optimisation and <b>(d)</b> after the parameter optimisation. . . . .	100
5.6	Evolution of the distributions of the IQR values calculated at each grid node. Mean of the IQR values dropped from 0.60 to 0.47 in automatically found parameters. . . . .	101
5.7	Variograms of the realisations and the experimental variogram of the borehole data: <b>(a)</b> Variograms before the optimisation; <b>(b)</b> variograms after the optimisation and <b>(c)</b> 99.9% confidence interval of the variograms before and after the optimisation. . . . .	102
5.8	Convergence comparisons of three different optimisers with three different set of initial DS parameters. . . . .	103
5.9	Convergence graphs obtained by running 30,000 function calls. . . .	104
5.10	Profile JS divergences as a function of 5 DS input parameters. . . .	105
6.1	The peaks and troughs of the actual ore/waste interface cannot be detected by an economically viable drilling spacing. This results in an inaccurate estimation of the ore/waste contact (after Erten (2012)) . . . . .	118
6.2	The main steps of the proposed methodology. See text for detailed explanations. . . . .	125

6.3	Steps followed to locate the pilot points and assign their initial values in the mining area . . . . .	128
6.4	Thicknesses used to calculate the losses due to an excavation surface	130
6.5	Steps used to determine the pilot points yielding optimum mining boundaries . . . . .	131
6.6	Conditioning data used in the simulations. Black dots represent the borehole data locations (primary variable) and the underlying image represents the GPR data (secondary variable) . . . . .	133
6.7	The constructed bivariate TI: (a) extraction surface of a previously mined out area and (b) extensive GPR survey carried out prior to the excavation of the mining area . . . . .	134
6.8	Average of 200 realizations representing the bauxite/ferricrete interface . . . . .	135
6.9	Locations of the pilot points in the mine area. Blue dots represent the pilot points located at every (a) 8th grid node, (b) 16th grid node and (c) 24th grid node. The dashed lines represent the coordinates where the cross-sections were taken as presented in Figures 6.10 and 6.11 . . . . .	136
6.10	Y-Y cross-sections of the optimum surfaces found by using (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points . . . . .	137
6.11	X-X cross-sections of the optimum surfaces found by using (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points . . . . .	138
6.12	Plan view of the excavation surface designed using (a) 251 pilot points, (b) 60 pilot points, (c) 26 pilot points and (d) the actual mined-out surface. The dashed lines show the sections where the cross-sections in Figures 6.10 and 6.11 were constructed. . . . .	139
6.13	Histograms of the losses calculated using the (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points and (d) the actual mined out surface . . . . .	141

6.14	Barplots representing (a) the mined bauxite ore percentage of the expected reserve and (b) dilution percentages of the total mined volumes . . . . .	142
A.1	Variables used to simulate the Kumbur mine area: (a) Kumbur GPR survey used to guide the simulations as a secondary variable, (b) Kumbur PCD used to condition the simulations as a primary variable (circles overlain the GPR map), (c) Oak mined-out floor topography (primary variable of the TI) and (d) Oak GPR survey (secondary variable of the TI) . . . . .	161
A.2	Variables used to simulate the contact topography at Oak mine area: (a) Oak GPR data, (b) Oak borehole data overlain the GPR map (c) floor survey variable of the TI (Kumbur Mine) and (d) GPR component of the TI (Kumbur mine) . . . . .	162
B.1	Trends inferred using the conditioning data available for both Oak and Kumbur mine areas . . . . .	164
B.2	Kumbur GPR trend and residuals: (1) Kumbur GPR data, (2) second order trend for the Kumbur GPR and (3) Kumbur GPR residuals obtained by subtracting the trend from the raw data . . .	164
B.3	Kumbur floor trend and residuals: (a) Kumbur mined-out floor survey, (b) second order trend fitted to the floor survey and (c) residuals of the floor survey . . . . .	165
B.4	Oak floor trend and residuals: (a) Oak mined-out floor survey, (b) first order trend fitted to the floor survey and (c) residuals of the floor survey . . . . .	165
B.5	Oak GPR trend and residuals: (a) Oak GPR data, (b) first order trend fitted to the Oak GPR data and (c) the residuals of the GPR data . . . . .	166
C.1	Experimental variogram calculated along the trend-free direction (N130) and the fitted model for the Kumbur GPR variable . . . . .	168



C.2	Experimental variogram calculated along the trend-free direction (N165) and the fitted model for the Kumbur floor survey variable . . . . .	168
C.3	Experimental variogram calculated along the trend-free direction (N20) and the fitted model for the Oak GPR variable . . . . .	169
C.4	Experimental variogram calculated along the trend-free direction (N30) and the fitted model for the Oak floor survey variable . . . . .	169
D.1	Variograms of the variables used to simulate the Kumbur mine (residuals) . . . . .	172
D.2	Variograms of the variables used to simulate the Oak mine (residuals)	172
D.3	Variogram models fitted to the PCD residuals for the univariate and bivariate Kumbur TB simulations. The difference in the experimental variograms is due to the differences in trends inferred using the borehole data for the univariate simulations and the GPR data (external drift) for the bivariate simulations. No anisotropy was detected for the univariate case. Therefore, only the omnidirectional variogram was modelled . . . . .	173
D.4	Variogram models fitted to the borehole residuals for the univariate and bivariate Oak TB simulations. As in Kumbur mine variables, the trend for the univariate simulations was inferred from the borehole and the trend for the bivariate simulations was inferred from the GPR as an external drift . . . . .	174
E.1	First four DS realisations of the Kumbur univariate simulations . . . . .	176
E.2	First four TB realisations of the Kumbur univariate simulations . . . . .	177
E.3	First four DS realisations of the Kumbur bivariate simulations . . . . .	178
E.4	First four TB realisations of the Kumbur bivariate simulations . . . . .	179
E.5	Average of 40 univariate Kumbur realisations and the IQR maps . . . . .	180
E.6	Average of 40 bivariate Kumbur realisations and the IQR maps . . . . .	181
F.1	First four DS realisations of the Oak univariate simulations . . . . .	183
F.2	First four TB realisations of the Oak univariate simulations . . . . .	184
F.3	First four DS realisations of the Oak bivariate simulations . . . . .	185

F.4	First four TB realisations of the Oak bivariate simulations . . . . .	186
F.5	Average of 40 univariate Oak realisations and the IQR maps . . . . .	187
F.6	Average of 40 bivariate Oak realisations and the IQR maps . . . . .	188
G.1	Reproduction of the experimental quantiles for Kumbur simulations	190
G.2	Reproduction of the experimental quantiles for Oak simulations . . .	191
G.3	Reproduction of the variograms for the Kumbur mine simulations .	192
G.4	Reproduction of the variograms for the Oak mine simulations . . .	193
G.5	Distribution of the interquartile ranges of the simulated elevation values at each grid node . . . . .	194
G.6	Distribution of the interquartile ranges of the simulated elevation values at each grid node . . . . .	195
G.7	Cumulant maps of the Kumbur mined-out floor (reference image) and the first 3 DS and TB univariate Kumbur realisations . . . . .	196
G.8	Cumulant maps of the Kumbur mined-out floor (reference image) and the first 3 DS and TB bivariate Kumbur realisations . . . . .	197
G.9	Cumulant maps of the Oak mined-out floor (reference image) and the first 3 DS and TB univariate Oak realisations . . . . .	198
G.10	Cumulant maps of the Oak mined-out floor (reference image) and the first 3 DS and TB bivariate Oak realisations . . . . .	199
G.11	Distributions of the cumulant errors for the univariate Kumbur sim- ulations . . . . .	200
G.12	Distributions of the cumulant errors for the univariate Oak simulations	201



# Chapter 1

## Introduction

## 1.1 Background

Estimation of recoverable resources and reserves is one of the major tasks to assess the feasibility of a mining project (Lipton et al., 1998; Abzalov, 2016). The primary goal of this work is to predict the tonnage and average grades of a mineral deposit that will be exploited during a specific period (Rossi and Deutsch, 2013). Such a goal can be achieved by utilising a variety of geostatistical techniques to model the spatial variability of an attribute within a mineral deposit. However, as the models constructed using these techniques are often based on sparsely-spaced borehole data, uncertainties associated with them are unavoidable (Rossi and Deutsch, 2013). These uncertainties can pose significant risks in the economic viability of the mining projects if they are not adequately assessed. Therefore, integration of these uncertainties in the financial assessment and decision-making processes can benefit the mine asset values significantly (Dimitrakopoulos et al., 2002).

Uncertainties can stem from the lack of knowledge and incomplete understanding of an attribute to be quantified (Caers, 2011). Considering the mineral resource estimation, the uncertainties can be due to the high natural variability of the attributes inherent to the geology such as grades or deposit geometries (Bardossy et al., 2003). Another source of uncertainty can be one related to the mineral exploration practices. Sampling errors, insufficient sampling, lack of geological experience and uncertainties in the models generated during the mineral exploration of a deposit also contribute to the uncertainties (Bardossy et al., 2003).

For the Weipa lateritic bauxite mine located in northern Queensland, Australia, modelling of the geological contact between the bauxite and ferricrete sections is rather a crucial task and incorporates both in-situ and operation-related uncertainties that need to be accounted for. The geology of the mine site is comprised of stratified geological layers and the bauxite ore exists within the soil horizons (Erten, 2012; Abzalov and Bower, 2014). The ore is covered by a topsoil layer and underlying the bauxite unit there exists an iron and silica-rich material called *ferricrete* (or locally known as ironstone). The formation process of the deposit comprises the intensive in-situ weathering/leaching of kaolinite, quartz and iron

oxide minerals (Loughnan and Bayliss, 1961; Grubb, 1971). Due to the weathering process involved in the formation of the deposit, the geological contact between the bauxite and ironstone units is rather uneven (Abzalov and Bower, 2009). The large lateral variability evident in the geological contact exhibits almost random frequencies, being the major source of the uncertainties in the estimation of the bauxite ore tonnage. An example of such a contact can be seen in Figure 1.1.

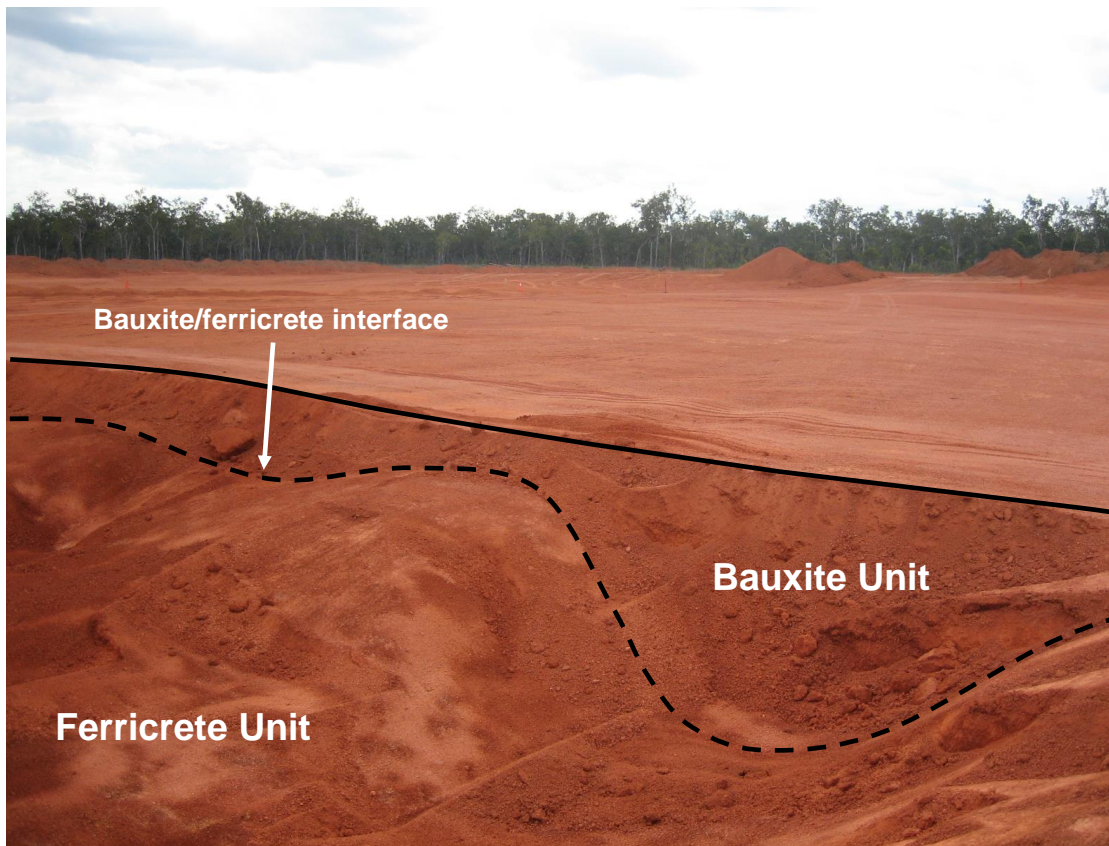


Figure 1.1: An example uneven surface illustrating the undulating contact topography between bauxite and ironstone units at the Weipa bauxite mine (after Erten (2012))

Resource estimation of a lateritic bauxite deposit is performed using sparsely-spaced boreholes drilled on a regular grid within the estimation domain (Erten, 2012; Abzalov and Bower, 2014). The chosen spacing between these boreholes is based heavily on the variation and continuity in grades (Hartman and Mutmanský, 2002). Hence, the defined drill spacing is sufficient to model the aluminium grade

( $\text{Al}_2\text{O}_3\%$ ) within the bauxite unit. However, as the geological interface between the bauxite and ironstone units is rather uneven, the chosen drill spacing fails to adequately infer the lateral variability in the footwall contact topography (Erten, 2012). Therefore, contact models created using interpolation techniques, such as the ones offered by geostatistics, become less detailed and smoother, as expected, than the reality. This situation leads to an imprecise resource estimation as well as volumetric uncertainties due to poorly defined deposit boundaries (Singh, 2007). Consequences of such uncertainties mainly include inaccurate quantification of the dilution, ore loss and total mined bauxite volumes. Figure 1.2 illustrates the underlying process of the unplanned dilution and ore losses.

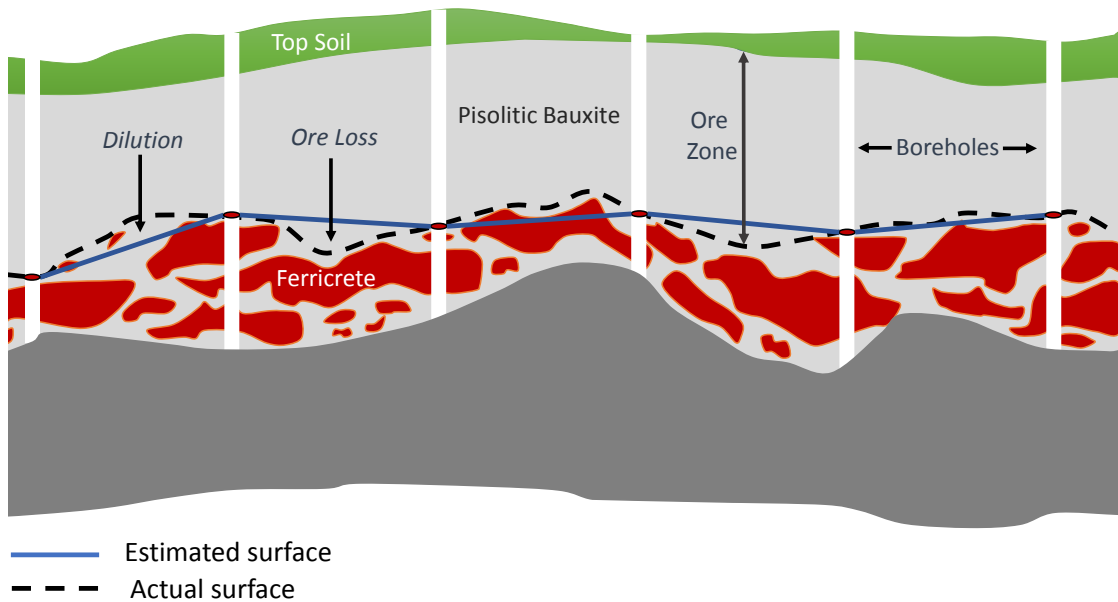


Figure 1.2: Borehole spacing determined based on the grade continuity cannot capture the lateral variability in the bauxite/ferricrete interface. Therefore, the modelled geological contact results in unplanned dilution and ore losses (after Erten (2012))

Mining strategies used to exploit the bauxite deposit also yield additional uncertainties in the estimated volumes. That is, a large front-end loader cannot track the actual ore/waste boundaries during the mining activity, which results in operational dilution. The bauxite layer at the Weipa mine is of pisolitic type and has free flow characteristics. Unlike many other mining methods, mining of the

bauxite deposit at the Weipa mine does not require any blasting operation since the entire deposit consists of loose pisolites. Therefore, extraction of the deposit is performed by a large front-end loader. Since the geological contact modelled during the resource estimation stage is poor and incorporates excessive uncertainties due to the use of sparse borehole data, the equipment operator is not given any prior information (i.e. spatial coordinates that can be uploaded to the GPS mounted on the front-end loader) related to the contact. To extract the bauxite ore, the operator primarily makes use of the hardness and colour differences between the bauxite and ferricrete units to follow the boundary. In other words, the decision on the ore boundaries has to be made by the operator at the time of mining. Furthermore, the digging ability of the mining equipment restricts tracking the actual interface due to substantial small-scale fluctuations apparent in the geological boundary. To sum up, subjective decisions of the equipment operator at the time of mining and the limited equipment selectivity cause critical operational uncertainties in resource estimations.

Accurate quantification of the dilution is critical to enhancing the profitability of the bauxite mining operations. Mineral processing of the bauxite is highly sensitive to the chemical composition of the ore being fed to the plant (Erten, 2012). Even a minor increase in the silica content of the ore results in a significant rise in the caustic soda consumption. Being a significantly expensive chemical compound, the quantity of the caustic soda used to treat the ore profoundly affects the processing costs, which in turn impacts on the overall mining costs. At the Weipa bauxite mine, for instance, the cost of one unit dilution is approximately 60 times more expensive than one unit ore loss (Erten, 2012). As silica is highly abundant in the Ferricrete unit, dilution of the Ferricrete unit acts as a major source of the silica in the mined volume. Therefore, accurate modelling of the contact between the bauxite and ferricrete as well as proper quantification of uncertainties is essential for risk management.

Mining strategies or resource estimation techniques allowing the incorporation and better quantification of the uncertainties can significantly benefit the management of the unplanned ferricrete dilutions. One way of reducing the uncertainties



in the bauxite base floor models would be the implementation of dense drilling applications. However, this solution would significantly increase the associated exploration costs, making the operation less profitable or not profitable at all. Another approach could be the use of geophysical methods, such as ground penetrating radar (GPR), to assist delineating the geological contact. GPR surveys can provide exhaustive secondary information about the location of the geological contact, and they are rather cost effective as compared to the drilling of additional exploration boreholes. Integration of GPR in contact modelling has demonstrated benefits in achieving better contact models in several applications (Erten, 2012; Erten et al., 2013, 2015). However, even though a better model is attained using the GPR survey to assist the modelling, the large spatial variations of the models attained cannot be accurately mined out due to the limited selectivity of the mining equipment. Therefore, any excavation surface inevitably causes dilution and ore losses that need to be accounted for. Even though this problem exhibits a resemblance with the dig-limit problems in open-pit mining, which has been a topic for a number of research studies such as Norrena and Deutsch (2001, 2002); Richmond (2004); Richmond and Beasley (2004); Isaaks et al. (2014); Ruiseco et al. (2016); Ruiseco and Kumral (2017); Sari and Kumral (2018), the problem encountered in lateritic bauxite deposits is different because of the free-digging mining method used. The geological contacts in lateritic bauxite are in the form of contact topographies and can be modelled using their elevation attributes. Therefore, determination of the optimum mining boundaries is achieved by finding the optimum mining elevation values. Research on determining the optimum mining elevations was conducted by McLennan et al. (2006) to optimise the dilution and ore losses in a lateritic nickel mine. However, due to the existence of good equipment selectivity, a special emphasis on the equipment flexibility was not vital in that study.

## 1.2 Motivation

Extraction of a bauxite ore at a particular location of the mine site exposes a mined-out surface that can be deemed representative of the actual geological interface. Given that the spatial variations of an attribute within a geological domain exhibit prominent similarities, mined-out areas can provide substantial prior knowledge about the structural information inherent in a future mining area. Therefore, it is possible to infer the spatial statistics of a future mining area without relying on the limited structural information derived from the sparse drill hole data. Such an inference can be done by the use of multiple-point statistics and can benefit the resource estimation in several aspects: (1) variogram modelling is not needed as the structural information is borrowed from a training image, (2) structural information that cannot be derived from the sparse drill hole can be retrieved from the mined out areas, (3) it allows building a catalogue of training images from which the most suitable training image can be chosen for a particular area. These advantages make it rather worthy to explore the potential benefits of what multiple-point statistics has to offer.

One common characteristic of the multiple-point statistical algorithms is the requirement for algorithmic input parameters to perform the simulations. The statistics of the generated models are highly dependent on the input parameters as well as the training image chosen. Once a decision on the training image has been made, the model qualities then become highly dependent on the input parameters. A sensitivity analysis comprising manual tuning of these parameters is the common approach to determining the parameters. However, this is a rather cumbersome task and requires one to devote significant time to discover the appropriate parameters. Therefore, a methodology for automatically determining the parameters is needed to reduce the time spent in the parameter tuning process.

Geological contact modelling in lateritic bauxite deposits poses significant challenges and incorporates uncertainties which stem from different sources. One can generate a model of uncertainties for the contact boundary using stochastic realisations. Although the generated realisations provide a set of possible scenarios which allow the quantification of the uncertainties, they cannot be directly used to

determine which boundaries to excavate. Therefore, there exists an apparent need for an optimum mining boundary that will be used to both guide the operators and to calculate the accurate volumes to be mined. Such a boundary should (1) lead to the reduction in economical losses due to dilution and ore loss, (2) incorporate the model of uncertainties and (3) have an adjustable degree of smoothness to increase the equipment flexibility.

## 1.3 Research Questions, Aims and Objectives

### Research questions

This thesis investigates approaches to improve the current ore boundary modelling practices for lateritic bauxite deposits. It attempts to address the following three major research questions:

1. Can the current geological contact models be improved by incorporating the MPS in the modelling framework?
2. How can the appropriate algorithmic input parameters of MPS algorithms be determined more efficiently? Can the parameter tuning process be automated so that faster tuning is achieved as well as improving the simulation qualities?
3. How can the optimum mining boundaries be obtained that:
  - are less detailed than the actual ore boundaries so that it is easier to excavate with front-end loader type mining equipment,
  - minimise the economic losses due to dilution and ore loss, and
  - incorporate the modelled uncertainty in the decision making process.

### Research aims and objectives

There are three main aims of this thesis. The first one is related to the spatial modelling of the geological interface utilising the previously mined-out areas as

analogue sites (which provide prior knowledge on the geology). Within the scope of this aim, the use of the Direct Sampling (DS) multiple-point statistical (MPS) algorithm and its potential benefits are explored in the thesis. Objectives set to achieve this aim are as follows:

1. Developing an overall framework to carry out the MPS simulations for the lateritic metal deposits.
2. Construction of the univariate and bivariate training images.
3. Setting up of the DS algorithm to perform the simulations.
4. Setting up of the Turning Bands (TB) geostatistical simulation technique to perform the simulations.
5. Comparing the DS simulation performances with the TB simulations using statistical and geostatistical performance indicators.

The second aim is to develop an approach which automatically determines the algorithmic input parameters of the DS algorithm. The following objectives were set for this aim of the thesis:

1. Determining a suitable technique to quantify the multiple-point statistics of the borehole data and the realisations.
2. Determining a metric to quantify the mismatch between the multiple-point statistics of the conditioning data and the realisations.
3. Setting up an optimisation framework to iteratively find the DS parameters. The decision variables to be used are the DS input parameters, and the objective function to be minimised is the mismatch between the input-output pattern statistics.
4. Determination of an optimisation method which efficiently approximates the global minimum of an objective function with many local minima.

The last aim of the thesis is about developing a methodology to design a smooth mining surface which minimises the dilution and ore losses as well as increasing the equipment flexibility. The following objectives were set to achieve this aim of the thesis:

1. Defining a loss function to calculate the expected losses of a given decision surface in all the contact realisations.
2. Determining an interpolation technique to construct the mining boundaries using a set of control points (pilot points). The chosen technique should allow adjusting the smoothness of the resulting surface to also incorporate the equipment flexibility.
3. Setting up of an optimisation framework to construct the optimum mining boundaries. The objective of the optimisation is to minimise the economical losses due to dilution and ore losses in each of the realisations. The decision variables comprise the pilot point values which determine the shape of the mining boundary.

## 1.4 Thesis Outline

The thesis presents several approaches related to the geological interface modelling of lateritic metal deposits. The relationship between the chapters is illustrated in Figure 1.3 and the content of the chapters are as follows:

### Chapter 2

Chapter 2 provides a comprehensive background on the methods which are used in this thesis. These methods will be referred to in the approaches presented in Chapter 4, 5 and 6.

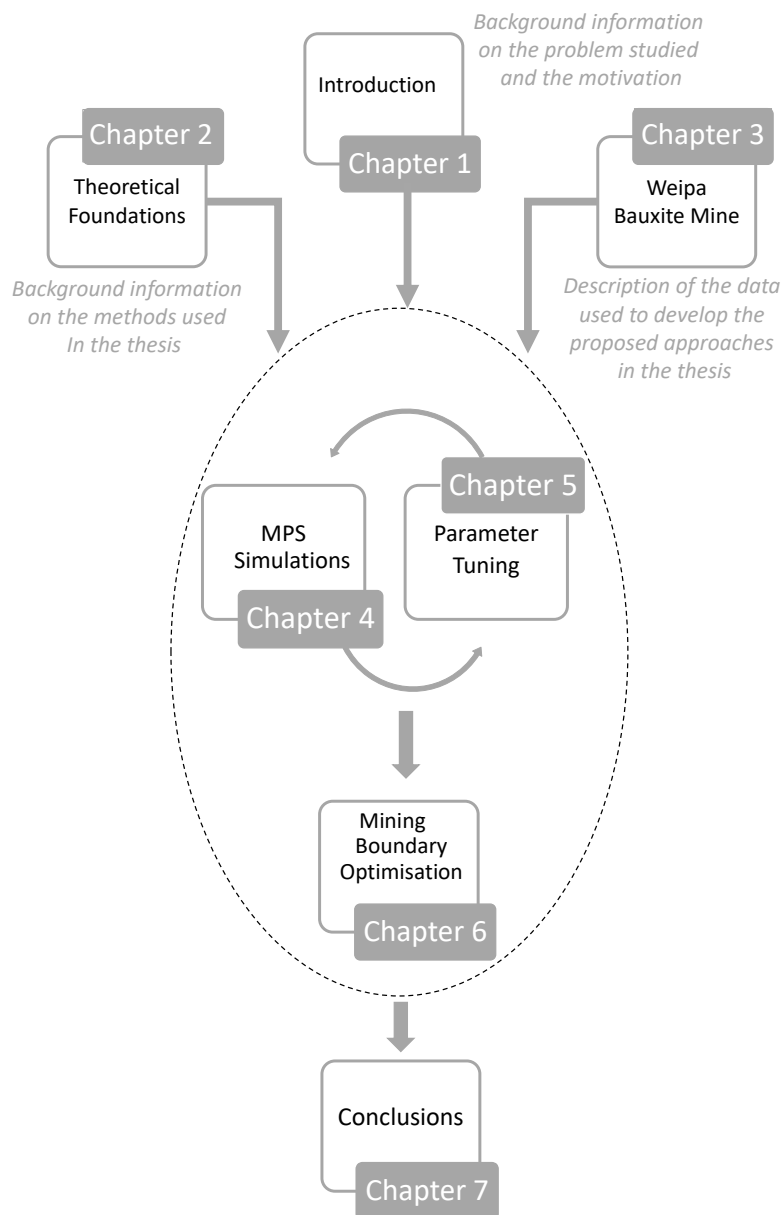


Figure 1.3: Flowchart of the thesis

## Chapter 3

Chapter 3 introduces the data used to develop the techniques presented in this thesis. The data comes from two mine areas within the Weipa bauxite mine in northern Queensland, Australia and it comprises borehole, GPR and the mined-out

floor survey data. Their exploratory data analyses are also given in this chapter. The data is used to perform the bivariate and univariate simulations for the elevation variable of the ore boundaries (footwall contact topography) in Chapter 4 and 5.

## **Chapter 4**

Chapter 4 presents the overall framework to carry out the geological contact simulations using multiple-point statistics. The step-by-step description of the methodology covers the data preparation as well as setting up the DS simulations in detail. The contact realisations created in this chapter are used as a model of uncertainty to create optimum mining boundaries in Chapter 6.

## **Chapter 5**

Chapter 5 presents a new strategy to automatically determine the algorithmic input parameters of the Direct Sampling algorithm. The automatically determined parameters are then used to perform the simulations in Chapter 4. Note that the process of tuning the parameters required for the MPS simulations is presented in Chapter 5 after the MPS simulations are presented in Chapter 4. This is because setting up the MPS simulations is required to both perform the simulations and tune the parameters. The mentioned relationship between the chapters can be seen in Fig. 1.3.

## **Chapter 6**

Chapter 6 presents a methodology inspired by a calibration technique called pilot points to construct optimum mining boundaries. The proposed methodology allows constructing an optimum mining surface using the model of uncertainties created in Chapter 4. It also allows the integration of the digging ability of the mining equipment through the degree of smoothness of the mining boundaries created.

The work done in Chapters 4 and 5 was utilised in Chapter 6, therefore, it is the final chapter before the conclusions of the thesis.

## **Chapter 7**

This chapter presents the concluding remarks and discussions on the advantages and drawbacks of all the components of the research. Recommendations for future work are also given here.



## References

- Abzalov, M. (2016). *Applied mining geology*, volume 12. Springer.
- Abzalov, M. and Bower, J. (2009). Optimisation of the drill grid at the weipa bauxite deposit using conditional simulation. In *Seventh international mining geology conference. AusIMM, Melbourne*, pages 247–251.
- Abzalov, M. and Bower, J. (2014). Geology of bauxite deposits and their resource estimation practices. *Applied Earth Science*, 123(2):118–134.
- Bardossy, G., Szabo, I., and Varga, G. (2003). A new method of resource estimation for bauxite and other solid mineral deposits. *BERG UND HUTTENMANNISCHE MONATSHEFTE*, 148(2):57–64.
- Caers, J. (2011). *Modeling uncertainty in the earth sciences*. John Wiley & Sons.
- Dimitrakopoulos, R., Farrelly, C., and Godoy, M. (2002). Moving forward from traditional optimization: grade uncertainty and risk effects in open-pit design. *Mining Technology*, 111(1):82–88.
- Erten, O. (2012). Profiling and mining control to mitigate dilution effect from sio<sub>2</sub> at the base of a bauxite deposit.
- Erten, O., Kizil, M. S., Topal, E., and McAndrew, L. (2013). Spatial prediction of lateral variability of a laterite-type bauxite horizon using ancillary ground-penetrating radar data. *Natural resources research*, 22(3):207–227.
- Erten, O., McAndrew, L., Kizil, M. S., and Topal, E. (2015). Incorporating fine-scale ground-penetrating radar data into the mapping of lateral variability of a laterite-type bauxite horizon. *Mining Technology*, 124(1):1–15.
- Grubb, P. (1971). Genesis of the weipa bauxite deposits, ne australia. *Mineralium Deposita*, 6(4):265–274.
- Hartman, H. L. and Mutmanský, J. M. (2002). *Introductory mining engineering*. John Wiley & Sons.

- Isaaks, E., Treloar, I., and Elenbaas, T. (2014). Optimum dig lines for open pit grade control. In *9th International Mining Geology Conference, Adelaide, South Australia*.
- Lipton, I., Gaze, R., Horton, J., and Khosrowshahi, S. (1998). Practical application of multiple indicator kriging and conditional simulation to recoverable resource estimation for the halley's lateritic nickel deposit. In *Proceedings of A 1 Day Symposium, Perth*, pages 88–105.
- Loughnan, F. and Bayliss, P. (1961). The mineralogy of the bauxite deposits near weipa, queensland. *American Mineralogist*, 46(1-2):209–217.
- McLennan, J., Ortiz, J., and Deutsch, C. (2006). Geostatistical simulation of optimum mining elevations for nickel laterite deposits. *CIM Magazine*, 1(6).
- Norrena, K. and Deutsch, C. (2001). Determination of dig limits subject to geostatistical, economic, and equipment constraints. In *2001 SME Annual Conference and Exhibition*, pages 133–148.
- Norrena, K. and Deutsch, C. (2002). Optimal determination of dig limits for improved grade control. In *APCOM 2002: 30th International Symposium on the Application of Computers and Operations Research in the Mineral Industry*, pages 329–339.
- Richmond, A. (2004). Integrating multiple simulations and mining dilution in open pit optimisation algorithms. In *Orebody Modelling and Strategic Mine Planning Conference*.
- Richmond, A. and Beasley, J. (2004). Financially efficient dig-line delineation incorporating equipment constraints and grade uncertainty. *International Journal of Surface Mining*, 18(2):99–121.
- Rossi, M. E. and Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.

- Ruiseco, J. R. and Kumral, M. (2017). A practical approach to mine equipment sizing in relation to dig-limit optimization in complex orebodies: multi-rock type, multi-process, and multi-metal case. *Natural Resources Research*, 26(1):23–35.
- Ruiseco, J. R., Williams, J., and Kumral, M. (2016). Optimizing ore–waste dig-limits as part of operational mine planning through genetic algorithms. *Natural Resources Research*, 25(4):473–485.
- Sari, Y. A. and Kumral, M. (2018). Dig-limits optimization through mixed-integer linear programming in open-pit mines. *Journal of the Operational Research Society*, 69(2):171–182.
- Singh, N. (2007). Ground penetrating radar (gpr) in mineral base profiling and orebody optimization. In *6th International Heavy Minerals Conference*, pages 185–194.

## **Chapter 2**

# **Theoretical Foundations**

## 2.1 Introduction

This chapter provides an overview of the methods used to investigate potential improvements in resource estimation of lateritic metal mines. Since Chapters 5 and 6 of the thesis are publications and written in article formats, some of the methods are already presented in these chapters. Therefore, to avoid any repetition, methods explained in those chapters are not again presented in Chapter 2. In other words, Chapter 2 presents methods which are not given in Chapters 4, 5 and 6.

The first aim of the thesis is to explore the potential benefits of using multiple-point statistics in geological contact modelling (covered in Chapter 4). Therefore, Chapter 2 starts with background information on geostatistics and available modelling techniques (estimation and simulation). This is then followed by describing the variogram related limitations of geostatistics and how MPS can be used to address such issues through TIs. The multiple-point statistical simulations based on a TI are demonstrated using the Direct Sampling algorithm in Chapters 5 and 6, which are based on the publications. Therefore, the Direct Sampling method was not included in this chapter. Similarly, the methods, such as smooth histograms, Jensen-Shannon divergence, spatial cumulants, pilot points, simulated annealing and multi-level B-spline methods, are originally presented in the publications given in Chapters 5 and 6; hence, they are not included in this chapter either.

## 2.2 Geostatistics in Resource Estimation

Initial knowledge of geological features such as grades, tonnages or boundaries of a deposit is often acquired through a limited number of samples collected from different parts of the ore body. Therefore, these attributes need to be estimated at locations where no samples are collected. Geostatistics provides a framework to perform the spatial modelling of these attributes by treating them as spatially correlated random variables. Assumptions and tools used in geostatistical modelling are explained in the following subsections.

### 2.2.1 Regionalised Variables and Random Functions

Geostatistics treats the measurable quantities of a mineralised phenomenon as *regionalised variables (ReV)* (Journel and Huijbregts, 1978). Given a mineral deposit, for instance, a ReV can be a grade, deposit thickness or the elevation variable of a foot-wall topography. A set of samples  $\{z(x_i), i = 1, \dots, N\}$  collected from  $x_i$  locations of a domain  $D$ , hence, represent a subset of ReV (Wackernagel, 2013). A ReV is considered to be the outcome of a random draw from a location-specific probability distribution. That is to say, at each of infinitely many  $x$  locations of a domain  $D$ , there exists a random variable  $Z(x)$  that can be described with a mean  $m$  and variance  $\sigma^2$  (Oliver, 2010). An RV can also be characterised by a certain distribution whose cdf can be defined as in the following:

$$F_x(z) = Prob\{Z(x) \leq z\} \quad (2.1)$$

where  $F$  defines the probability of the outcome  $Z$  of the RV to be lower than a specific  $z$  value. A family of all the random variables within a domain is called a *random function (RF)* (Armstrong, 1998). An RF  $Z(x)$  can be decomposed into components which are a deterministic mean  $m$  (drift) and a stochastic part fluctuating around the mean (residual)  $R(x)$ :

$$Z(x) = m(x) + R(x) \quad (2.2)$$

### 2.2.2 Stationarity

The term stationarity refers to the assumption that the data over a region exhibits the same degree of variation, and its distribution is invariant under translation  $h$  (Oliver and Webster, 2015). It can be considered as a decision made for the choice of the RF and allows one to pool the data over an area (Wackernagel, 2013). The strict version of the stationarity requires all the moments of the RV to be translation invariant (Armstrong, 1998). That is, the multivariate CDF within an

area should stay the same under any translation, which is called *strict stationarity*:

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = F_{x_1+h, \dots, x_n+h}(z_1, \dots, z_n) \quad (2.3)$$

However, due to the limited number of samples available, inference of multivariate cdf is usually not practical and such an assumption cannot be verified. Therefore, a weaker form of stationarity assumption called *second order stationarity* is often employed. Second order stationarity assumes that the expected value of  $E[Z(x)] = m$ , which is the mean, is constant throughout a given area. Moreover, the covariance function, which can be used to characterise the spatial variation, should only depend on the vector  $h$  rather than the coordinates:

$$\begin{aligned} C(h) &= E[\{Z(x) - m(x)\}\{Z(x+h) - m(x)\}] \\ &= E[\{Z(x)Z(x+h) - m(x)\}] \end{aligned} \quad (2.4)$$

In situations where the mean varies rather than being constant, the covariance function cannot be calculated. Therefore, a weaker form of stationarity called *intrinsic stationarity* can be used. This form of stationarity is based on the increments  $Z(x+h) - Z(x)$  and requires the expected value of the increments to be equal to 0.

$$E[Z(x) - Z(x+h)] = 0 \quad (2.5)$$

Intrinsic hypothesis uses the variance of the differences to characterise the behaviour of the spatial variations, rather than the covariance of the residuals. Similar to the second order stationarity, the variances should depend on the lag distance  $h$  rather than the  $x$  locations:

$$\text{var}[Z(x) - Z(x+h)] = E[\{Z(x) - Z(x+h)\}^2] = 2\gamma(h) \quad (2.6)$$

$\gamma(h)$  represents the value of the semivariance at lag distance  $h$ . In situations where the second order stationarity exists, the covariance and variogram become

equivalent:

$$\begin{aligned}
\gamma(h) &= \frac{1}{2}E\{[Z(x) - Z(x+h)]^2\} \\
&= \frac{1}{2}E[Z(x)^2] + \frac{1}{2}E[Z(x+h)^2] - E[Z(x)Z(x+h)] \\
&= E[Z(x)^2] - E[Z(x)Z(x+h)] \tag{2.7} \\
&= \text{Var}[Z(x)] - \{E[Z(x)Z(x+h)]E[Z(x)]E[Z(x+h)]\} \\
&= C(0) - C(h)
\end{aligned}$$

However, the opposite is not true; the variogram may not be equal to the covariance if the covariance does not exist.

### 2.2.3 Spatial Correlation and the Variogram

The stochastic component  $R(x)$  of the RF exhibits a structured random behaviour around the drift  $m(x)$  (Armstrong, 1998). Geostatistical interpolation techniques require this structural information be derived to perform estimations at unsampled locations of a domain. Although both semi-variograms and covariances can be used to express the spatial variability as a function of the separation distance  $h$  between the samples, semi-variograms are frequently used as a measure of spatial correlation. Given sample data  $\{z(x_i), i = 1, \dots, n\}$ , the experimental semi-variogram can be calculated using the average of squared differences between data pairs:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{2.8}$$

where  $N(h)$  is the total number of pairs separated by a lag distance  $h$ . Once the experimental variogram was constructed using a set of  $h$  lag distances, the next step is then to fit one or more authorized models to characterise the spatial variability. Examples of commonly used authorised variogram models include spherical, power, exponential and Gaussian (Norrena, 2007), as can be seen in Fig. 2.1.

Variogram models are usually controlled by the *range* and *sill* parameters. The sill parameter refers to the variance  $C$  of the random process whereas the



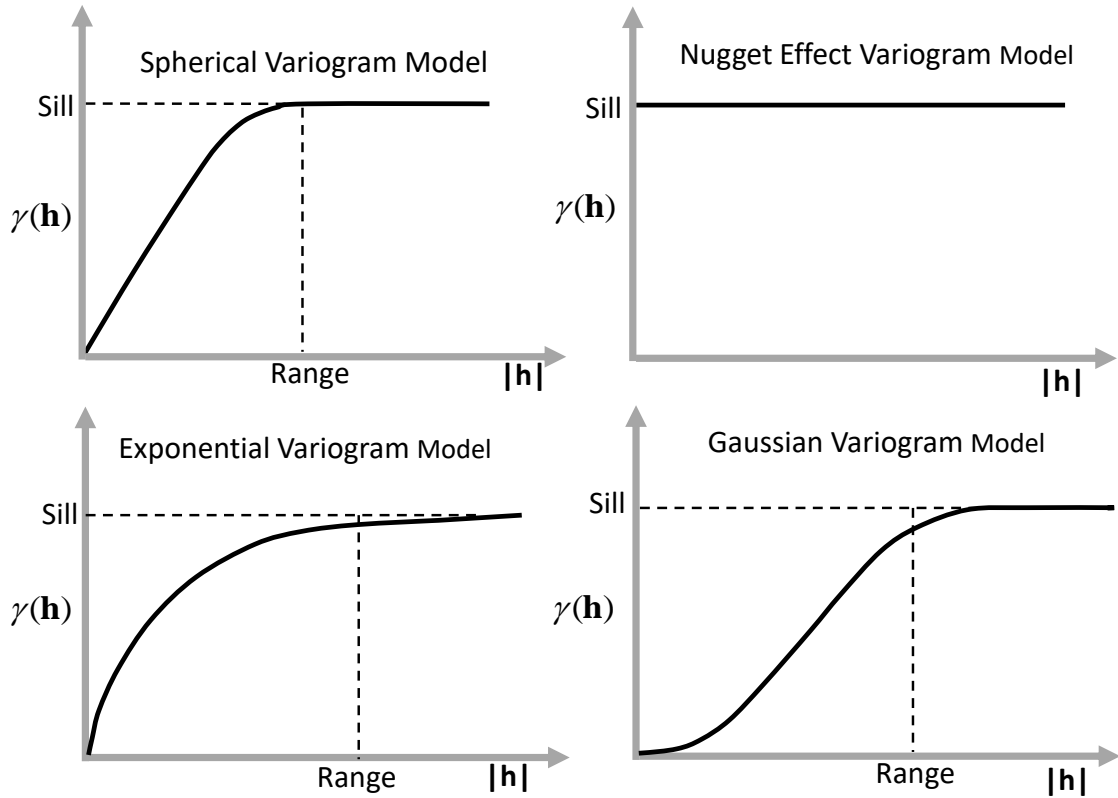


Figure 2.1: Examples of commonly used semi-variogram models to fit an experimental variogram (After Norrena (2007))

range  $a$  corresponds to the lag distance beyond which no spatial correlation exists. These variograms can be expressed as following:

### Spherical Model

Spherical function is commonly used in geostatistics and can be defined by the range and the sill values:

$$\gamma(h) = \begin{cases} C \left( 1.5 \frac{|h|}{a} - 1.5 \frac{|h|^3}{a^3} \right) & |h| \leq a \\ C & |h| > a \end{cases} \quad (2.9)$$

### Nugget Effect Model

Also called pure nugget, the nugget effect model implies no spatial correlation between the samples:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C & |h| > 0 \end{cases} \quad (2.10)$$

### Exponential Model

The exponential model is defined as:

$$\gamma(h) = C(1 - e^{-\frac{|h|}{r}}) \quad (2.11)$$

where  $r$  represents a distance parameter. In this model, a finite range cannot be defined as the model approaches the sill asymptotically. Therefore, the lag distance value corresponding to the 95<sup>th</sup> percentile of the sill variance is considered as the practical range, also called the *effective range* and corresponds to the value of  $3r$ .

### Gaussian Model

Similar to the exponential model, this function approaches the sill asymptotically and the lag distance value corresponding to the 95<sup>th</sup> percentile of the sill becomes  $\sqrt{3}r$ . The function defining the model is described as follows (Wackernagel, 2013):

$$\gamma(h) = C(1 - e^{-\frac{|h|^2}{r^2}}) \quad (2.12)$$

## 2.2.4 Kriging

Kriging is a geostatistical estimation technique which was first put in a framework by Matheron et al. (1965). It allows the estimation of an attribute at an unsampled location as a linear combination of available neighbouring data (Goovaerts, 2000). Given a set of samples  $\{z(x_i), i = 1, \dots, n\}$ , for instance, the estimation at an  $x$  location of a domain is performed using the weighted average of the samples:

$$z^*(x) = \sum \lambda_i z(x_i) \quad (2.13)$$

where  $\lambda_i$  represents the weight assigned to the random variable  $z(x_i)$ . The weights in kriging are chosen based on two criteria. First, the estimation should be unbiased, that is, the difference between the mathematical expectation and the true value should be equal to zero:

$$E[Z^*(x) - Z(x)] = 0 \quad (2.14)$$

Second, the prediction variance  $Var[Z^*(x) - Z(x)]$ , also called the kriging variance, should be the minimum. Due to these characteristics of the estimations, kriging is also called the best linear unbiased estimator (BLUE).

Although there are various types of kriging techniques suitable for various applications, commonly used ones which are *Ordinary Kriging* and *Simple Kriging* will be covered in the following subsections.

### Ordinary Kriging

In Ordinary Kriging (OK), the mean is assumed to be unknown. Considering a random function  $Z$  from which a set of samples were collected at  $x_i$  locations  $i = 1, \dots, N$ , the value at location  $x_0$  is estimated as follows:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i(x_0) z(x_i) \quad (2.15)$$

where  $n$  represents the data points in the local neighbourhood around  $x_0$ , which is a subset of  $N$ . The estimation unbiasedness is guaranteed by summing the weights  $\lambda_i$  to one:

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.16)$$

In order to obtain the weights, the following linear equations need to be solved:

$$\begin{aligned} \sum_{i=1}^n \lambda_i(x_0) \gamma(x_j - x_i) + \mu(x_0) &= \gamma(x_j - x_0) & \text{for } i = 1, 2, \dots, n \\ \sum_{i=1}^n \lambda_i(x_0) &= 1 \end{aligned} \quad (2.17)$$

where  $\mu$  represents the Lagrange multiplier used to achieve the variance minimisation and  $\gamma(x_j - x_i)$  is the average semi-variance between the data points. The equation can also be written in a matrix form as follows:

$$K\lambda = k \quad (2.18)$$

$$\begin{bmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1(x_0) \\ \vdots \\ \lambda_n(x_0) \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix} \quad (2.19)$$

where  $K$  represents the  $(n+1) \cdot (n+1)$  matrix of data semivariances,  $\lambda$  represents the kriging weights and the Lagrange multiplier.  $k$  on the right hand side of the equation represents the data-to-unknown covariances. The kriging variance can be obtained by:

$$\sigma^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \mu(x_0) \quad (2.20)$$

### Simple Kriging

Simple Kriging (SK) is similar to OK except that the mean  $m$  needs to be specified as a priori knowledge (Abzalov, 2016). Therefore, the SK estimator utilises  $(n+1)$  number of information including the defined mean  $m$ :

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i z(x_i) + \left\{1 - \sum_{i=1}^n \lambda_i\right\} m \quad (2.21)$$

Unlike the OK, the sum of the SK weights need not be equal to one. Due to these characteristics of the method, the SK can only be expressed in terms of covariances (Goovaerts, 2000). Hence, the equation that needs to be solved becomes

$$\sum_{i=1}^n \lambda_i C(x_j - x_j) = C(x_0, x_i) \quad \text{for } i = 1, 2, \dots, n \quad (2.22)$$

The minimum error variance for the SK is estimated by:

$$\sigma^2(x_0) = C(0) - \sum_{i=1}^n \lambda_i C(x_i - x_0) \quad (2.23)$$

### 2.2.5 Geostatistical Simulations

The most common problem with the kriging methods is that the large values tend to be underestimated whereas the smaller values are overestimated (Webster and Oliver, 2001). In other words, the kriging results represent the smoothed version of the reality due to the loss of original data variance. Geostatistical simulation techniques can address this issue by reproducing the variation inferred from the data without the smoothing effect. An additional benefit of the simulations includes the incorporation of the uncertainties in the model through the equiprobable realisations.

Geostatistical simulations generate realisations of a random field with certain statistical and geostatistical properties such as spatial variability and histogram. The primary aim is to construct a set of realisations representing possible scenarios for the variable distribution. Altogether, these realisations serve as a model of uncertainty of the attribute within the simulation domain (Norrena, 2007). If any data on the attribute of interest is available, the realisations can be conditioned to that data (conditional simulation). Similarly, if no samples have been collected, one can also perform unconditional simulations based on a priori knowledge on the expected statistical properties. In such situations, the focus is frequently on the reproduction of the specified mean and the variogram (Webster and Oliver, 2001). In conditional simulations, apart from the reproduction of the original variability, the realisations honour the observations at the sample locations (Gebbers and De Bruin, 2010). Considering a set of observations  $z(x_i), i = 1, 2, \dots, N$ , for instance, the simulated values  $z_c^*(x_j), j = 1, 2, \dots, T$  should be the same at the sample locations:

$$z_c^*(x_i) = z(x_i), \quad \text{for all } i = 1, 2, \dots, N \quad (2.24)$$

### Sequential Gaussian Simulations

In Sequential Gaussian Simulations (sGs), the values at each location of a domain is simulated sequentially based on a conditional cumulative distribution function (ccdf) computed at each location. The conditioning data used includes the observations as well as the previously simulated values. Given a set of  $k$  observations  $Z = \{Z(x_i)|i = 1, \dots, k\}$ , the steps of the algorithm can be described as follows (Gebbers and De Bruin, 2010; Olea, 2012):

1. Pre-processing:
  - (a) Check if the conditioning data  $Z(x_i)$  is normally distributed. Perform a normal transformation to obtain normal scores if necessary.
2. Initial definitions:
  - (a) Define the variogram: Compute the experimental variogram and fit a model. For the unconditional simulations, select a model and define the parameters.
  - (b) Initialise a simulation grid: Create a regular grid comprised of  $m$  nodes at which the variable of interest will be simulated.
  - (c) Move the data to the closest grid nodes.
  - (d) Define a path: Generate a random path to be sequentially visited to simulate the values at all nodes.
3. Simulation of the grid node  $n$  at  $x_0$  location:
  - (a) Kriging: Perform a kriging estimation of  $Z^*(x_0)$  and the variance  $\sigma_K^2(x_0)$  at the node  $n$  based on the observations  $Z(x_i)$  and the previously simulated data  $Z_{s(1,2,\dots,n-1)}$  within the search radius.
  - (b) Modelling of the ccdf: Establish a Gaussian cdf using the kriged estimate  $Z^*(x_0)$  and the variance  $\sigma_K^2(x_0)$ ,  $N(Z^*(x_0), \sigma_K^2(x_0))$ .
  - (c) Random number generation: Create a random number using the constructed ccdf and assign this number  $Z_s(x_0)$  to the grid node.

- (d) Visit the next grid number  $(n + 1)$  on the path defined.
- (e) Repeat steps 3.a to 3.d until all the nodes have been informed.
- (f) Back transformation if any transformation has been carried out.

### Turning Bands Simulations

The Turning Bands method (TB) is a computationally efficient simulation technique working in reduced dimensionality (Olea, 2012). The method creates a set of independent one-dimensional realisations with a certain covariance and averages them to generate the realisations of a multi-dimensional random function. The steps of the algorithm include the following:

1. Perform a normal score transformation of the data, if required.
2. Fit a covariance model  $Cov_n(h)$  to the transformed data using the standard variogram modelling techniques.
3. Derive the covariance  $Cov_1(h)$  that will be used to produce one-dimensional realisations.
4. Create a number of one-dimensional independent realisations centralised at the origin of a multidimensional sphere.
5. Sum up all the line realisations to produce the multi-dimensional realisation  $Z_n(x_i)$ .
6. Back transform if necessary.

## 2.3 Multiple-Point Statistics

Traditional geostatistics explained in the preceding sections rely on the variogram or covariance-based random function theory (Journel, 2005). Therefore, the geological continuity is defined in second-order statistics, which is based on 2-point correlations in space. Although the second-order statistics can be suitable to

characterise Gaussian processes, it is inadequate to identify complex structures exhibiting non-linear physical realities (Guardiano and Srivastava, 1993; Mariethoz and Caers, 2014). An example of these is the curvilinear structures which are strongly connected and frequently define sub-surface preferential flow paths (Feyen and Caers, 2005). In addition to the lack of structural information that covariances/variograms can identify, prior knowledge on them may be rather difficult to infer and model (Journel, 2005; De Iaco and Maggio, 2011; Mariethoz and Caers, 2014). In traditional geostatistics, the structural information is often derived from the observations collected. In situations where the number of available observations are scarce, the experimental variograms may not reveal a clear structure. Therefore, model parameters for the variograms cannot be accurately inferred. Under these circumstances, a common approach becomes the derivation of model parameters from a geologically similar area, from an outcrop or based on the prior geological knowledge (Journel, 2003). Another limitation of the traditional geostatistics is that the mathematical representation of geostatistics may not be non-expert friendly and require substantial knowledge in statistical modelling and the prevailing geology to be modelled (Caers and Zhang, 2004; Mariethoz and Caers, 2014).

The above-mentioned limitations of traditional geostatistics can be addressed by multiple-point geostatistics (MPS). MPS provides a framework to go beyond the second-order statistics through the use of Training Images (TI). A TI can be considered as an analogue of a variogram/covariance model. It essentially serves as a conceptual geological model and contains spatial patterns which are deemed representative of the spatial variations of the geology of interest (Mariethoz and Caers, 2014). One of the benefits of using a TI as structural information is that it accommodates higher-order rich structural information that cannot be inferred from a set of observations through variograms. Secondly, the use of TI avoids the problem of having unclear structural information due to undersampling. Unlike traditional geostatistics in which the structural information is derived from the observations, a TI offers complete and rich structural information. Therefore, utilisation of a TI in modelling yields much more realistic results than traditional



geostatistics. Structures such as curvilinear features are very well reproduced with remarkable connectivity. Lastly, the use of a TI as an analogue image is rather non-expert friendly as the spatial continuity is not expressed in mathematical models as in geostatistics (Mariethoz and Caers, 2014). Examples of the sources of a TI can be digitised hand-drawn sketches of the geology, outcrops or present-day depositions (Journel, 2003).

The spatial patterns existing in the TI can be extracted by a number of MPS simulation algorithms that are available (Guardiano and Srivastava, 1992; Zhang et al., 2006; Arpat and Caers, 2007; Gloaguen and Dimitrakopoulos, 2009; Dimitrakopoulos et al., 2010; Honarkhah and Caers, 2010; Mariethoz et al., 2010; Straubhaar et al., 2011; Tahmasebi et al., 2012). These algorithms reproduce the patterns contained in a TI in a puzzling manner and differ in the way they perform this operation. Therefore, each of the algorithms has its own advantages and disadvantages over other algorithms. Among all the algorithms currently available, Direct Sampling offers plenty of advantages including its capability to simulate high numbers of facies, non-stationary fields and continuous or multivariate simulations in a very fast manner without scanning the whole TI (Mariethoz et al., 2010). In this thesis, incorporation of the MPS framework to model the geological interface within lateritic metal mines is investigated with the Direct Sampling algorithm.

### 2.3.1 Higher-Order Spatial Statistics with Cumulants

In statistics, cumulants are used to describe the probability distribution of a random variable with the linear combination of statistical moments (Mariethoz and Caers, 2014). Application of cumulants in spatial statistics has been put in a framework by Dimitrakopoulos et al. (2010) as an alternative way to characterise non-linear and non-Gaussian ergodic spatial random fields. Spatial cumulants serve as the parametric representation of the interaction of more than two-points in a mathematical framework. Due to these characteristics, they can also be considered as an extension to the covariance functions (Mustapha and Dimitrakopoulos, 2010).

Given a random variable  $Z$ , the moment-generating function can be expressed using its Taylor expansion about the origin (Rosenblatt, 1985):

$$M(w) = E[e^{wZ}] = \sum_{r=0}^{\infty} \frac{w^r \overbrace{Mom[Z, \dots, Z]}^r}{r!} \quad (2.25)$$

the  $r$ th moment can be obtained by taking the derivative of  $M$  at the origin:

$$\frac{d^r}{dw^r} [E(e^{wZ})] = \overbrace{Mom[Z, \dots, Z]}^r \quad (2.26)$$

similarly, the cumulant-generating function  $K$  can be defined as the Neperian logarithm of the moment generating function  $M$  and can also be expressed as a Taylor expansion:

$$K(w) = \ln(E[e^{wZ}]) = \sum_{r=0}^{\infty} \frac{w^r \overbrace{Cum[Z, \dots, Z]}^r}{r!} \quad (2.27)$$

Therefore, the cumulants can be expressed as functions of moments. The relationship between the cumulants and moments can be expressed as follows:

$$\begin{aligned} Cum[Z] &= Mom[Z] \\ Cum[Z, Z] &= Mom[Z, Z] - Mom[Z]^2 \\ Cum[Z, Z, Z] &= Mom[Z, Z, Z] - 3Mom[Z, Z]Mom[Z] + 2Mom[Z]^3 \end{aligned} \quad (2.28)$$

Let us now consider a zero-mean, stationary and ergodic random field  $\{Z(x), i = 0, 1, \dots, N\}$ ,  $x_i \in \Omega \subseteq \mathbb{R}^r$  ( $r = 1, 2, \text{ or } 3$ ) where  $N$  represents the number of points in a discrete grid  $D_N$ . The spatial cumulant of  $Z(x)$  of order  $r$  becomes:

$$c_r^Z(h_1, \dots, h_{r-1}) = Cum[Z(x), Z(x+h_1), \dots, Z(x+h_r)] \quad (2.29)$$

Based on the relationship between the moments and the cumulants given in Eq.2.28,

the first order cumulant for a zero-mean random field  $Z(x)$  can be expressed as:

$$c_1^Z = E[Z(x)] = 0 \quad (2.30)$$

The second-order cumulant of a non-centred random function  $Z(x)$  is the covariance:

$$c_2^Z(h) = E[Z(x)Z(x+h)] - E[Z(x)]^2 \quad (2.31)$$

The third cumulant becomes as follows:

$$\begin{aligned} c_3^Z(h_1, h_2) = & E[Z(x)Z(x+h_1)Z(x+h_2)] \\ & - E[Z(x)]E[Z(x+h_1)Z(x+h_2)] \\ & - E[Z(x)]E[Z(x+h_1)Z(x+h_3)] \\ & - E[Z(x)]E[Z(x+h_2)Z(x+h_3)] + 2E[Z(x)]^3 \end{aligned} \quad (2.32)$$

where  $h_3$  represents the interaction between the vectors  $h_1$  and  $h_2$ . Given a set of conditioning data, the cumulants can be experimentally calculated by scanning the given data with a spatial template that has a specific geometry (Minniakhmetov and Dimitrakopoulos, 2017). Considering an  $x$  location, for instance, a spatial template centred around  $x$  with a set of directional vectors  $\{h_1, \dots, h_n\}$  can be defined as  $d_n(x, L) = \{Z(x+h_1), \dots, Z(x+h_n)\}$ . For the third-order cumulant, the template becomes  $d_2(x, L) = \{Z(x+h_1), Z(x+h_2)\}$  and the third-order cumulant is calculated as:

$$c_3^Z(h_1, h_2) = \frac{1}{N_{h_1, h_2}} \sum_{k=1}^{N_{h_1, h_2}} Z(x)Z(x+h_1)Z(x+h_2) \quad (2.33)$$

where  $N_{h_1, h_2}$  is the number of replicates of the template. In the case of data on an irregular grid, the tolerances in distances, angles and bands are included.

## 2.4 Summary and Discussion

This chapter provided a review of the methods that are utilised to develop the approaches presented in Chapter 4, 5 and 6. The review started with geostatistics to model an attribute using the theory of regionalised variables. This was then followed by the theories of geostatistical estimation and simulation techniques such as OK, SK, sGs and TB. Use of variogram models in geostatistics has some limitations, particularly for scarce data sets. MPS can address such limitations of geostatistics by incorporating analogue data sets in modelling practices through the TI concept. Therefore, MPS exhibits a potential to improve the contact modelling in lateritic metal deposits.

Cumulants allow quantifying the multiple-point statistics of a partial or complete data set using a spatial template that has a particular spatial configuration. The way it performs such a task is different than that of the smooth-histogram technique presented in Chapter 5. For instance, the smooth histogram technique utilises pattern similarities for a given spatial template. Whereas, the cumulant method computes higher-order statistics using a spatial template that has a particular spatial configuration. The computation of cumulants is performed for different lag distances. On the other hand, the smooth histogram method uses only one spatial template with a certain size. Hence, it takes longer to generate a cumulant map as compared to computing pattern statistics through the smooth histogram. Detailed information on the smooth histogram method is given in Chapter 5.

## References

- Abzalov, M. (2016). *Applied mining geology*, volume 12. Springer.
- Armstrong, M. (1998). *Basic linear geostatistics*. Springer Science & Business Media.
- Arpat, G. B. and Caers, J. (2007). Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203.
- Caers, J. and Zhang, T. (2004). Multiple-point geostatistics: a quantitative vehicle for integrating geologic analogs into multiple reservoir models.
- De Iaco, S. and Maggio, S. (2011). Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. *Mathematical Geosciences*, 43(4):483–500.
- Dimitrakopoulos, R., Mustapha, H., and Gloaguen, E. (2010). High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-gaussian and non-linear phenomena. *Mathematical Geosciences*, 42(1):65–99.
- Feyen, L. and Caers, J. (2005). Multiple-point geostatistics: a powerful tool to improve groundwater flow and transport predictions in multi-modal formations. *Geostatistics for environmental applications*, pages 197–208.
- Gebbers, R. and De Bruin, S. (2010). Application of geostatistical simulation in precision agriculture. In *Geostatistical applications for precision agriculture*, pages 269–303. Springer.
- Gloaguen, E. and Dimitrakopoulos, R. (2009). Two-dimensional conditional simulations based on the wavelet decomposition of training images. *Mathematical Geosciences*, 41(6):679–701.
- Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, 228(1):113–129.

- Guardiano, F. and Srivastava, R. (1992). Borrowing complex geometries from training images: The extended normal equations algorithm. *Stanford Center for Reservoir Forecasting Report, Stanford University*.
- Guardiano, F. B. and Srivastava, R. M. (1993). Multivariate geostatistics: beyond bivariate moments. In *Geostatistics Troia92*, pages 133–144. Springer.
- Honarkhah, M. and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42(5):487–517.
- Journel, A. G. (2003). Multiple-point geostatistics: A state of the art.
- Journel, A. G. (2005). Beyond covariance: the advent of multiple-point geostatistics. *Geostatistics Banff 2004*, pages 225–233.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining geostatistics*. Academic press.
- Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.
- Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11).
- Matheron, G. et al. (1965). Les variables régionalisées et leur estimation.
- Minniakhmetov, I. and Dimitrakopoulos, R. (2017). Joint high-order simulation of spatially correlated variables using high-order spatial statistics. *Mathematical Geosciences*, 49(1):39–66.
- Mustapha, H. and Dimitrakopoulos, R. (2010). A new approach for geological pattern recognition using high-order spatial cumulants. *Computers & Geosciences*, 36(3):313–334.
- Norrena, K. P. (2007). *Decision Making Using Geostatistical Models of Uncertainty*. ProQuest.

- Olea, R. A. (2012). *Geostatistics for engineers and earth scientists*. Springer Science & Business Media.
- Oliver, M. (2010). An overview of geostatistics and precision agriculture. In *Geostatistical applications for precision agriculture*, pages 1–34. Springer.
- Oliver, M. A. and Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging*. Springer.
- Rosenblatt, M. (1985). Stationary sequences and random fields.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.
- Tahmasebi, P., Hezarkhani, A., and Sahimi, M. (2012). Multiple-point geostatistical modeling based on the cross-correlation functions. *Computational Geosciences*, 16(3):779–797.
- Wackernagel, H. (2013). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.
- Webster, R. and Oliver, M. A. (2001). *Geostatistics for environmental scientists (statistics in practice)*.
- Zhang, T., Switzer, P., and Journel, A. (2006). Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38(1):63–80.

## Chapter 3

# Weipa Bauxite Mine and Data Description



## 3.1 Introduction

This chapter provides background information about the data used to develop the methods for the resource estimation of lateritic bauxite deposits. The data used comes from two mine areas, namely Oak and Kumbur mine areas, located within the Weipa bauxite mine. Therefore, the chapter starts with brief information on the Weipa bauxite mine. This is then followed by the geology of the Weipa bauxite mine and the mining methods used to extract the bauxite ore. Lastly, the data collected from the Oak and Kumbur mine areas is introduced. The exploratory data analysis has also been performed and presented in this chapter as well.

## 3.2 Weipa Bauxite Mine

Weipa bauxite mine is one of the biggest open-cut bauxite mine operations in the world. It is located in the west coast of Cape York peninsula in northern Queensland, Australia (Fig.3.1). Having been discovered in 1955 by an oil prospecting company, further investigations revealed the suitability of the deposit for alumina production (Evans, 1959; Duncan, 1961). The operations started in 1961 and the mine is currently being operated by Rio Tinto Alcan. There are mainly two major sites located within the mine. These sites are named as Andoom and Weipa, which are divided by the Mission River. Bauxite formation is distributed approximately 10,000 km<sup>2</sup> over the Western Cape York with a depth ranging between 1m and 6m (Evans, 1975).

### 3.2.1 Geology of the Mine Site

The western part of Cape York Peninsula is mainly comprised of two types of sediments, which are Rolling Down Group sediments and Bulimba Formation sediments (Schaap, 1990). Rolling down group sediments form the parent rock formation of Andoom mine site and are estimated to be 50 to 100 million years old. Bulimba formation, on the other hand, dominates the Weipa site and dates back

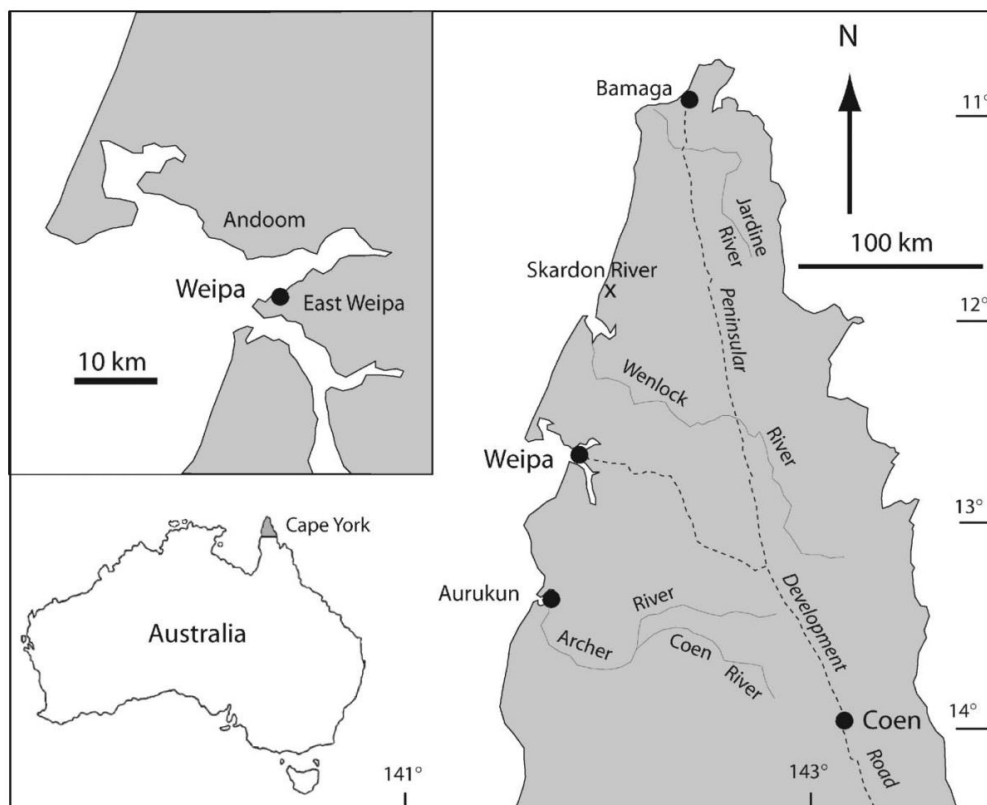


Figure 3.1: Location of the Weipa bauxite mine. After Taylor et al. (2008)

to Tertiary time period, which is estimated to be 50 million years old. The sediments of Bulimba overly the Rolling down group and contain tertiary or upper Cretaceous unconsolidated fluvial sands and clays. Rolling down group basically comprise marine claystone, mudstone, siltstone, lithic sandstone and minor quartz sandstone.

The bauxite deposit at the Weipa is thought to have formed from the in-situ chemical weathering of kaolinite, quartz and iron oxide minerals (Loughnan and Bayliss, 1961). Primary controlling factors for the weathering comprised climate, vegetation cover, ground water circulation, bedrock composition and texture (Gow and Lozej, 1993). Since the geology is originated from such a process, the bauxite deposit exists within a regolith profile comprised of several geological units (Francke, 2012). The overall weathering profile has a thickness ranging between 20 to 35 m (Schaap, 1990). The components that make up of the weathering profile are, from bottom to top, top-soil, pisolitic bauxite, ferricrete (ironstone), kaolin and sandstone units, respectively (Fig. 3.2). The thickness of the major units are: 0.5 m for the top-soil, 0.25-8 m for the bauxite unit and 1-2m for the ferricrete unit (Schaap, 1990). The alumina-rich bauxite is composed of loose pisolithes with a size range of 0.5 to 30 mm (Bárdossy and Aleva, 1990).

The geological interface between the bauxite/ferricrete is believed to correspond to the water table level at the time of bauxite formation (Eggleton et al., 2005). In order to visually observe the degree of lateral variability of the contact, a test excavation with a small-scale excavator had previously been performed (Erten, 2012). This specific excavation has demonstrated that the geological interface separating the bauxite from the ferricrete is rather wavy, as can be seen in Fig. 3.3. It was also observed that the revealed surface undulates with almost random frequencies. Furthermore, the frequency of the short-scale fluctuations observed in the exposed surface were higher than the exploration borehole spacing, making the chosen drill spacing (76.2 x 76.2 m) insufficient to reliably model the ore boundaries.

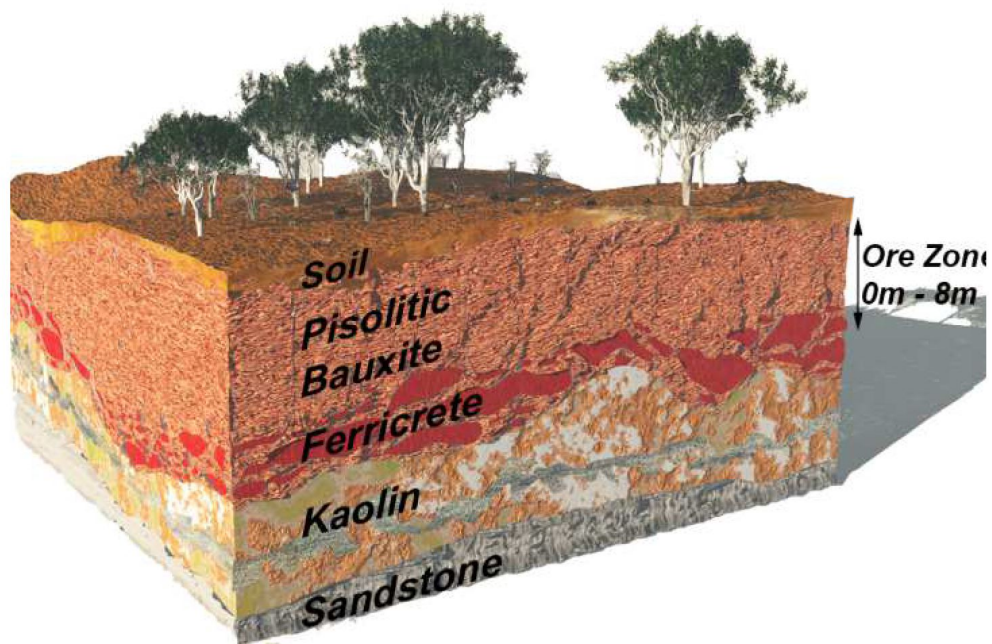


Figure 3.2: Geological units present at the Weipa bauxite mine. After Francke (2012)



Figure 3.3: The test excavation conducted to reveal the characteristics of the lateral variability inherent in the geological interface (Erten, 2012)

### 3.2.2 Bauxite Mining Practices at Weipa

#### Exploration

Mining of bauxite at Weipa mine pre-requires an exploration programme to be conducted to model the deposit. This is mainly done to quantify the bauxite tonnages and average grades of a planned mining area by creating a geological resource model. The drilling programme conducted at Weipa comprises regularly spaced boreholes with a spacing of  $76.2 \times 76.2$  m (Erten, 2012). Air-core drilling method is utilised to drill the boreholes with an average depth of 3.5m. The core samples are collected in 0.25 m vertical intervals and assayed for  $Al_2O_3$ ,  $SiO_2$ ,  $Fe_2O_3$ ,  $TiO_2$ ,  $LoI$ ,  $TAA$  (total available alumina),  $MHA$  (monohydrate alumina or boehmitic or diasporitic alumina) and  $THA$  (Trihydrate alumina or gibbsitic alumina). Modelling of the grades is performed by first constructing a block model inside a pre-defined ore boundaries. The grade values are then interpolated at each discrete location of the blocks. The ultimate model is then used to generate short, medium and long term mine plans.

#### Mining

The bauxite ore at the Weipa exists within the soil horizons and it has free-flow characteristics. Therefore, no drilling and blasting is required to fragment the ore material and the overlying geological units (Bárdossy and Aleva, 1990). Extraction of the bauxite requires, first, a site preparation work involving the removal of the top-soil material. This task is carried out through Cat-657 type scrapers with a  $33 \text{ m}^3$  capacity (Erten, 2012). The stripped top-soil material is then dumped at a spot which is next to the mining area for the future rehabilitation purposes.

Having completed the displacement of the top-soil, the exposed bauxite unit is extracted using rubber-tired front-end loaders, with a bucket capacity of  $10 \text{ m}^3$  as shown in Fig. 3.4. The extracted material is hauled by Cat bottom-dump trucks with a 150-tonne capacity as shown in Fig. 3.5. Depending on the location of a specific mine area within the Weipa, the ore is either directly dumped to a stockpile or transferred to the wagons of a train by the trucks and then dumped to a stockpile.



Figure 3.4: Front-end loader type of mining equipment used to extract the bauxite ore at the Weipa bauxite mine (Erten, 2012)

### 3.3 Oak and Kumbur Mine Areas

Oak and Kumbur mine areas are located within the Weipa bauxite mine and are used in this thesis to develop and investigate techniques to improve current resource estimation practices. Oak mine is located in Andoom mine site and covers an area of  $360 \times 800$  m, whereas the Kumbur mine area exists within the north of Andoom mine site and is  $215 \times 500$  m in size (Erten, 2012). Major components of the laterite profile at Oak and Kumbur mine areas are: overburden layer, red soil layer, Al-rich pisolitic bauxite layer, transitional zone and ferricrete layer.

#### 3.3.1 Data

There are three types of data collected from both Oak and Kumbur mine areas to create a resource model: (1) exploration boreholes (2) saturation boreholes and (3) dense GPR survey data to help locating the ore boundaries. The exploration boreholes are drilled on a regular grid of  $76.2 \times 76.2$  m in both Oak and Kumbur



Figure 3.5: Cat bottom-dump type truck used to haul the extracted bauxite material. The capacity is 150 tonne (Erten, 2012)

mine areas. These boreholes contain a variety of information including grades and lithologies defined in three-dimensional space. The number of boreholes drilled are 33 for the Oak and 13 for the Kumbur mines. In addition to the regular exploration boreholes, 218 saturation boreholes, also called as production control boreholes (PCD), were drilled in the Kumbur area. The spacing between these holes were 19.05 m and they are only analysed for the lithologies to aid delineating the ore boundaries. Therefore, the PCD only contain the elevation values corresponding to the bauxite/ferricrete interface. Location maps of the exploration and production boreholes can be seen in Fig. 3.6.

As for the GPR data, both mine areas were surveyed using an UltraGPR radar system with a 80-MHz antenna mounted. The data collection was performed by towing a GPR device along a predefined virtual grid for each mine area. The bauxite/ferricrete boundary was sampled every 0.25 m along the profiles. The spacing between the profiles were approximately 15 m for the Oak mine area and 9 m for the Kumbur mine area. The total area covered for the Oak mine was 14.23 hectares with 17.94 km GPR profile. As for the Kumbur mine, the GPR surveys covered 6.78 hectares with 15.17km of GPR profile. The number of GPR

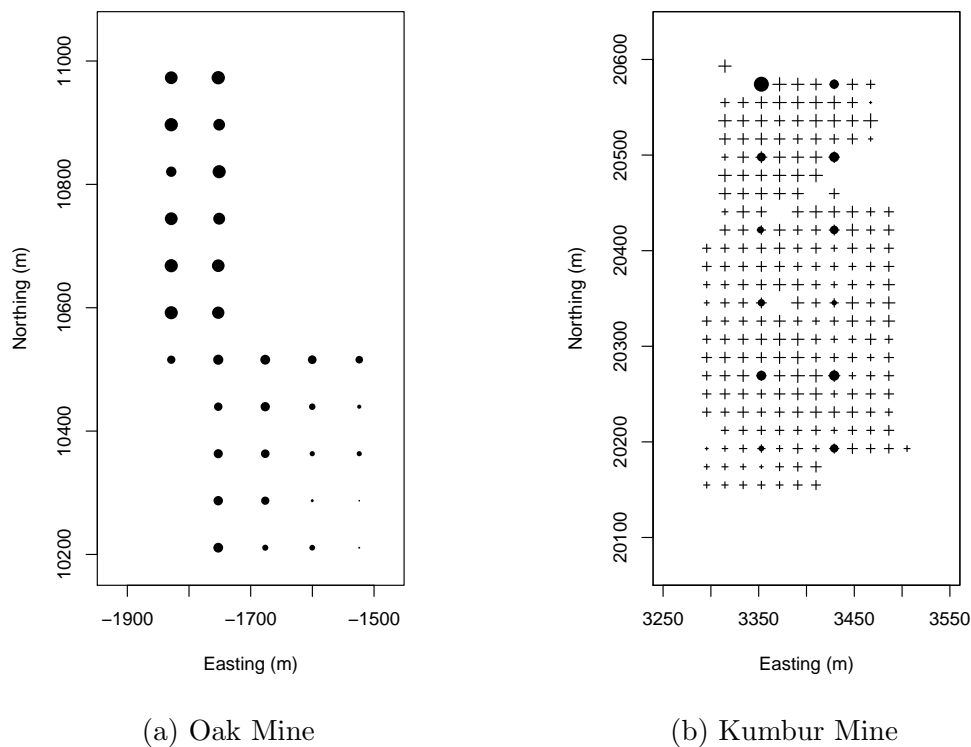


Figure 3.6: Location maps for borehole data of the Oak and Kumbur mine areas. Circles represent the exploration borehole locations whereas "+" signs represent the saturation drill hole locations

pick-points collected was 83,861 for the Oak and 60,451 for the Kumbur. The locations of the GPR pick-points are given in Fig. 3.7. Following the extraction of the bauxite ore at each mine area, the exposed topographical surfaces were surveyed. These resulting mined-out surfaces can be deemed representative of the combination of lateral variability of the geological interface as well as the mining equipment selectivity. The survey points collected after the bauxite extraction can be seen in Fig. 3.8.



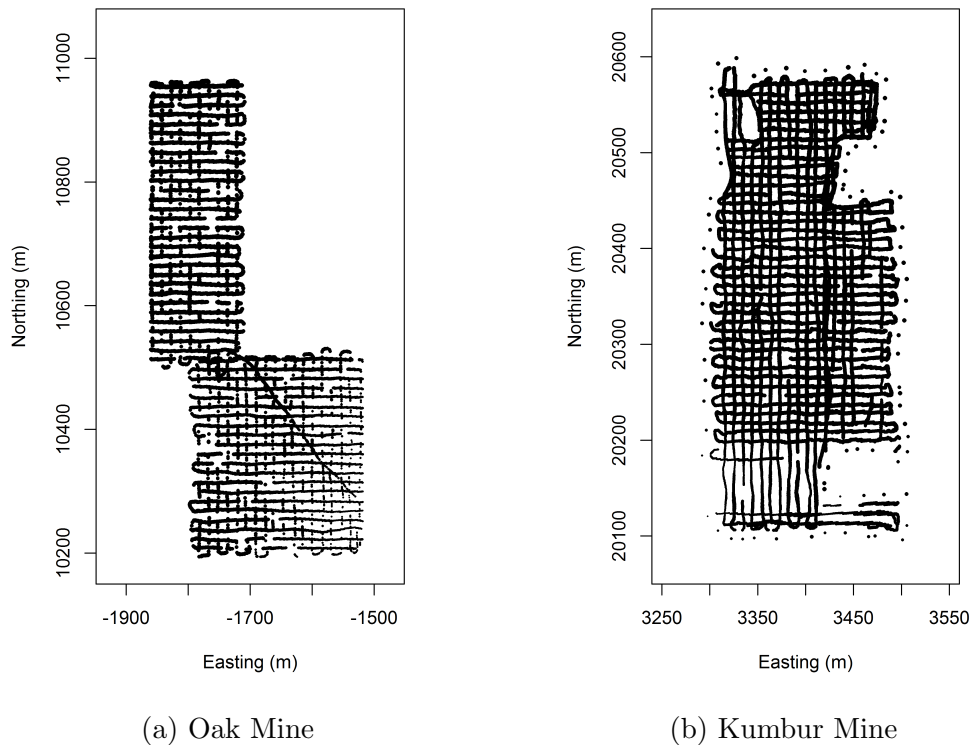


Figure 3.7: Location maps of the GPR pick-points collected from the Oak and Kumbur mine areas

### 3.4 Exploratory Data Analysis

Exploratory data analysis (EDA) serves as one of the essential tasks to evaluate a resource estimation project (Tukey, 1977). It particularly helps to understand the statistical properties of the datasets through a number of statistical tests performed prior to the estimation practices (Abzalov, 2016). In this thesis, the statistical properties of the available data is explored using a number of commonly used EDA tools. Before conducting any statistical analysis, the GPR data of both mines are first pre-processed to remove any duplicate pick-points. This operation is mainly performed to avoid any possible matrix instability that may take place during geostatistical estimation or simulations.

As one of the tools used in a EDA, histograms are computed for each of the

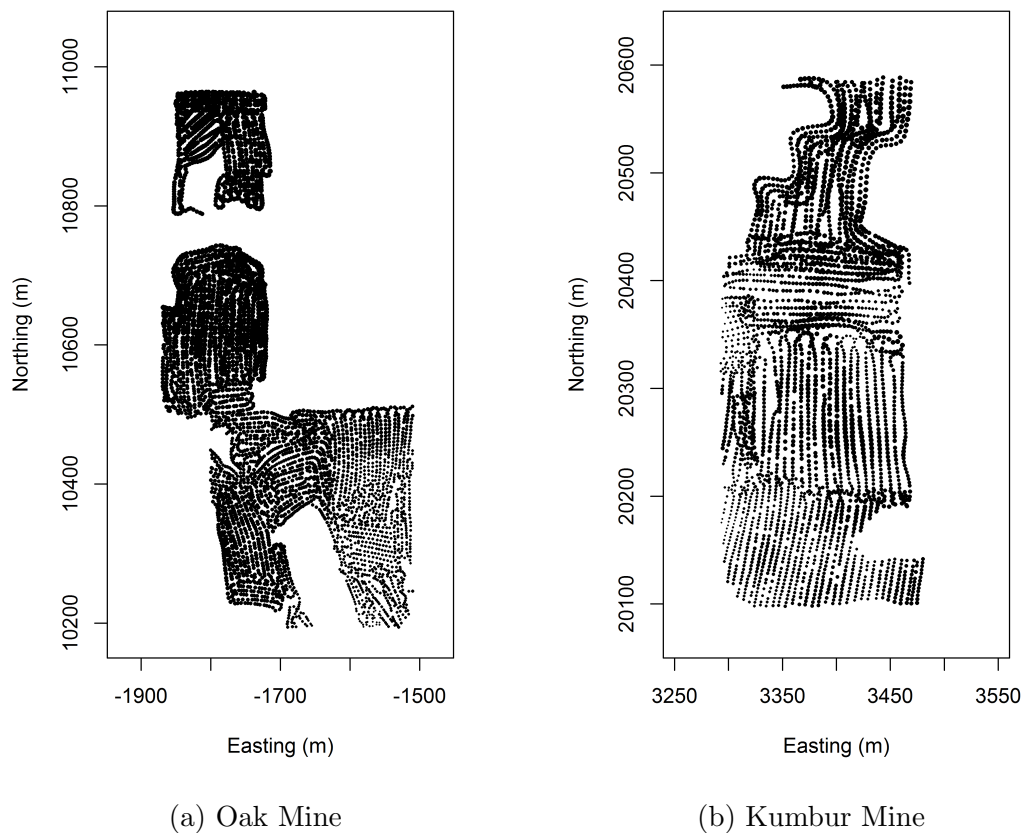


Figure 3.8: Location maps of the survey pick-points collected after the extraction of bauxite from Oak and Kumbur mine areas (mined-out surfaces)

variables to visually check the statistical distributions. The frequency distributions of the Oak and Kumbur mine variables are presented in Fig. 3.9 and Fig. 3.10 respectively.

It can clearly be observed from the Oak GPR elevation data that the bauxite/ferricrete interface can be multi-modal. Such a multi-modality is weakly apparent in the borehole and floor data, since they relatively comprise less number of samples. As for the Kumbur data, there is not an apparent evidence of multi-modality. However, the Kumbur GPR data seems to have a left skewed distribution.

An interesting point that can be observed from these histograms is that the

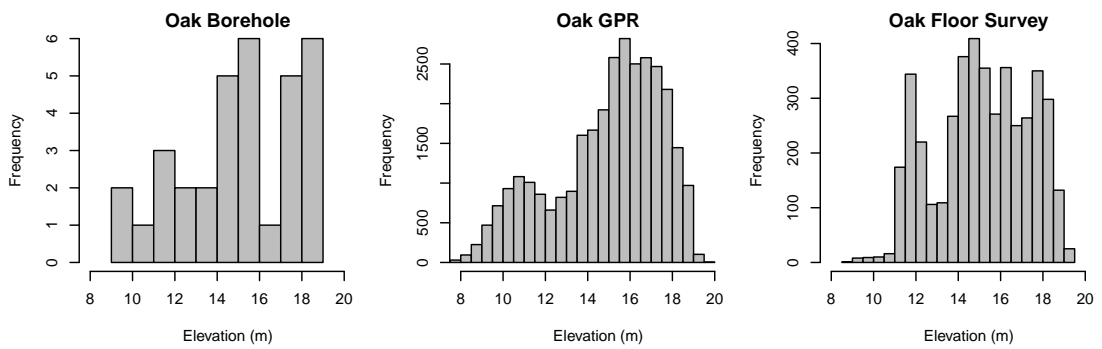


Figure 3.9: Histogram of the elevation data collected from Oak mine area.

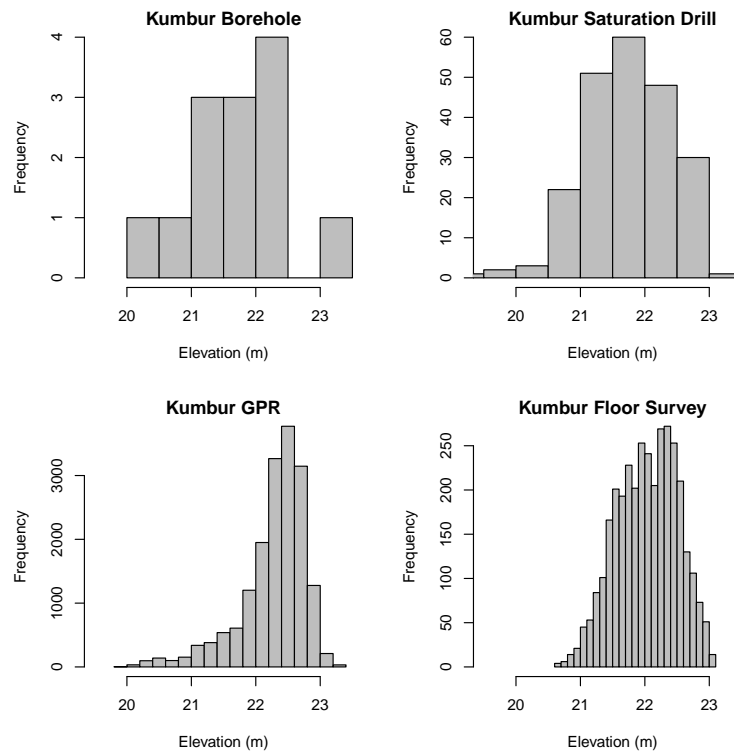


Figure 3.10: Histogram of the elevation data collected from Kumbur mine area.

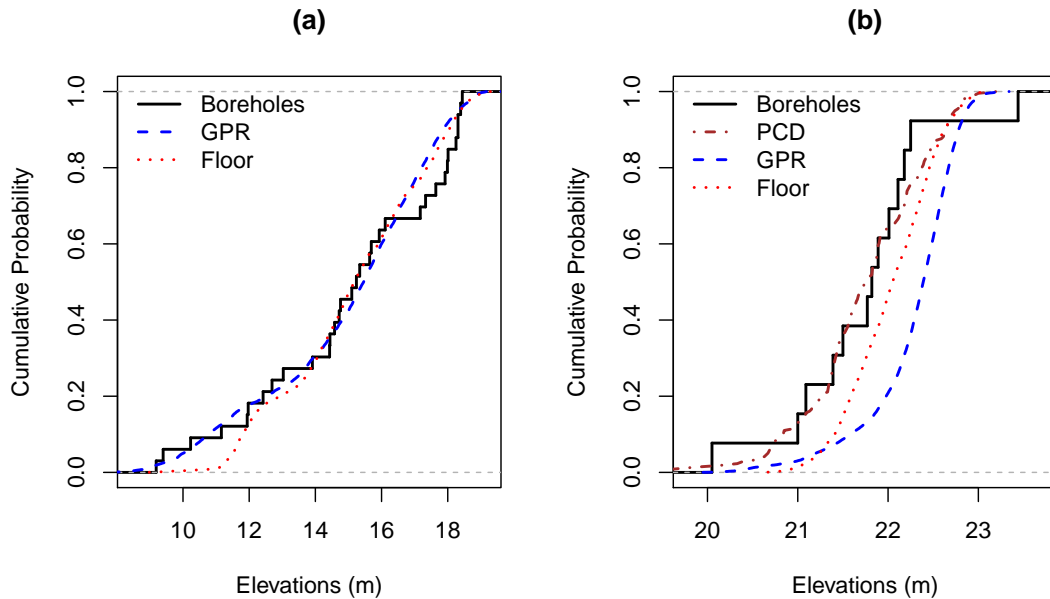


Figure 3.11: Empirical cumulative distribution function for the elevation variables collected from Oak (a) and Kumbur (b) mine areas

floor data contains less number of elevation values within the low histogram bins, as compared to the borehole and GPR variables. Histogram-based statistics presented in Table 3.1 also confirm this situation. In both Oak and Kumbur data, minimum elevation values of the floor survey data are higher than those of the GPR data. This could be related to the decision made by the front-end loader operator at the time of mining. Since the dilution is rather costly, the operator might intuitively have been more cautious at lower bauxite/ferricrete elevations.

Table 3.1: Descriptive Statistics of the Variables for Oak and Kumbur Mine Areas

Variable (m)	Count	Min	Max	Mean	Median	Variance	Std.	IQR	C. of Var.	C. of Skew.
Oak Borehole	33	9.19	18.44	15.02	15.24	7.55	2.75	4.61	0.18	-0.49
Oak GPR	30630	7.81	19.90	14.92	15.46	6.60	2.57	3.46	0.17	-0.64
Oak Floor	4350	8.98	19.34	15.13	15.18	4.85	2.20	3.26	0.15	-0.22
Kumbur Borehole	13	20.05	23.44	21.73	21.82	0.63	0.80	0.72	0.04	-0.02
Kumbur PCD	218	19.45	23.20	21.76	21.77	0.45	0.67	0.88	0.03	-0.42
Kumbur GPR	17245	19.95	23.34	22.28	22.39	0.26	0.51	0.53	0.02	-1.45
Kumbur Floor	3395	20.67	23.09	22.04	22.05	0.22	0.47	0.71	0.02	-0.20

The cumulative distribution function (CDF), also called the probability plot, is a convenient tool to illustrate the probability distribution of variable. Unlike the histogram, a CDF is not depended on the choice of the bin size as in histogram calculations (Rossi and Deutsch, 2013). The probability plots created for both mine areas are presented in Fig.3.11. According to these plots, the Oak mine elevation variables exhibit rather similar probability distributions. However, as also discussed previously, the floor elevations slightly deviate from the borehole and GPR elevations within the low elevation range. For the Kumbur mine, borehole and PCD distributions are rather similar. On the contrary, GPR and floor elevations noticeably deviate from the borehole and GPR data distributions. This could be due to the fact that the bauxite deposit at Kumbur mine is relatively shallower compared to the oak mine. Since the performance of the GPR surveys are affected by the depth of the deposit, the GPR surveys at Kumbur might have been affected by the existence of a shallow deposit thickness. The same reason could also be valid for the mined-out floor survey; existence of a shallow deposit could have made it difficult to accurately tracking the deposit boundaries.

In order to check the normality of the data, Q-Q plots of the raw data against the calculated normal scores are constructed as in Fig. 3.12 and Fig. 3.13. The Q-Q plot of the Oak data indicates the existence of a deviation from the normal distribution. The same is true for the Kumbur mine data as well, the elevation values exhibit slight to moderate deviations from the theoretical distribution quantiles. However, when the coefficient of variation values in Table 3.1 are concerned, all the values are below 1. Hence, the distributions do not pose a significant erratic values that might cause issues in geostatistical estimations (Issaks and Srivsatava, 1989).

### 3.5 Summary and Discussion

In this chapter, a background information on the Weipa bauxite mine as well as the prevailing geology existing in the mine are given. Moreover, two mine areas within the Weipa mine, namely Oak and Kumbur mine areas, are presented. The data of

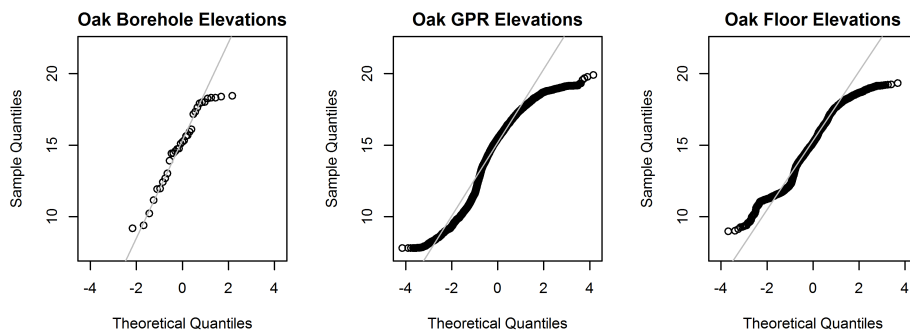


Figure 3.12: Q-Q plots of the Oak mine area variables.

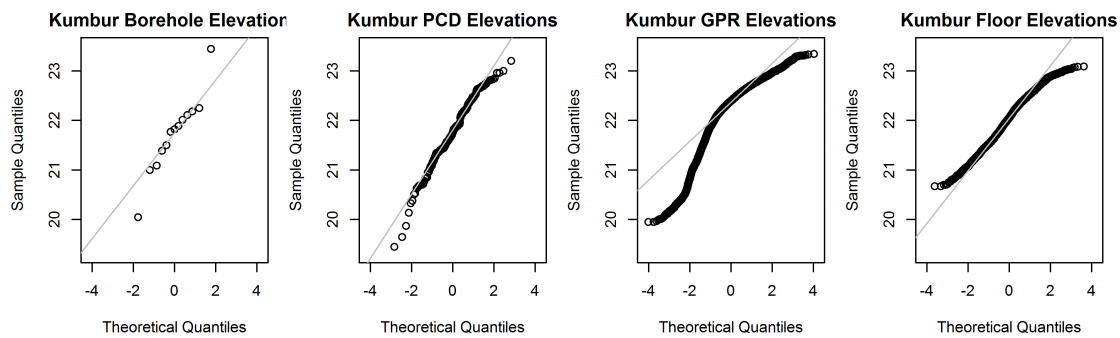


Figure 3.13: Q-Q plots of the Kumbur mine area variables.

these two mines are used in this thesis to develop and investigate the techniques to improve the current resource estimation practices for lateritic metal mines. Therefore, an exploratory data analysis of the variables from these mines is also performed to analyse the statistical characteristics of the data. The exploratory data analysis included plotting the location maps, calculating the histograms, cumulative distribution functions, Q-Q plots and the descriptive statistics.

## References

- Abzalov, M. (2016). *Applied mining geology*, volume 12. Springer.
- Bárdossy, G. and Aleva, G. J. J. (1990). *Lateritic bauxites*, volume 27. Elsevier Science Ltd.
- Duncan, C. (1961). The aluminum industry in australia. *Geographical Review*, 51(1):21–46.
- Eggleton, T., Taylor, G., et al. (2005). Bioturbation of the weipa bauxite. In *Regolith*, pages 83–85.
- Erten, O. (2012). Profiling and mining control to mitigate dilution effect from sio<sub>2</sub> at the base of a bauxite deposit.
- Evans, H. (1959). *The geology and exploration of the Cape York Peninsula bauxite deposits in northern Queensland*.
- Evans, H. (1975). Weipa bauxite deposit, queensland. *Knight CL, Economic geology of Australia and Papua New Guinea, Australasian Institute of Mining and Metallurgy: Melbourne*, pages 968–980.
- Francke, J. (2012). The role of ground penetrating radar in bauxite resource evaluations. In *Ground Penetrating Radar (GPR), 2012 14th International Conference on*, pages 459–463. IEEE.
- Gow, N. N. and Lozej, G. P. (1993). Bauxite. *Geoscience Canada*, 20(1).
- Issaks, E. H. and Srivastava, R. M. (1989). *Applied geostatistics*. Oxford University Press.
- Loughnan, F. and Bayliss, P. (1961). The mineralogy of the bauxite deposits near weipa, queensland. *American Mineralogist*, 46(1-2):209–217.
- Rossi, M. E. and Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.

- Schaap, A. t. (1990). Weipa kaolin and bauxite deposits. *Geology of the Mineral Deposits of Australia and Papua New Guinea, Australasian Institute of Mining and Metallurgy: Melbourne*, pages 1669–1673.
- Taylor, G., Eggleton, R., Foster, L., and Morgan, C. (2008). Landscapes and regolith of weipa, northern australia. *Australian Journal of Earth Sciences*, 55(S1):S3–S16.
- Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, Mass.



## Chapter 4

# Multiple-point statistical simulation of the ore boundaries for lateritic metal deposits

## 4.1 Introduction

This chapter presents the framework used to implement MPS on modelling the ore-body boundaries in lateritic metal deposits. The proposed approach was demonstrated on a lateritic bauxite deposit using the borehole and GPR data sets as simulation variables. The TIs used to perform the simulations were obtained from the footwall topographies exposed after the extraction of the bauxite ore at two historic mine areas, which were the Oak and Kumbur areas. In addition to the contact topographies exposed, the mine areas also had the GPR data collected prior to the extraction. Therefore, the simulations were performed in the form of both univariate and bivariate simulations using the DS MPS algorithm. The data from one mine area was used as a TI to perform the simulations for the other mine area and vice-versa. In order to compare the performance of the MPS simulations, the same simulations were also conducted using the TB geostatistical simulation technique. The comparison has been made using several statistical and spatial statistical indicators.

## 4.2 Background

The objective of resource estimation practice for mineral deposits is to accurately forecast the grades and tonnages that will be extracted within a specified time period (Bardossy et al., 2003; Rossi and Deutsch, 2013). This kind of a goal involves defining the geological boundaries inside which blocks models are constructed. A conventional practice to accomplish this is to outline the boundaries of the geological lithologies by explicitly drawing the geological contacts using the available borehole data (Osterholt and Dimitrakopoulos, 2018). This technique is considerably subjective and leads to an over-smoothed interpretation of the geology. Alternatively, geostatistical estimation methods may be utilised to define the geological domains (de Freitas Silva and Dimitrakopoulos, 2016). However, estimations performed using geostatistical techniques also exhibit a smoothing effect and do not take into account the uncertainties in the contact models.

Geostatistical conditional simulation approaches address the smoothing problems of the estimation techniques by creating equiprobable realisations of the orebody boundaries (Dimitrakopoulos, 1998). Each individual orebody realisation exhibits statistical properties and a variogram structure similar to what is observed in the borehole data. Therefore, the resulting realisations do not go through any smoothing effect, as in kriging. Furthermore, when taken jointly, the realisations created can be used as a model of uncertainty for the orebody boundaries. However, the geostatistical simulation approaches utilise variogram/covariance models for the structural information, which cannot capture and reproduce complex curvilinear geological features (Journel, 2005; De Iaco and Maggio, 2011; Mariethoz and Caers, 2014). Another drawback of the use of variograms is that the spatial continuity is inferred and modelled from a set of borehole data. Hence, in the case of a scarce data set, the structural information cannot be adequately identified.

As far as the MPS modelling framework is concerned, the structural information is not derived from a variogram model. It is instead obtained from a TI which could be regarded as an analogue of a variogram/covariance model used in classical geostatistics. A TI is a conceptual geological model containing spatial patterns which are believed to exist in the modelling domain (Mariethoz and Caers, 2014). These patterns are extracted from the TI and simulated in the modelling domain through a number of MPS algorithms available (Guardiano and Srivastava, 1992; Strebelle, 2002; Zhang et al., 2006; Arpat and Caers, 2007; Gloaguen and Dimitrakopoulos, 2009; Dimitrakopoulos et al., 2010; Honarkhah and Caers, 2010; Mariethoz et al., 2010; Straubhaar et al., 2011; Tahmasebi et al., 2012).

Incorporation of TIs in orebody modelling can offer some advantages. First of all, a TI comprises high-order and rich structural information that cannot be sufficiently described in the form of second-order statistics. Hence, obtaining the structural information from a TI makes it possible to reproduce complex orebody structures which better represent the reality. A second advantage of the use of TI is that it alleviates the difficulties experienced in the case of limited conditioning data sets. Rather than depending on the borehole data for the structural information, the spatial information is acquired from a TI which is composed of rich spatial

patterns. The last benefit of the TI is that it provides a user-friendly modelling framework as the structural information is not defined in mathematical expressions (Mariethoz and Caers, 2014).

There are very few applications of MPS for resource estimation of mineral deposits compared to the oil-gas industry (Pasti et al., 2012). This is mainly due to the difficulties experienced when obtaining a suitable TI which represents the geology of interest well (Mery et al., 2017). Nonetheless, a number of approaches have previously been employed to create TIs for mineral deposits. For instance, the exploration boreholes or blastholes collected from a mine site can be utilised to conceptualise the geology (Jones et al., 2013). Another example of deriving a TI in mineral deposits is to use deterministic orebody models generated using the borehole data (Goodfellow et al., 2012; van der Grijp and Minnitt, 2015). Knowledge obtained from the previously mined-out areas can also be utilised as representative geologies (Rezaee et al., 2013; Osterholt and Dimitrakopoulos, 2018). Finally, the TI can also be constructed using TI generator software if the properties of the expected geological features are known (Pasti et al., 2012).

In lateritic metal mines, the deposits are frequently shallow and the geological contact defining the ore boundaries can be observed after extracting the ore unit. Although the shape of a mined-out surface is affected by how well the interface is traced during the extraction, the resulting exposed topography still reflects the variations inherent in the contact surface. Therefore, if these surfaces are surveyed, they can be used as a TI to model the ore boundaries of another mining area with a similar geology. Due to such geological characteristics and the nature of mining operations involved, it is believed that the use of MPS in modelling the geological contacts in lateritic or stratified deposits have the potential to improve resource modelling practices; hence, it is definitely worth investigating. The following sections will describe the steps and the preliminary analysis conducted to implement the MPS simulations in a lateritic bauxite deposit.

## 4.3 Structural Analysis and Pre-Processing of the Data

As the first step, the available data was analysed in terms of its spatial statistics. This analysis revealed the existence of a trend in the data set and allowed to determine to perform the simulations using the residuals. The following two subsection illustrate: (1) how the trends were detected and mathematically expressed; and (2) the approach used to convert the point-type the data sets into the gridded ones.

### 4.3.1 Structural Analysis and Trend Detection

In order to perform the structural analysis of the elevation variables, the experimental variograms need to be computed and analysed first. These variograms can normally be calculated along a given direction. Alternatively, they can be also calculated independent of a direction, which is called an *omnidirectional* variogram. Omnidirectional variograms are useful for preliminary analysis of the structural variability of a RV. However, in the presence of an anisotropy, the sill and the variance of a variogram may vary in different directions. In such cases, the directions in which the spatial continuity are maximum and minimum need to be identified to better describe the spatial continuity.

To get an initial understanding of the structural continuity, omnidirectional variograms were first computed using the average differences of data pairs  $Z(x_i)$  and  $Z(x_i + h)$ , separated by a specified  $h$  increment in the Oak and Kumbur mine areas as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (4.1)$$

where  $N$  represents the number of pairs for a given  $h$  increment. For the Oak borehole data, the experimental omnidirectional variogram was computed using a lag distance of 76.2 m, lag tolerance of 38.1 m and 10 lags. Since the GPR and floor survey data of the Oak mine are densely sampled, a lag distance of 15 m,

a lag tolerance of 7.5 m and 50 lags were chosen to compute the omnidirectional variograms. As for the Kumbur borehole and PCD variables, the lag distance was chosen based on the spacing of the PCD, which is 19.05 m. Lag tolerance for these variables was chosen as 9.53 m and the variograms had 50 lags. The floor survey and the GPR data variograms for the Kumbur area were computed using the same lag distance and tolerances as in the Oak GPR data, which were 15 m and 7.5 m, respectively. The resulting omnidirectional variograms of the raw point-type data can be seen in Fig. 4.1.

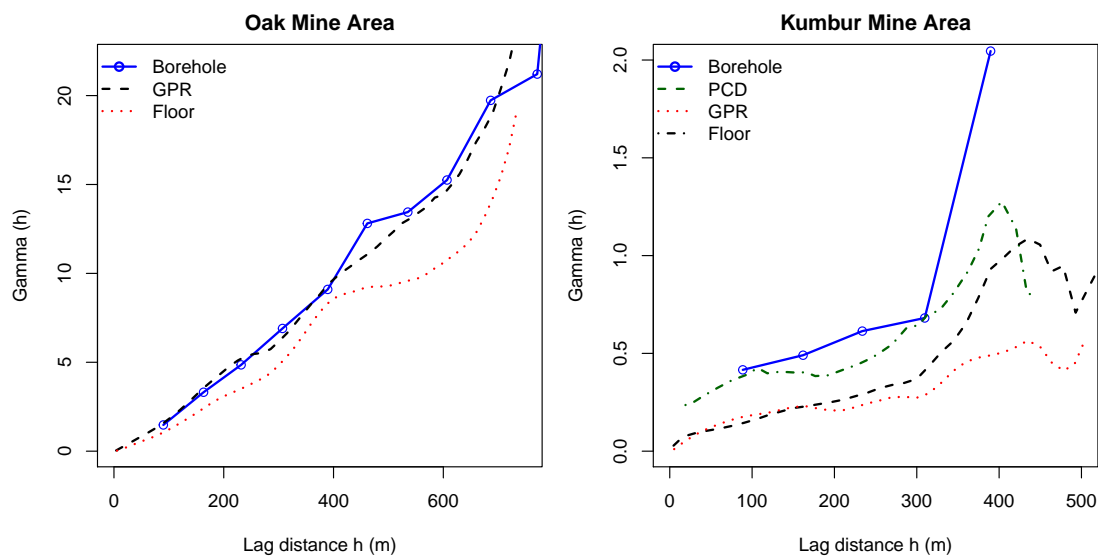


Figure 4.1: Omnidirectional variograms computed for the Oak and Kumbur mine areas

The variograms computed for the Oak mine indicate the existence of a non-stationarity as there is a lack of evidence for a finite sill. Therefore, the geological interface between the bauxite/ferricrete units appears to have a spatial trend. As for the Kumbur mine, the variograms also exhibit a non-stationarity behaviour. However, the non-stationarity for this mine seems to be weak and the rise in the variogram values tend to stop and drop down after the lag distance of 450 m.

Determination of the optimum drifts was carried out by using the non-stationarity modelling tool of the ISATIS software package (Bourassi et al., 2016). The types

of the drifts analysed include the universality condition, the linear drift (first order) and the quadratic drift (second order). The degrees of polynomials providing the most suitable estimates were identified using the cross-validation technique. The type of the neighbourhood used for the GPR and the floor survey data was of moving type, since these data sets were dense and irregularly spaced. The borehole data and PCD comprised of limited number of points. Hence, unique neighbourhood type was utilised for these data sets. The search neighbourhoods used had a maximum distance of 300m. Moreover, eight angular sectors with six points per sector were utilised. The most suitable drift types using this configuration were the first order for the Oak mine variables and the second order for the Kumbur variables. Having identified the drift types, the coefficients of the drifts were then determined through the methods of least squares, as can be seen in Table 4.1. The trend surfaces constructed using these coefficients were then used to obtain the residuals.

Table 4.1: Drift Coefficients Determined for the Oak and Kumbur Mine Variables

Variable	UC	X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY
Oak Borehole	-62.37	$-15.54 \times 10^{-5}$	$4.85 \times 10^{-5}$	-	-	-
Oak GPR	-58.59	$-17.68 \times 10^{-5}$	$4.11 \times 10^{-5}$	-	-	-
Oak Floor	-65.09	$-11.25 \times 10^{-5}$	$5.80 \times 10^{-5}$	-	-	-
Kumbur PCD	-1649.55	$-14.90 \times 10^{-4}$	$1.39 \times 10^{-3}$	$-8.41 \times 10^{-10}$	$-3.03 \times 10^{-10}$	$-4.43 \times 10^{-10}$
Kumbur GPR	-4355.40	$29.80 \times 10^{-4}$	$3.80 \times 10^{-3}$	$-2.29 \times 10^{-9}$	$-8.65 \times 10^{-10}$	$-6.89 \times 10^{-10}$
Kumbur Floor	-1588.10	$72.74 \times 10^{-5}$	$1.44 \times 10^{-3}$	$-2.39 \times 10^{-9}$	$-3.91 \times 10^{-10}$	$4.64 \times 10^{-10}$

### 4.3.2 Migrating the Data Points to the Grid Nodes and Constructing the TI

The simulations require the data to be located on a predefined grid. Therefore, if the data is in point form, it should be converted to grid data by migrating the data points to the nodes of the grid prior to the simulations. As far as the primary simulation variable is concerned, the migration process is handled within the DS algorithm itself. On the other hand, the migration of the other variables, which are the secondary simulation variable and the variables of the bivariate TI, are

performed outside of the DS algorithm. Such a preliminary step is required to both fill the gaps between the data point locations as well as converting the point data into the grid type data.

The raw data from the Oak and Kumbur mine areas were in the form of point data. Therefore, they had to be migrated to the grids created for each mine area. For this reason, two grids for both Oak and Kumbur mine areas were created. The constructed grids have a single grid cell (pixel) size of  $2.38 \times 2.38$  m. The sizes of the grids in terms of number of grid nodes are  $180 \times 400$  for the Oak mine and  $97 \times 214$  for the Kumbur mine.

Since the GPR and the mined-out survey data are sampled irregularly at finite locations, the collected data points cannot be directly used to inform all the grid nodes. In other words, after assigning the available data points to the closest grid nodes, some of the nodes may still be uninformed. Therefore, reconstruction of the missing elevation variable is needed to create the full picture of the TIs and auxiliary variables (GPR). The elevation values could be assigned through interpolation techniques by utilising the neighbouring data points. However, such an approach would lead to a smoothing effect and the resulting re-constructed images might have different spatial statistical properties than the original data sets. The reconstruction task can be achieved by the use of geostatistical simulation techniques, such as the sGs. A simple example illustrating the process involved in such a reconstruction approach can be seen in Fig. 4.2.

Informing the uninformed grid node locations using a simulation technique has a benefit of preserving the original statistical and spatial statistical properties of the data. However, since the sGs utilises a variogram model, which is two-point statistics, to perform the simulations, the assigned values might lack higher-order statistics. Nevertheless, since the raw data set is rather densely sampled, the majority of the grid nodes are informed by the collected samples. Hence, the extent of the multiple-point statistical loss can be ignored and the resulting simulations are still considered to exhibit the multiple-statistical information of the raw data.

An important aspect that needs to be considered is the trends evident in the



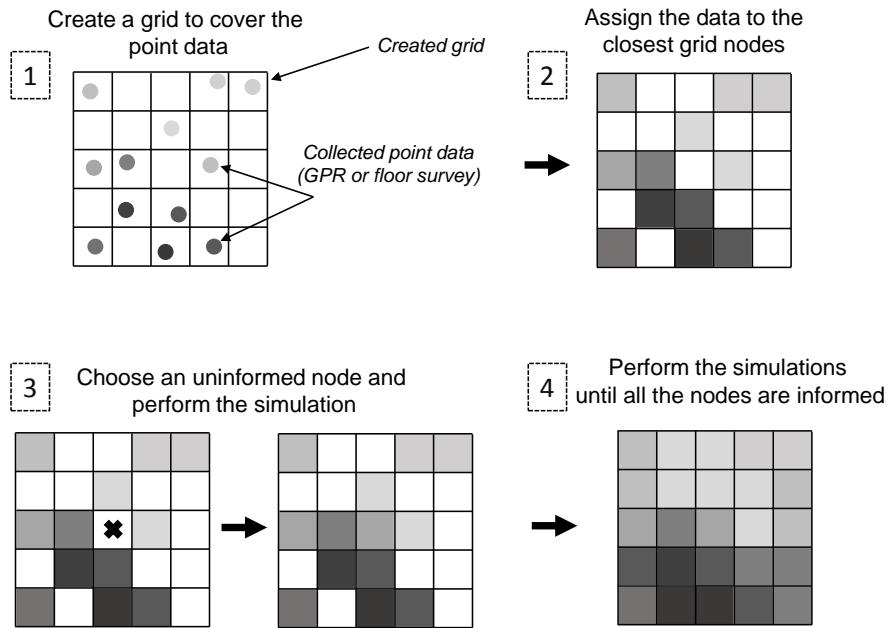


Figure 4.2: Process used to both convert the point data into the gridded type and reconstruct the images

datasets, as they can have an effect on the accuracy of the simulated values. However, since the GPR and floor surveys are densely sampled, the data points used to simulate a grid node are always within close proximity of the uninformed grid nodes. Hence, the effect of the trend is less pronounced in the simulated values of the GPR and floor surveys. This fact can be confirmed by the study presented in Dagasari et al. (2018a) (given in Appendix J). The study comprised the performance comparison of stationary and non-stationary modelling techniques to reproduce the elevation field in the Oak mine area using the GPR data. Universal Kriging and IRF-k were used as non-stationary geostatistical techniques. Hence, the trend was taken into consideration. On the other hand, the OK was used to perform the same task by ignoring the trend. The variogram for the OK was modelled by first identifying the variogram-free direction and fitting a model to the experimental variogram computed along this direction. The results of this study have shown that the performances of the three methods were almost identical due to the existence of an abundance of data points. Therefore, the variograms to

perform reconstruction work through the sGs in this thesis were modelled using the experimental variograms calculated along the variogram-free directions.

In order to identify the variogram-free directions, the variogram maps of the GPR and floor surveys of both mine areas were generated first, as can be seen in Appendix C. The variogram map created for the Kumbur GPR variable is also illustrated in Fig. 4.3. This map reveals the trend free direction as N130. The experimental variogram calculated along the N130 direction and the fitted model can also be seen in 4.3. The model fitted is comprised of two spherical structures with ranges of 19.33 m and 306.05 m. The sill values of these are 0.065 and 0.056, respectively. The same procedure was also followed to migrate the points of the Kumbur floor, Oak GPR and Oak floor survey variables. The data before and after the re-construction process can be seen in Fig. 4.4.

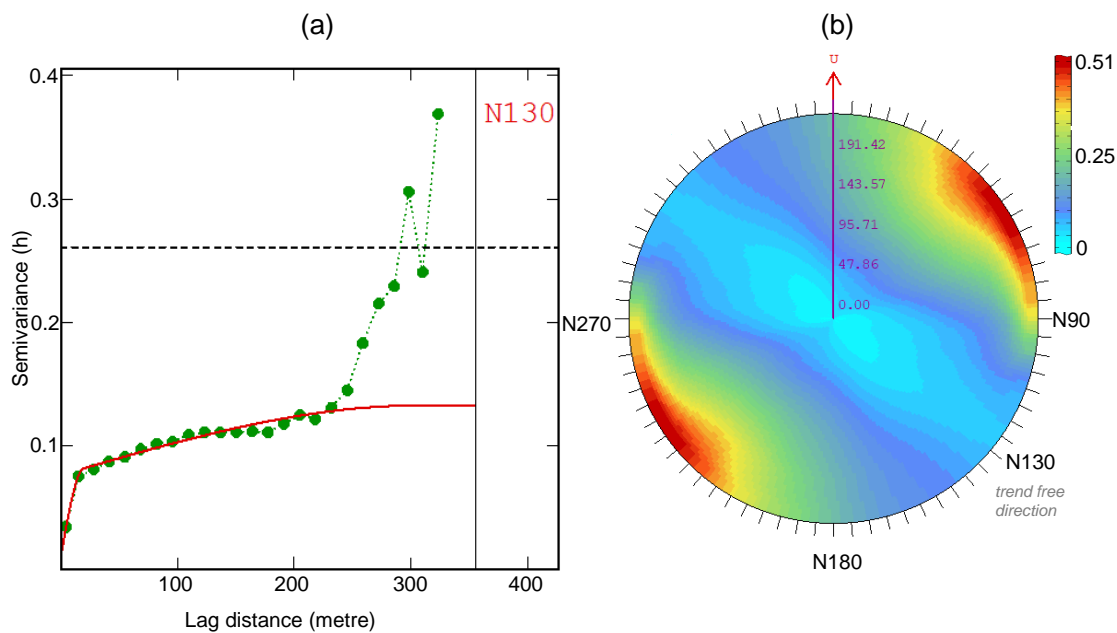


Figure 4.3: (a) Experimental variogram calculated along the trend-free direction (N130) and (b) the variogram model fitted for the Kumbur GPR variable

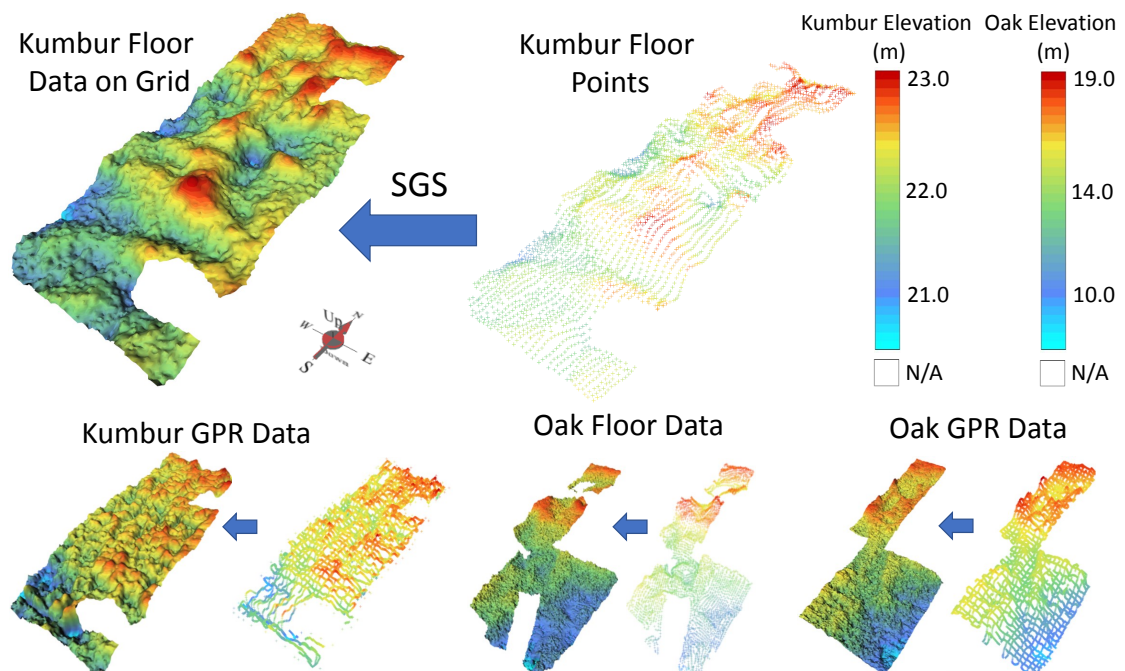


Figure 4.4: Re-construction of the GPR and floor surveys of the Oak and Kumbur mine areas

## 4.4 Setting up the MPS Simulations

The MPS simulations in this study are performed using the DS algorithm. Selecting the DS was mainly due to the benefits it offers such as the ability to perform multivariate and continuous simulations. Furthermore, since the DS does not scan the whole TI to store the conditional probabilities, it performs the simulations rather fast.

As in all the MPS algorithms, the DS requires a preliminary set-up process in which the information about the simulations and the algorithmic input parameters are specified. Examples of such information include the number of variables used in the simulations, the locations where the data are stored in the computer, and the size and origin of the simulation grid. The following subsections will provide explanations on the selection of the simulation variables as well as the specification of algorithmic input parameters to perform the simulations in this research. An overview of the inputs and outputs of the bivariate DS simulations can be seen in

Fig. 4.5.

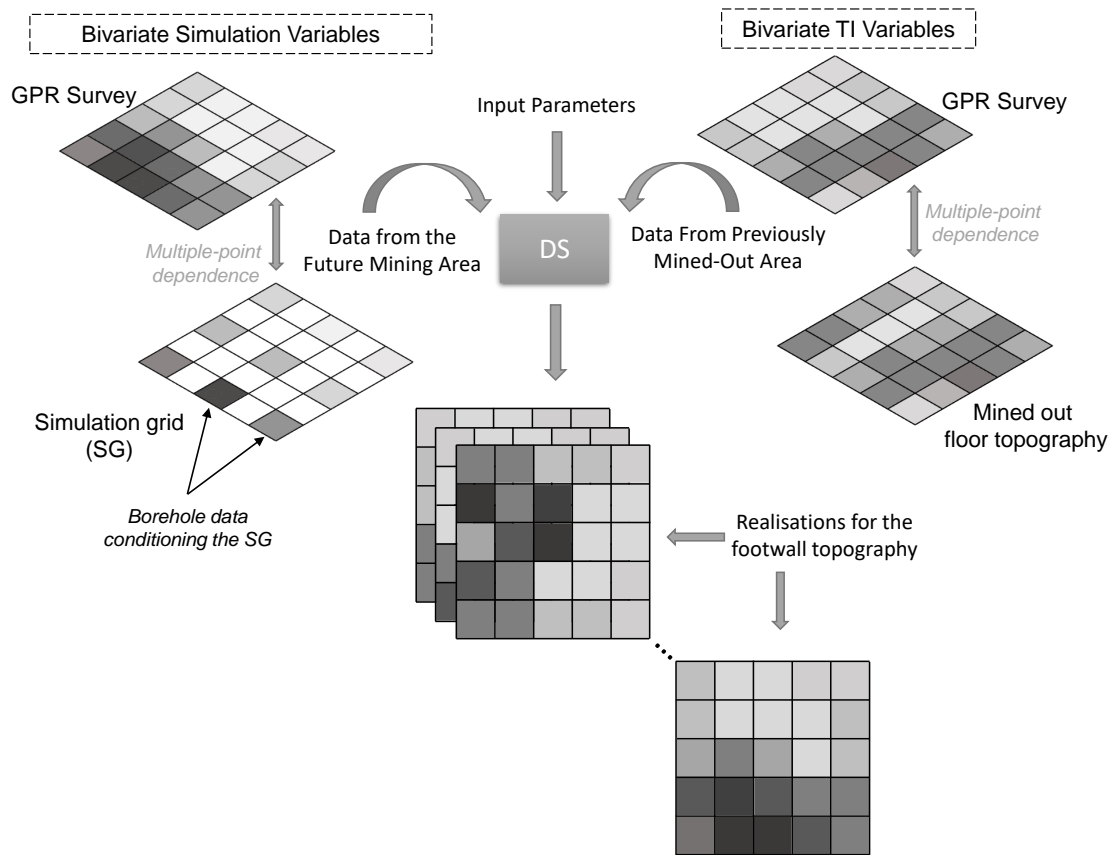


Figure 4.5: Inputs and outputs of the bivariate DS simulations

#### 4.4.1 Choosing the Simulation Variables

There are mainly two types of data collected from the Oak and Kumbur mine areas for the resource modelling. The first type is the drill hole data obtained from the exploration boreholes or PCD. The second type is the geophysical data acquired from the GPR surveys. The GPR data provides the depth to the geological interface from the surface through two-way travel time. However, this information is not as reliable as the borehole data information. Therefore, the GPR data has been used as an auxiliary variable to guide the simulations whereas the PCD/borehole data has been used as the primary variable to condition the

simulations. Conditioning the simulations, as also described in Section 4.3.2, is carried out internally within the DS algorithm prior to the simulation process. Since the borehole and PCD data are drilled on a regular grid (equally spaced), their locations are co-located with the grid node locations. Hence, the original raw data locations do not undergo any translation during the migration process.

Due to the existence of two kinds of data representing the elevation variable of the geological contact, the simulations are performed in both univariate and bivariate forms. For the univariate simulations, the mined-out surface of one mine area is used as a TI to perform the simulations for the other mine area. As for the bivariate simulations, the constructed TIs comprise the mined-out topographies as the primary variable and the GPR surveys as the secondary variable. Therefore, the multiple-point dependence between the variables is also utilised to guide the simulations. This is carried out by constructing bivariate data events to scan the TI until the distances in each variable fall below the defined thresholds or the maximum scan fraction is reached.

#### 4.4.2 Algorithmic Input Parameter Selection

Although the MPS simulations are predominantly affected by the TI selection and the simulation variables, the choice of the input parameters also plays a significant role on the simulation results. Therefore, the appropriate simulation parameters need to be identified prior to the simulations. A common way to achieve this is to perform a manual sensitivity analysis using different input parameters and seek the ones yielding good simulation quality. To carry out such an analysis, five DS parameters were selected. For the bivariate simulations, the parameters include the weighting factor  $w_{hd}$  assigned to the conditioning data, acceptance thresholds ( $t_{hd}$  and  $t_{GPR}$ ), and the number of neighbours ( $n_{hd}$  and  $n_{GPR}$ ) of the DS algorithm. Some of the parameters were not selected for the sensitivity analyses. These include the weighting factor  $w_{GPR}$  of the GPR data and the scan fraction  $f$ . The reason why  $w_{GPR}$  was not selected was that the GPR grid was already exhaustively informed and the change in the GPR weight would not have had an effect on the simulations. As for the scan fraction, the preliminary parameter trials

conducted have shown that it had a minor effect on the simulations. Therefore, it was chosen as 0.5 for both the sensitivity analysis and the simulations. It should be noted that 0.5 for the scan fraction is also suggested by Meerschman et al. (2013) for continuous simulations.

Having tested a number of different DS parameter combinations manually, the appropriate DS parameters were finally determined. However, during this sensitivity analysis, it was understood that the manual determination of the parameters were rather cumbersome and time-consuming. Therefore, an approach to automatically tune the input parameters was developed. The main idea of the proposed approach was to find the parameters which best reproduce the patterns of the borehole data. This is carried out by first computing the pattern statistics of the conditioning data and comparing it with that of the simulations after each and every parameter trial. This process is automatically performed by utilising the optimisation framework of the SA algorithm. The parameters are used as the decision variables of the SA and they are iteratively optimised to achieve minimum input-output statistical dissimilarity. The details of the proposed approach are presented in both Chapter 5 and Dagasan et al. (2018b). Table 4.2 shows the automatically determined parameters using the proposed approach for the Kumbur area. All the parameters used to perform the simulations can be seen in Table 4.3.

Table 4.2: Optimised DS parameters used to simulate the position of the geological contact in Kumbur mine area

Parameters	$w_{hd}$	$n_{hd}$	$n_{GPR}$	$t_{hd}$	$t_{GPR}$
Tuned Kumbur Bivariate Simulations	13.08	14	45	0.009	0.305

## 4.5 Results and Comparison

The simulation results were analysed using 40 bivariate and univariate realisations generated for each mine area. Considering the fact that the simulations were carried out using the DS and TB methods, each mine area contained four sets of simulation results (univariate and bivariate simulations using the DS and TB

Table 4.3: DS parameters used to simulate the position of the geological contact in Oak and Kumbur mine areas

Parameters	$w_{hd}$	$n_{hd}$	$n_{GPR}$	$t_{hd}$	$t_{GPR}$
Kumbur Bivariate	13.08	14	45	0.009	0.305
Kumbur Univariate	13.08	14	-	0.009	-
Bivariate Oak	10.00	20	21	0.001	0.01
Univariate Oak	10.00	6	-	0.001	-

methods). Therefore, the results occupy a large volume. Due to this reason, only the bivariate simulations for the Kumbur mine are presented in the chapter. The rest of the results are, therefore, presented in the appendices.

The omnidirectional variograms exhibited non-stationarity in the variables. Therefore, the simulations were performed using the residuals. To do this, the trend surfaces constructed using the coefficients in Table 4.1 were subtracted from the raw data to obtain the residuals, as can be seen in Appendix B. The resulting residual data sets were then used as the simulation and TI variables. The residual simulation variables and the TI variables for the bivariate Kumbur simulations can be seen in Fig. 4.6. After performing the simulations, the trends were added back to each of the realisations generated.

For the geostatistical simulations, the TB with external kriging method in the RGeostats R package (Renard et al., 2017) was utilised. The GPR variable was used as an external drift and facilitated obtaining the residuals of the PCD. The resulting residuals were used to create the variogram model to perform the simulations. Experimental variograms computed in various directions demonstrated a geometric anisotropy;  $N0^\circ$  and  $N90^\circ$  directions appeared to be the main anisotropy directions. Two nested models consisting of exponential and spherical structures were fitted to the computed experimental variograms in the major anisotropy directions. The sill value of the spherical component was 0.137 and the ranges were 112 and 237 in the  $N0^\circ$  and  $N90^\circ$  directions, respectively. On the other hand, the sill value of the exponential component was 0.2 and the ranges were 38 m along  $N0^\circ$  and 35 m along  $N90^\circ$ . The variogram model created using the given range and sill parameters can be seen in Fig. 4.7. After modelling the variogram, the TB

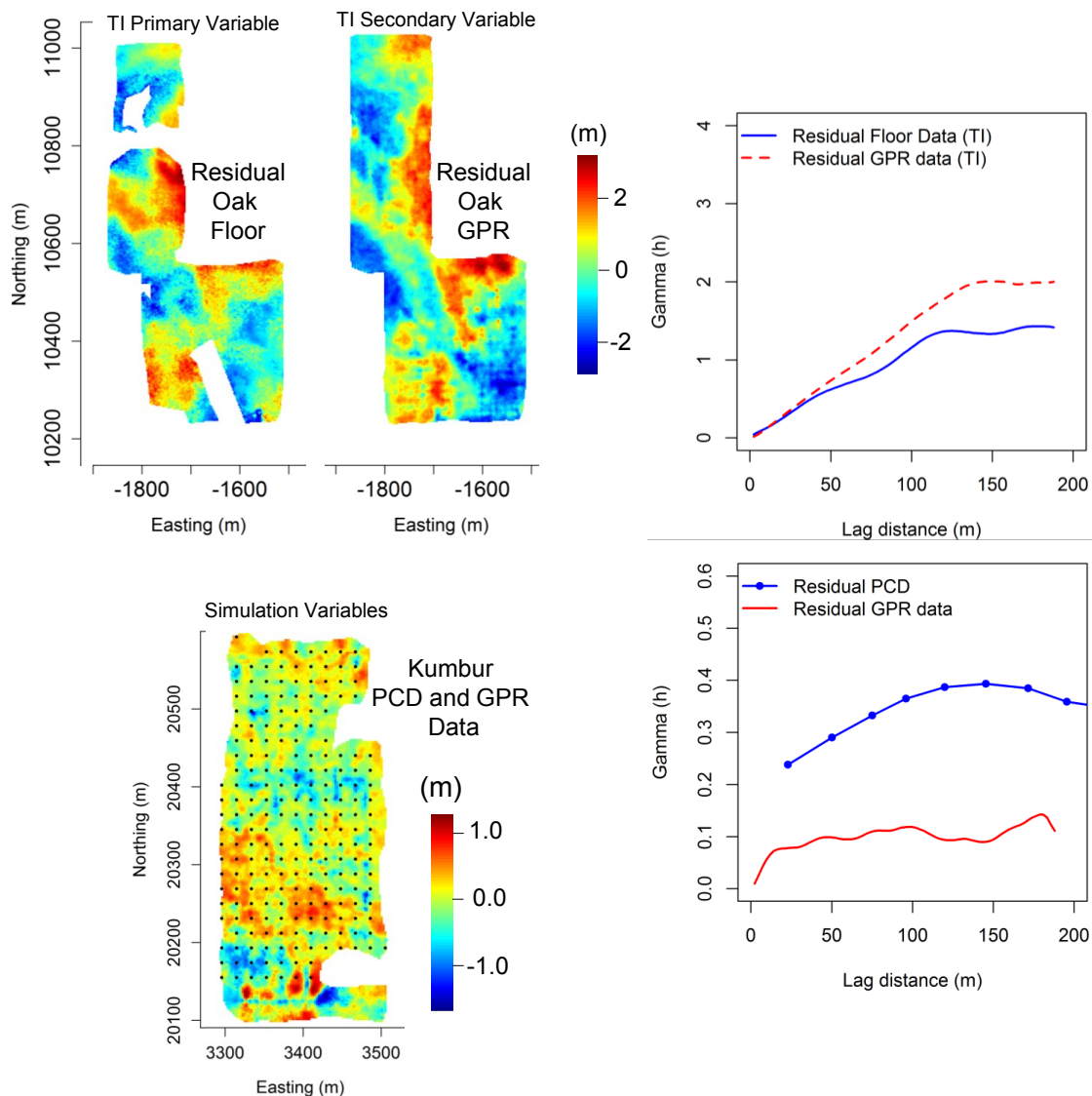


Figure 4.6: Bivariate TI, simulation variables and the omnidirectional residual variograms for the Kumbur mine area

simulations were performed utilising 400 turning bands in the simulations. Using the setup explained, 40 realisations were generated using the DS and TB methods. The first three realisations of the simulations can be seen in Fig. 4.8. The average of the realisations and the variograms of the realisations are illustrated in Fig. 4.9.

Performance evaluations of the DS and TB simulations have been undertaken using various statistical indicators. The first performance indicator used was the



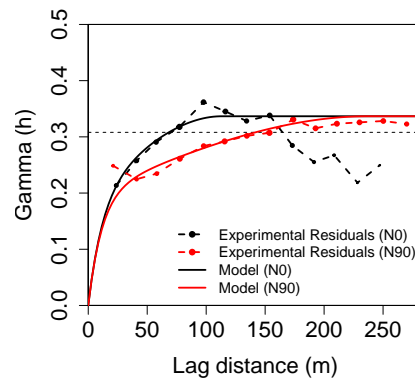


Figure 4.7: Experimental variogram of the PCD residuals along with the fitted model

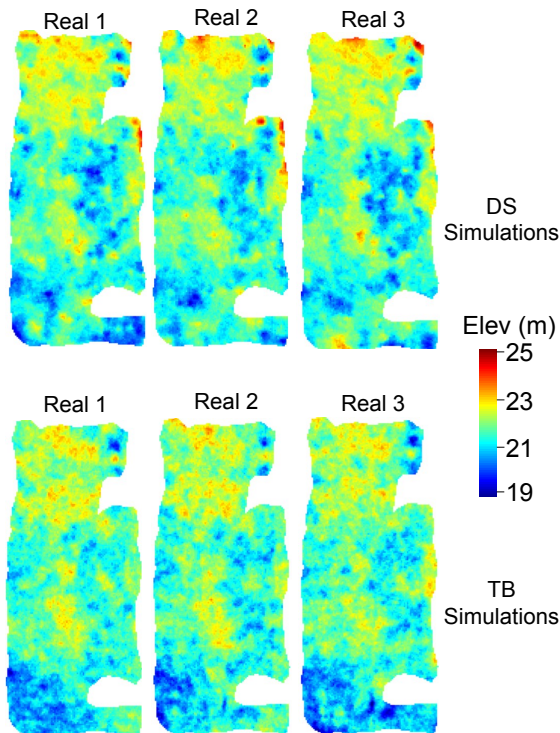


Figure 4.8: First three realisations of the DS and TB simulations

variogram reproduction of the conditioning data used. Omnidirectional variograms of the residuals in Fig. 4.6 reveal the difference between the two-point statistics

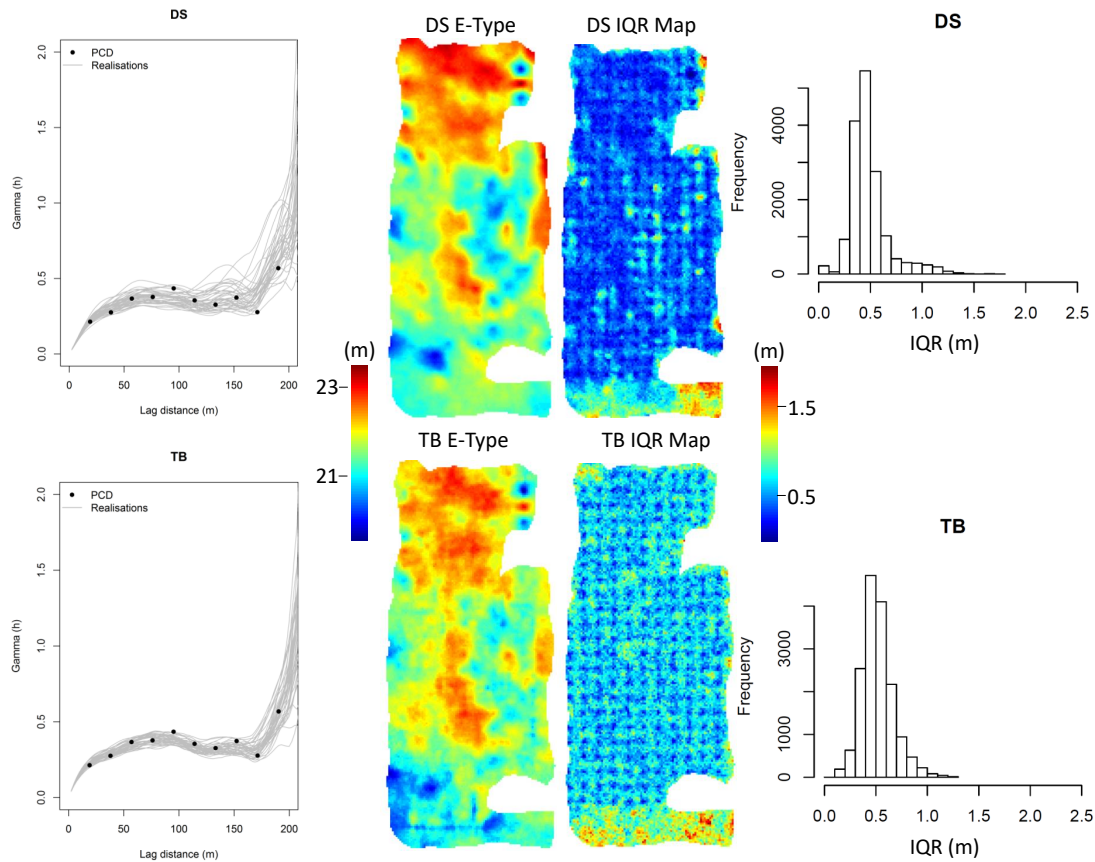


Figure 4.9: Average of the TB and DS realisations together with the IQR distributions and the variograms of the realisations

of the simulation variables and the TI used to perform the simulations. Such a difference can rightfully be interpreted as an indication of a compatibility issue. Nevertheless, in spite of such a difference, variograms of the realisations illustrated in 4.9 have shown that the DS was capable of reproducing two-point statistics of borehole data successfully. It is also observed from the variograms in Fig. 4.9 that both methods reproduce the variogram of the conditioning data well and the simulations are rather similar.

To be able to observe the visual differences in the results between the two simulation techniques, the mean of the simulations and the variations around the mean of the simulated values are analysed. Considering the generated  $L$  realisations, the average of the simulations (E-type maps) were created by taking the average of

the simulated values at each  $x$  discrete grid node location:

$$z_E^*(x) = \frac{1}{L} \sum_{l=1}^L z^l(x) \quad (4.2)$$

where  $z_E^*(x)$  represents the expected elevation values and  $z^l(x)$  represents the elevation values at  $x$  locations of the  $l^{th}$  realisation. The created maps in Fig. 4.9 exhibit only a slight difference between them. Hence, it is realised that the average of the DS and TB realisations are fairly similar.

The variation of the simulated values around the mean was used to analyse the generated uncertainty model. To obtain such variations, the interquartile-range (IQR) of the simulated elevations at each grid node was calculated by subtracting the lower quantile values from the upper quantile ones:

$$q_R(x) = q_{0.75}(x) - q_{0.25}(x) \quad (4.3)$$

The IQR maps shown in Fig. 4.9 indicate that the DS tends to produce lower IQR values; hence, less uncertainty. The calculated average of the IQR values are 0.426 for the DS simulations and 0.501 for those of the TB.

Comparison of the simulation performances has also been made in terms of higher-order statistics. Computation of higher-order statistics has been carried out by the *hosc* software, which utilises spatial cumulants (Dimitrakopoulos et al., 2010). Third-order experimental cumulant maps were computed using an L-shaped spatial template constructed along the  $N0^\circ$  and  $N90^\circ$  directions. Lag separations used were 2.38 m and the cumulant maps were computed for 50 lag distances in both directions. The resulting cumulant maps of the realisations as well as the mined-out surface referenced are shown in Fig. 4.10.

The resulting cumulant maps look fairly different and do not give adequate information on the resemblance between the realisations and the reference mined-out surface. Hence, the similarity between the cumulant maps needed to be expressed quantitatively. This calculation was performed by subtracting the cumulant maps of each of the 40 realisations from that of the mined-out surface pixel-by-pixel, as

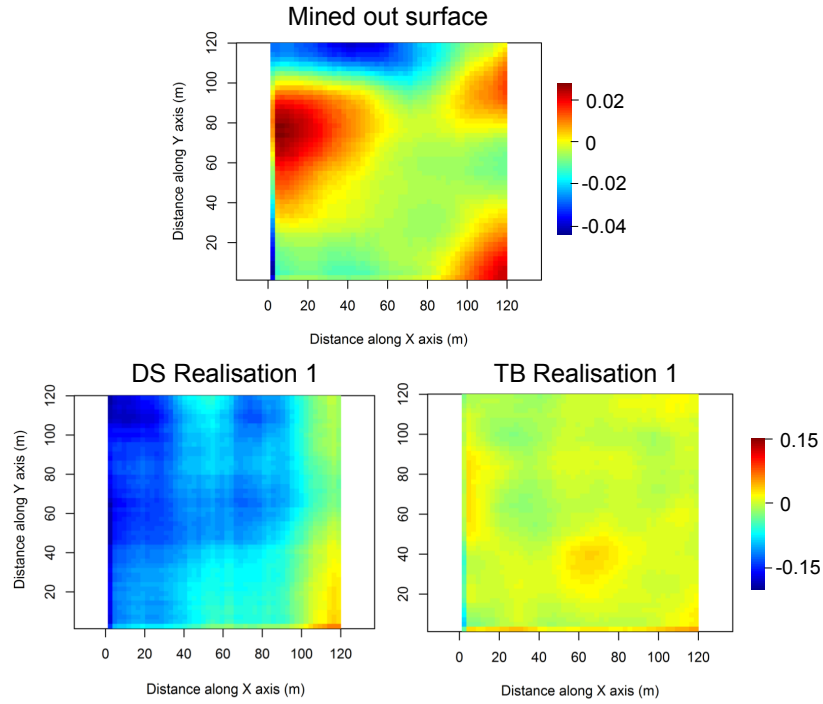


Figure 4.10: Cumulant maps of the Kumbur mined-out floor and the first realisations of the DS and TB Kumbur simulations

follows:

$$error_C^l = \frac{1}{N} \sum_{j=1}^N |c^l(j) - c^{(ref)}(j)| \quad (4.4)$$

where  $c^l(j)$  and  $c^{(ref)}(j)$  are the calculated cumulant values at the  $j^{th}$  pixel of the maps and  $N$  is the number of grids (pixels) that the cumulant maps comprise. Distribution of the mean absolute errors  $error_C^l$  for each realisation  $l$  can be seen in Fig. 4.11.

These histograms demonstrate that both simulation methods yield similar results in terms of the generated higher-order statistics. The means of the errors in 40 realisations are both  $5 \times 10^{-7}$ . Hence, both methods produce the same multiple-point statistical errors.

The last performance comparison was carried out by measuring the similarity

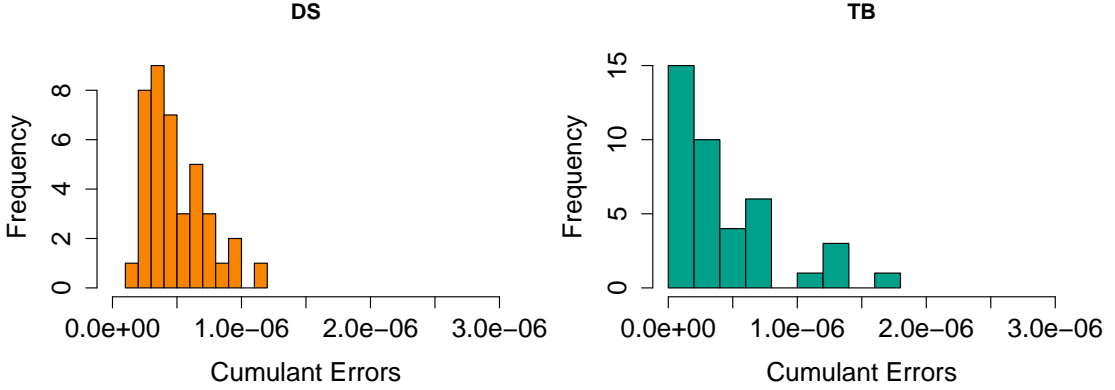


Figure 4.11: Histograms of the mean cumulant errors for each realisation

between the realisations and the mined-out surface of the simulation area. This similarity was calculated by subtracting the realisations from the mined-out surface pixel-by-pixel and taking the mean of the absolute differences:

$$error_{Ref} = \frac{1}{LK} \sum_{l=1}^L \sum_{i=1}^K |Z^l(i) - Z^{(ref)}(i)| \quad (4.5)$$

where  $K$  represents the number of grids in the realisations/mined-out image, and  $Z^l(i)$  and  $Z^{(ref)}(i)$  are the elevation values in the  $l^{th}$  realisation and the mined-out image, respectively. The calculated error for the DS and TB are 0.534 and 0.528, respectively. Based on these results, it is also obvious that the difference between the DS and TB simulations is insignificant for the bivariate Kumbur simulations.

The above analyses have also been performed for the Bivariate Oak, Univariate Oak and Univariate Kumbur simulations. The results of all the simulations are summarised using three performance indicators with respect to the mined-out reference surfaces: IQR, cumulant error and MAE. The results presented in Table 4.4 and Fig. 4.12 show that the DS consistently produced less uncertainties than the TB. As for the cumulant errors, the DS yields less cumulant errors in three out of four cases. This is intuitively expected since the simulations in DS are performed within the MPS framework. The direct comparison with the mined-out surfaces indicates that the DS is again slightly better than the TB as the errors it

yields are less in two cases (univariate and bivariate Oak), is the same in one case (bivariate Kumbur) and is greater in only one case (univariate Kumbur).

Table 4.4: Summary of the simulation results

Simulation Type	Mine Area	Method	Uncertainty (IQR)	Cumulant Error	MAE
Univariate	Oak	DS	0.69	0.0013	0.92
		TB	0.99	0.0028	0.94
	Kumbur	DS	0.48	1.5e-6	0.54
		TB	0.52	3.0e-7	0.53
Bivariate	Oak	DS	0.80	0.0011	0.87
		TB	0.96	0.0122	0.94
	Kumbur	DS	0.43	5e-7	0.53
		TB	0.50	5e-7	0.53

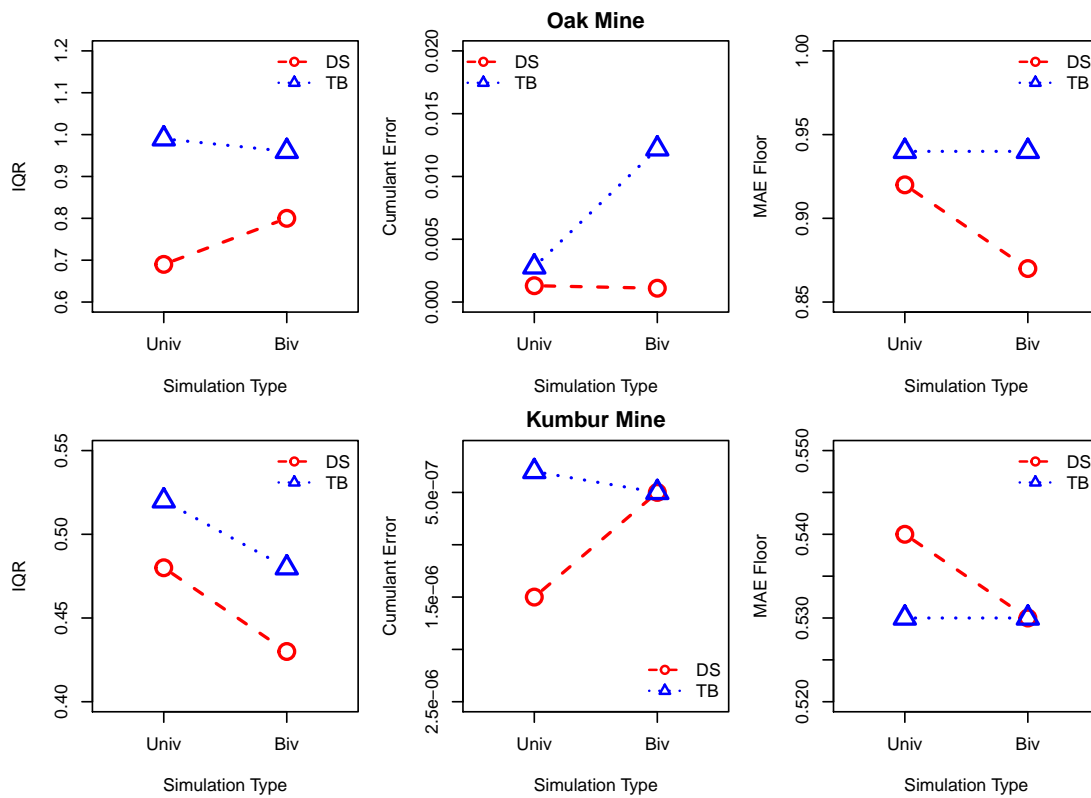


Figure 4.12: Summary of the performance indicators used to compare the TB and DS techniques for the Oak and Kumbur mine areas

## 4.6 Summary and Discussion

This chapter investigated the use of multiple-point statistics in modelling the foot-wall topographies in lateritic metal deposits. The training images used to perform the simulations comprised the mined-out floor and the ground penetrating radar surveys collected from previously mined-out areas. Due to the existence of two types of simulation variables representing the bauxite/ferricrete interface (bore-hole and geophysical data), the simulations were performed in the form of both univariate and bivariate types. The resulting multiple-point statistical simulations were compared with classical geostatistical simulations through various statistical indicators.

Concerning the comparisons made between the direct sampling and turning bands simulations, both techniques produced reasonably similar results. Therefore, it can be deduced that the use of multiple-point statistics did not remarkably enhance the contact models. However, it has been observed that it can be a promising tool to be used in modelling lateritic bauxite deposits. Supporting evidence to this is the successful variogram reproduction in spite of a non-compatibility between the TI and the simulation variables. Therefore, if a better training image in terms of the training image-conditioning data compatibility is derived, it is believed that substantial simulation performance could be achieved. Obtaining a compatible training image in lateritic metal mines is rather straightforward since extracting the ore at a mining area results in a mined-out topography. Provided that a group of mined-out regions are surveyed, they can be used to construct a catalogue of training images to simulate the future mining areas. Such a catalogue can provide geologists with a variety of training images with different spatial statistics. Therefore, a suitable training image with similar statistical and structural characteristics observed in the conditioning data set can be chosen easily. Lastly, the use of multiple-point statistics provides a non-parametric modelling framework. Therefore, it is rather non-expert friendly as the structural model is not expressed in mathematical definitions.

One important aspect that is considered to affect the quality of the training images is the digging precision of the mining machinery used to extract the deposit.

Since the bauxite ore is of pisolitic type and is located within the soil horizons, it is extracted using a front-end loader type mining machine. Bucket dimensions of the machinery are thought to influence the shape of the resulting mined-out surface significantly; the smaller the bucket dimensions of the machinery, the better the digging precision is in tracking the bauxite/ferricrete interface. Furthermore, the actual interface is sensed by the mining machinery operator utilising the hardness and colour differences between the bauxite and ferricrete units at the time of mining. Since the decision on the location of the orebody boundaries is made by the operator at the time of mining, the resulting reference geology is subjective to some extent. In other words, the resulting mined-out surface might be slightly off from the actual geological interface due to the decision made by the operator. To alleviate this problem, a pilot excavation area can be dug more carefully using smaller mining equipment next to a future mining area. Hence, the geological variations can be better illuminated. This approach could allow generating better quality training images and is believed to improve the quality of the training images produced.



## References

- Arpat, G. B. and Caers, J. (2007). Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203.
- Bardossy, G., Szabo, I., and Varga, G. (2003). A new method of resource estimation for bauxite and other solid mineral deposits. *BERG UND HUTTENMANNISCHE MONATSSHEFTE*, 148(2):57–64.
- Bourassi, A., Foucher, B., Geffroy, F., Marin, J., Martin, B., Meric, Y., Perseval, S., Renard, D., Robinot, L., Touffait, Y., and Wagner, L. (2016). Isatis software user’s guide. *Geovariances, Ecole des Mines de Paris, Paris*.
- Dagasan, Y., Erten, O., and Topal, E. (2018a). Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study. *The Journal of South African Institute of Mining and Metallurgy*, 118(2):173–184.
- Dagasan, Y., Renard, P., Straubhaar, J., Erten, O., and Topal, E. (2018b). Automatic parameter tuning of multiple-point statistical simulations for lateritic bauxite deposits. *Minerals*, 8(220).
- de Freitas Silva, M. and Dimitrakopoulos, R. (2016). Simulation of weathered profiles coupled with multivariate block-support simulation of the puma nickel laterite deposit, brazil. *Engineering Geology*, 215:108–121.
- De Iaco, S. and Maggio, S. (2011). Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. *Mathematical Geosciences*, 43(4):483–500.
- Dimitrakopoulos, R. (1998). Conditional simulation algorithms for modelling ore-body uncertainty in open pit optimisation. *International Journal of Surface Mining, Reclamation and Environment*, 12(4):173–179.
- Dimitrakopoulos, R., Mustapha, H., and Gloaguen, E. (2010). High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-gaussian and non-linear phenomena. *Mathematical Geosciences*, 42(1):65–99.

- Gloaguen, E. and Dimitrakopoulos, R. (2009). Two-dimensional conditional simulations based on the wavelet decomposition of training images. *Mathematical Geosciences*, 41(6):679–701.
- Goodfellow, R., Consuegra, F. A., Dimitrakopoulos, R., and Lloyd, T. (2012). Quantifying multi-element and volumetric uncertainty, coleman mcreedy deposit, ontario, canada. *Computers & geosciences*, 42:71–78.
- Guardiano, F. and Srivastava, R. (1992). Borrowing complex geometries from training images: The extended normal equations algorithm. *Stanford Center for Reservoir Forecasting Report, Stanford University*.
- Honarkhah, M. and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42(5):487–517.
- Jones, P., Douglas, I., and Jewbali, A. (2013). Modeling combined geological and grade uncertainty: application of multiple-point simulation at the apensu gold deposit, ghana. *Mathematical Geosciences*, 45(8):949–965.
- Journel, A. G. (2005). Beyond covariance: the advent of multiple-point geostatistics. *Geostatistics Banff 2004*, pages 225–233.
- Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.
- Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11).
- Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., and Renard, P. (2013). A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. *Computers & Geosciences*, 52:307–324.

- Mery, N., Emery, X., Cáceres, A., Ribeiro, D., and Cunha, E. (2017). Geostatistical modeling of the geological uncertainty in an iron ore deposit. *Ore Geology Reviews*, 88:336–351.
- Osterholt, V. and Dimitrakopoulos, R. (2018). Simulation of orebody geology with multiple-point geostatistics application at yandi channel iron ore deposit, wa, and implications for resource uncertainty. In *Advances in Applied Strategic Mine Planning*, pages 335–352. Springer.
- Pasti, H. A., Costa, J. F. C. L., and Boucher, A. (2012). Multiple-point geostatistics for modeling lithological domains at a brazilian iron ore deposit using the single normal equations simulation algorithm. In *Geostatistics Oslo 2012*, pages 397–407. Springer.
- Renard, D., Bez, N., Desassis, N., Beucher, H., Ors, F., and Freulon, X. (2017). *RGeostats: Geostatistical Package*. R package version 11.0.6.
- Rezaee, H., Mariethoz, G., Koneshloo, M., and Asghari, O. (2013). Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. *Computers & geosciences*, 54:293–308.
- Rossi, M. E. and Deutsch, C. V. (2013). *Mineral resource estimation*. Springer Science & Business Media.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical geology*, 34(1):1–21.
- Tahmasebi, P., Hezarkhani, A., and Sahimi, M. (2012). Multiple-point geostatistical modeling based on the cross-correlation functions. *Computational Geosciences*, 16(3):779–797.

- van der Grijp, Y. and Minnitt, R. (2015). Application of direct sampling multi-point statistic and sequential gaussian simulation algorithms for modelling uncertainty in gold deposits. *Journal of the Southern African Institute of Mining and Metallurgy*, 115(1):73–85.
- Zhang, T., Switzer, P., and Journel, A. (2006). Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38(1):63–80.

## Chapter 5

# Automatic Parameter Tuning of Multiple-Point Statistical Simulations for Lateritic Bauxite Deposits<sup>1</sup>

---

<sup>1</sup>This chapter has been published in *Minerals* as:

Dagasan, Y., Renard, P., Straubhaar, J., Erten.O and Topal.E. Automatic Parameter Tuning of Multiple-Point Statistical Simulations for Lateritic Bauxite Deposits *Minerals*, DOI: 10.3390/min8050220 and reprinted with permission in Appendix I

**Abstract**

The application of multiple-point statistics (MPS) in the mining industry is not yet widespread and there are very few applications so far. In this paper, we focus on the problem of algorithmic input parameter selection, which is required to perform MPS simulations. The usual approach for selecting the parameters is to conduct a manual sensitivity analysis by testing a set of parameters and evaluating the resulting simulation qualities. However, carrying out such a sensitivity analysis may require significant time and effort. The purpose of this paper is to propose a novel approach to automate the parameter tuning process. The primary criterion used to select the parameters is the reproduction of the conditioning data patterns in the simulated image. The parameters of the MPS algorithm are obtained by iteratively optimising an objective function with simulated annealing. The objective function quantifies the dissimilarity between the pattern statistics of the conditioning data and the simulation image in two steps: the pattern statistics are first obtained using a smooth histogram method; then, the difference between the histograms is evaluated by computing the Jensen–Shanon divergence. The proposed approach is applied for the simulation of the geological interface (footwall contact) within a laterite-type bauxite mine deposit using the Direct Sampling MPS algorithm. The results point out two main advantages: (1) a faster parameter tuning process and (2) more objective determination of the parameters.

**5.1 Introduction**

Multiple-point statistics (MPS) allows the simulation of spatial or temporal random functions by reproducing pattern statistics from an exhaustive data set — the training image (TI) — built from conceptual knowledge (Strebelle, 2002) or borrowed from analogue sites (Oriani et al., 2014; Pirot et al., 2014). Due to the capabilities and the power of MPS, its use has been rather widespread for simulating complex structures in earth sciences. This fact paved the way for the development of a large number of MPS algorithms and their application to various contexts (Guardiano and Srivastava, 1992; Strebelle, 2002; Zhang et al., 2006;

Arpat and Caers, 2007; Gloaguen and Dimitrakopoulos, 2009; Dimitrakopoulos et al., 2010; Honarkhah and Caers, 2010; Mariethoz et al., 2010; Straubhaar et al., 2011; Tahmasebi et al., 2012).

In this paper, we focus on a novel application of MPS: geological modelling of lateritic bauxite deposits. The objective is to simulate the footwall contact topography, which constitutes the base of the exploitable deposit. The reason why MPS could be interesting in this particular case is that several sites have already been mined out, and the topographies of the footwall contact exposed after the mining operations can be considered as analogues to the footwall contact of a future mining area. In addition, the data sets collected from such mines usually include several variables such as exploration boreholes, exhaustive geophysical data, and production control boreholes (Erten, 2012; Dagasan et al., 2018a,b). In order to account for the multivariate nature of the modelling problem and the necessity to simulate a continuous variable, it was decided to use the Direct Sampling (DS) algorithm (Mariethoz et al., 2010) in this work.

As in all MPS algorithms (Mariethoz and Caers, 2014), DS requires some input data such as the training image (TI) and a set of specific input parameters. The choices of the TI and the input parameters significantly influence the simulation results (Boucher, 2007). Assuming that the selection of a TI has already been made, identifying a correct set of input parameters plays a direct role on the quality of the simulations and the spatial uncertainty quantification. One of the most common approaches to determine the MPS parameters is to manually tune them by conducting a sensitivity analysis (Mariethoz et al., 2010; Boisvert et al., 2010; Mariethoz and Caers, 2014). Computation time and simulation quality are the key performance indicators used in the sensitivity analysis (Liu, 2006). Because these two indicators often counteract each other, the optimality of a set of input parameters is usually assessed based on the balance achieved between them. Applications of such sensitivity analyses can be found in Meerschman et al. (2013); Rezaee et al. (2013); Liu (2006); Huysmans and Dassargues (2011).

More generally, the quality of the simulations is often difficult to assess since many different criteria can be used and the parameter identification is based on a

manual procedure and a visual inspection of the results. More rigorous methods define a set of indicators to quantify the similarity between the input data and the simulations. Quantification of the similarity between the TI, conditioning data and the resulting simulations is usually based on the reproduction of first and second order statistics. For instance, Meerschman et al. (2013) quantify the error between the TI and model statistics based on the calculated errors in connectivity, histograms and variograms. Quantifying the multiple-point statistics error, on the other hand, can be done by multiple-point histograms (Boisvert et al., 2010; De Iaco and Maggio, 2011; Tan et al., 2014), cumulants (Dimitrakopoulos et al., 2010), connectivity functions (Renard and Allard, 2013), spatial patterns (Pérez et al., 2014), smooth histograms (Melnikova et al., 2015), or connectivity indicators (Pirrot et al., 2014). However, carrying out such a sensitivity analysis manually requires running many simulations and is cumbersome and time-consuming.

The specific aim of the present paper is to present a new automated technique to determine the appropriate input parameters of the MPS algorithm for the case of bauxite deposits. Some of the techniques and ideas that are developed in this context can easily be re-used and extended for more general situations (stratified deposits), as we will discuss at the end of the paper. The benefits of an automated method are twofold: it leads to higher quality simulations through the enhancement of pattern reproduction; and it provides less labour intensive and more objective parameter tuning. The approach presented in this paper makes use of a stochastic optimisation framework to automate the parameter tuning process. Input parameters are utilised as the decision variables to minimise an objective function which quantifies the mismatch between the pattern statistics of the conditioning data and the generated realisations. Computing the objective function is performed in two steps: First, the smooth histogram formulation of Melnikova et al. (2015) is used to quantify the pattern statistics of the boreholes and that of the simulations. We have selected that approach since it allows working with continuous variables, while other simpler techniques would not allow computing histograms of patterns of continuous variables. Second, the dissimilarity between pattern histograms is evaluated utilising the Jensen-Shannon divergence



(JS) (Cover and Thomas, 2012). In order to observe the effect of the optimisation algorithm on the tuned parameters, a number of optimisation techniques have been utilised to minimise the objective function. Nevertheless, the Simulated Annealing (SA) algorithm (Kirkpatrick et al., 1983) has proven to be the most efficient method amongst the ones tested, as the objective function may contain many local minima. After hundreds of iterations, the SA algorithm converges and provides the parameters yielding the minimum JS divergence value.

The remainder of the paper is organised as follows: Section 5.2 reviews the underlying methods used to develop the proposed approach, Section 5.3 presents the setup of the problem using the data from a bauxite mine and implementation of the approach, Section 5.4 is dedicated to the analyses of the results and Section 5.5 concludes the key features of the automated parameter tuning process presented.

## 5.2 Overview of Underlying Methods

In this section, the methods that are used in the proposed methodology are reviewed. The aim is to provide the required background information to understand the implementation of the methodology described in Section 5.3.

### 5.2.1 Direct Sampling Algorithm

Direct Sampling is a pixel-based MPS algorithm which is used to simulate a random function  $Z$  (Mariethoz et al., 2010). Being a sequential simulation algorithm, it successively visits all the locations  $x$  of a regular simulation grid (SG) and generates simulated  $Z^*(x)$  values, until all the grid nodes are informed. Once all the conditioning data (if available) are assigned in the SG, the algorithm follows a predefined random path to visit all the non-informed grid nodes  $x$  to perform the simulations. Having chosen the  $n$  maximum number of closest grid nodes parameter of the DS, the algorithm finds  $n$  number of informed neighbours at grid nodes  $\{x_1, x_2, \dots, x_n\}$  around  $x$  and computes the lag vectors  $L = \{h_1, h_2, \dots, h_n\}$  to define the data event  $d_n(x, L) = \{Z(x + h_1), \dots, Z(x + h_n)\}$ . The TI is randomly scanned at  $y$  locations until the distance between the patterns

$d_n(x, L)$  and  $d_n(y, L) = \{Z(y+h_1), \dots, Z(y+h_n)\}$  falls below a predefined threshold  $t$  or until the maximal scan fraction  $f$  of the TI is reached. Then, the value  $Z(y)$  at the scanned node  $y$  is taken as the best match and is pasted on the grid node  $x$  of the SG. These steps are repeated until all the nodes in the SG are simulated.

There are different ways of computing the distance between two data events (or patterns) in the DS algorithm and the choice depends mainly on the type of the variable. For instance, the distance for continuous variables can be calculated by using the Manhattan distance:

$$d\{d_n(x, L), d_n(y, L)\} = \frac{1}{n} \sum_{i=1}^n \frac{|Z(x_i) - Z(y_i)|}{d_{\max}} \in [0, 1] \quad (5.1)$$

where

$$d_{\max} = \max_{y \in TI} Z(y) - \min_{y \in TI} Z(y) \quad (5.2)$$

When calculating the distances between the patterns retrieved from the SG and the TI, different weightings can be given to the data event nodes based on their distances  $\|h_i\|$  to the central node. This is carried out by a weighting factor  $\alpha_i$  applied to each data event node such that:

$$\alpha_i = \frac{\|h_i\|^{-\delta}}{d_{\max} \sum_{j=1}^n \|h_j\|^{-\delta}} \quad (5.3)$$

where  $\delta$  is the power for computing the weight. A specific weighting  $w$  can also be used to achieve pattern consistency in the neighbourhood of the conditioning data. The weighting factor  $\alpha_i$  then becomes:

$$\alpha_i = \frac{\beta_i \|h_i\|^{-\delta}}{d_{\max} \sum_{j=1}^n \beta_j \|h_j\|^{-\delta}} \quad (5.4)$$

where

$$\beta_i = \begin{cases} w & \text{if the } i^{\text{th}} \text{ node (i.e., } x_i) \text{ is a conditioning location (hard data)} \\ 1 & \text{otherwise} \end{cases}$$

The DS algorithm also allows multivariate simulations utilising the multiple-point relationship between the variables in the training data set. The training image TI used in multivariate simulations is comprised of  $m$  variables. Considering the variables  $Z_1(x), \dots, Z_m(x)$  ( $k = 1, \dots, m$ ) used in the multivariate analysis, each variable  $Z_k(x)$  may have a different data event  $d_{n_k}^k(x, L) = \{Z_k(x + h_1^k), \dots, Z_k(x + h_{n_k}^k)\}$  with different  $L_k$  lag vectors. The joint data event  $d_n(x) = \{d_{n_1}^1(x, L^1), \dots, d_{n_m}^m(x, L^m)\}$  is then used to scan the TI to find a compatible pattern.

In the original version of the DS (Mariethoz et al., 2010), the algorithm combines all the specific distances using a weighted average. However, in this work, we use a different implementation of the DS called DeeSse (Straubhaar, 2016). Instead of using a global threshold and a set of weights, DeeSse uses a variable specific threshold  $t_k$  for each variable  $Z_k$ . The TI is scanned until the distance for each pattern is below the corresponding threshold, or the maximum scan fraction is reached (Straubhaar, 2016; Straubhaar et al., 2016). More detailed information on the implementation and application of the DS algorithm can be found in Mariethoz et al. (2010) and Meerschman et al. (2013).

## 5.2.2 Comparing Patterns with Smooth Histograms

The smooth histogram is a pseudo-histogram reflecting the pattern statistics of an image (Melnikova et al., 2015). Since it is based on the pixel values rather than the pattern counts, it can be computed for both discrete and continuous images. Melnikova et al. (2015) use such a histogram to compare the multiple-point statistics of a continuous model with a discrete training image for an inverse problem. The comparison is carried out by first defining a search template  $T$  to collect the unique patterns  $pat_j^{TI,un}$  of the categorical training image. These patterns are then used to construct the categories (classes) of the pseudo-histograms  $H^{d,m}$

and  $H^{d, TI}$  of the model image and the TI, respectively. Because the categories of the constructed pseudo-histograms are discrete, any pattern  $pat_i^m$  observed in the continuous model image  $m$  would not fall into one of these categories. Instead, it contributes to all  $N^{TI, un}$  numbers of unique categories with a value between 0 and 1. This value is calculated based on the similarity of the patterns in the histogram categories and the patterns of the model image. The  $j$ th bin of the model histogram  $H^{d, m}$  is then computed by:

$$H_j^{d, m} = \sum_{i=1}^{N^m} \frac{1}{(1 + A \left\| pat_i^m - pat_j^{TI, un} \right\|_2^k)^s} \quad (5.5)$$

Similarly, the pseudo-histogram of the TI can be computed by calculating the patterns of the categories and all the patterns in the TI, as in the following:

$$H_j^{d, TI} = \sum_{i=1}^{N^{TI}} \frac{1}{(1 + A \left\| pat_i^{TI} - pat_j^{TI, un} \right\|_2^k)^s} \quad (5.6)$$

where  $A$ ,  $k$  and  $s$  are user-defined parameters to shape the pattern similarity function. Melnikova et al. (2015) state that these parameters not only define how well the pseudo-histogram approximates the true frequency distribution, but they also control the degree of smoothing. However, there is a trade-off between the degree of smoothing achieved and the true frequency distribution approximated. Therefore, optimal values balancing these are required. Melnikova et al. (2015) define  $A = 100$ ,  $k = 2$  and  $s = 2$  as optimal parameters balancing this trade-off.

The comparison of two pseudo-histograms requires a dissimilarity metric. In our case, we chose the Jensen–Shannon (JS) divergence to calculate the dissimilarity of two pseudo-histograms. The JS is used to quantify the dissimilarity between two density distributions  $p$  and  $q$  by averaging the Kullback–Leibler divergences, as in the following (Endres and Schindelin, 2003):

$$d_{JS}(p, q) = \frac{1}{2} \sum_i p_i \log \left( \frac{p_i}{q_i} \right) + \frac{1}{2} \sum_i q_i \log \left( \frac{q_i}{p_i} \right) \quad (5.7)$$

where  $p_i$  and  $q_i$  represent the probability densities at the  $i$ th bins. Calculation of the dissimilarity also requires the pseudo-histograms to share the same (mutual) classes.

### 5.2.3 Generalised Simulated Annealing

Simulated Annealing (SA) is a stochastic optimisation tool to solve complex optimisation problems by mimicking the annealing process of a molten metal (Reeves, 1993; Xiang et al., 2017). Being inspired by the Metropolis algorithm (Metropolis et al., 1953), it was first proposed by Kirkpatrick et al. (1983) to find the global optimum of a complex objective function. The method approximates the global minimum by exploring the solution space at finite locations. The combination of stochastic exploration and a cooling scheme controls the probability of accepting worse solutions and avoiding remaining trapped in local minima. The artificial temperature is high in the initial stages of the optimisation, therefore, worse solutions have a higher probability of acceptance. As the optimisation progresses, the temperature is lowered and the focus is shifted toward accepting only better solutions to identify the minimum more accurately.

Given an objective function  $f(x)$  such that  $x = (x_1, x_2, \dots, x_n)$ , a standard SA algorithm uses the following steps to find the global minimum (Sun and Sun, 2015):

1. Choose a high initial temperature  $T_0$  value, an initial solution  $x^0$  and evaluate the objective function,  $E^0 = f(x^0)$ .
2. Propose a new solution  $x^{i+1}$ :
  - Generate a candidate solution  $x^{i+1}$  from the current one ( $x^i$ ) using a predefined visiting distribution.
  - Evaluate the change in the objective function for the candidate solution,  $\Delta = f(x^{i+1}) - f(x^i)$ .
  - Accept the iteration if it reduces the objective function,  $\Delta < 0$ .
  - Otherwise, accept or reject it based on a probability of acceptance criterion.

3. Repeat step 2 for  $L$  number of iteration times keeping  $T$  constant.
4. Reduce the temperature to  $T_{n+1}$  using a cooling function.
5. Repeat steps 2–4 until the convergence criteria is satisfied.

The choice of the visiting distribution in step 2 has a significant effect on the efficiency of the SA algorithm. Therefore, different visiting distributions for the SA have been investigated by some researchers such as Szu and Hartley (1987) and Tsallis and Stariolo (1996). In this research, the Generalised Simulated Annealing algorithm (GSA) (Tsallis and Stariolo, 1996) with the distorted Cauchy–Lorentz distribution is used. As for the stopping criteria, the GSA offers different options such as maximum running time, maximum function calls, maximum iteration number or a threshold value for the objective function.

### 5.3 Problem Setup and Methodology

Before describing the details of the proposed methodology, let us present the problem more precisely. The ultimate aim is to determine the MPS parameters to simulate the interface between the ore and waste within a laterite-type bauxite deposit.

The data to perform the simulations comes from two mine sites. The first one (Figure 5.1a,b) comprises a finite set of depth measurements collected from regularly spaced boreholes (conditioning data and primary variable) and an exhaustive Ground Penetrating Radar (GPR) survey (soft data and secondary variable). The position of the interface in this mine site is aimed to be simulated by DS, using these data sets as conditioning information. The second site is a mined-out analogue area providing a training data set. It includes the topographical survey of the exposed deposit base (primary variable) as well as a GPR survey done before the mining operation. This training data set, which is used as a bivariate training image, can be seen in Figure 5.1. In addition to this data, the simulation of the interface with DS requires the input parameters (number of neighbours, distance

thresholds, etc.) to be determined. The aim of the proposed methodology is, therefore, to automatically obtain an optimal set of parameters for that purpose.

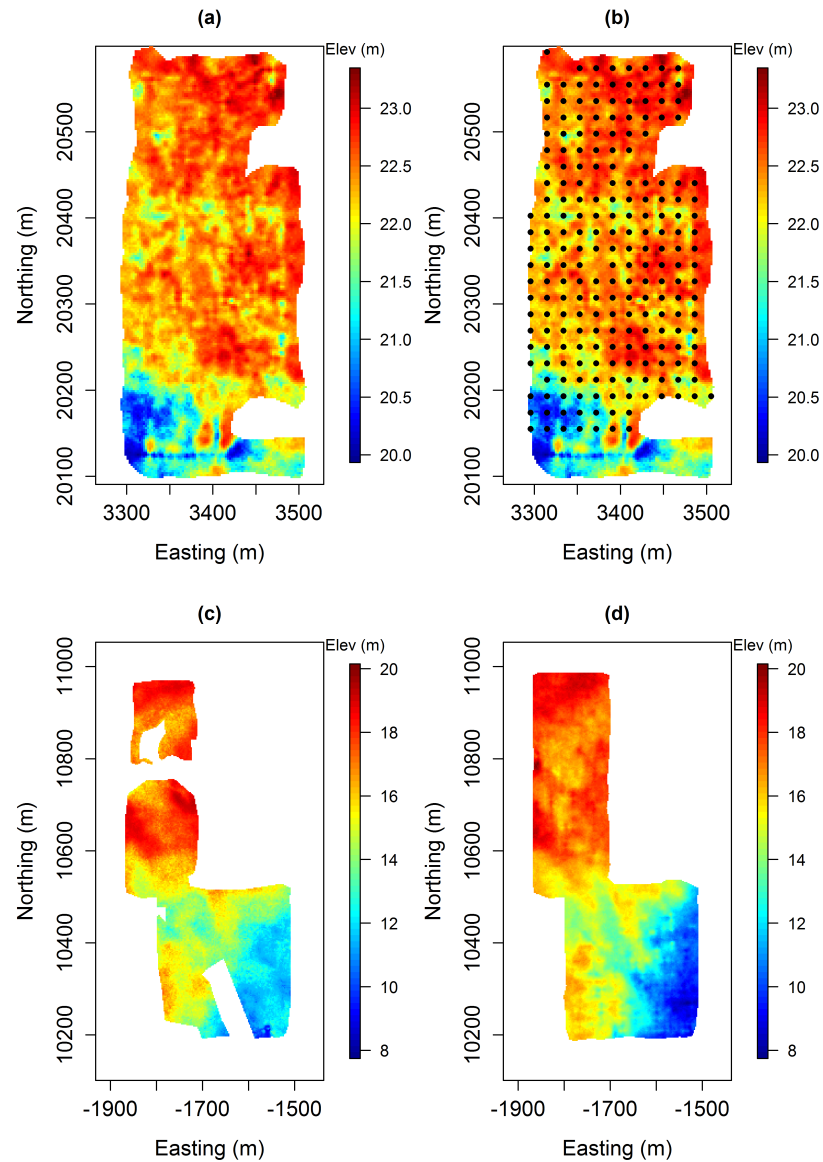


Figure 5.1: Simulation and the TI variables: (a) GPR variable used as the secondary simulation variable; (b) borehole data used to condition the simulations (primary variable-black dots overlying the GPR survey); (c) floor survey of a previously extracted mine area (primary TI variable) and (d) GPR data of the previously mined area collected prior to the extraction (Secondary TI variable).

In lateritic bauxite deposits, the data sets may be dense and regularly organised. This offers the possibility to define a specific quality metric in which the patterns observed in the conditioning data can be taken as the target. Therefore, the general principle of the proposed methodology is to minimise an objective function representing the dissimilarity of the conditioning data with the simulated image in terms of pattern statistics. The calculation of the dissimilarity requires the pattern statistics to be computed. This is only done once for the conditioning data. For the simulated image, however, the pattern statistics change with each perturbation of the parameters and needs to be updated.

### 5.3.1 Methodology

The computation of the conditioning data pattern statistics first requires the migration of the borehole datapoints into the SG. This step is needed to convert the punctual data into the gridded type. Once this step is done, a search template  $T$  is defined to retrieve the patterns contained in the borehole data. Since we are comparing the pseudo-histograms of one fully informed simulation image and a partially-informed borehole data image, a special search template needs to be constructed. Therefore, the construction of the search template  $T$  is based on two parameters: (1) the number of grids between borehole data and (2) the number of borehole data desired to be used in one category of the pseudo-histogram. This is illustrated in the following example: say we have borehole data located in a grid at every five nodes in the x and y directions. If we want to capture a four boreholes data pattern, we choose the size of the search template to be  $5 \times 5$ . The constructed search template scans the borehole data grid and collects the patterns once four borehole data is captured by the search template. The collected patterns serve as the categories of the pseudo-histogram to be constructed. In order to account for the non-stationarity, the means  $\overline{pat_i^C}$  are subtracted from each  $pat_i^C$ , focusing on the variation of each pattern around its mean. The steps mentioned above are illustrated in Figure 5.2.



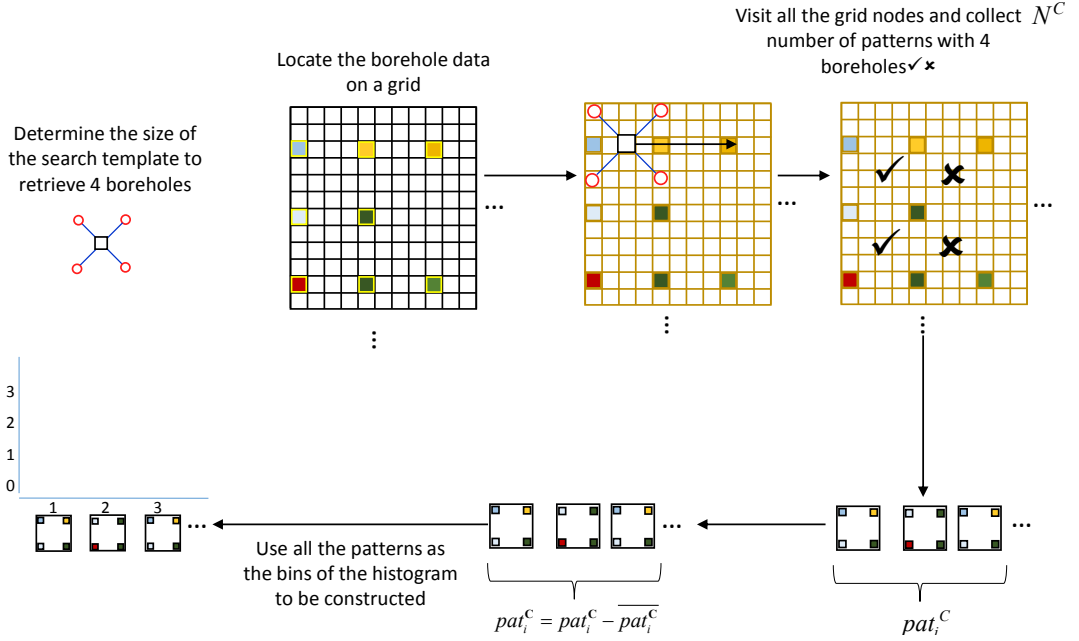


Figure 5.2: Construction of the smooth histogram categories of the of borehole data.

Following the construction of the smooth histogram categories for the conditioning data, the pseudo-histogram  $H^{d,C}$  for the conditioning data is computed by the following:

$$H_j^{d,C} = \sum_{i=1}^{N^C} \frac{1}{(1 + A \|pat_i^C - pat_j^C\|_2^k)^s} \quad (5.8)$$

where  $N^C$  represents the number of patterns captured from the SG.

Once the smooth histogram of the borehole data is generated, the SA algorithm runs the DS algorithm to produce a realisation. The smooth histogram of this realisation is then computed using the categories created for the conditioning data histogram. In order to calculate the contribution of each pattern  $pat_i^{SIM}$  in the histogram categories  $H_j^{d,SIM}$ , the means  $\overline{pat_i^{SIM}}$  are again subtracted from each  $pat_i^{SIM}$ . The pseudo-histogram  $H^{d,SIM}$  of a realisation is calculated as follows:

$$H_j^{d,SIM} = \sum_{i=1}^{N^{SIM}} \frac{1}{(1 + A \|pat_i^{SIM} - pat_j^C\|_2^k)^s} \quad (5.9)$$

where  $N^{SIM}$  is the number of patterns captured in the realisation image generated. Because there are more patterns in the simulation image than the conditioning data image, the weights in each category of the pseudo-histograms of the realisations are expected to be higher than those of the conditioning pseudo-histogram. In order to compare the pseudo-histograms, they are normalised before the computation of the Jensen–Shannon divergence. The normalisation is carried out through the following:

$$H_j^{d_{Nor}} = \frac{H_j^d - 1}{\sum_{i=1}^{NC} (H_i^d - 1)} \quad (5.10)$$

where  $H_j^{d_{Nor}}$  is the normalised weight corresponding to the weight of the  $j^{th}$  category  $H_j^d$  of either  $H_j^{d_{SIM}}$  or  $H_j^{d_{C}}$ . Subtraction of 1 from each individual pseudo-histogram category has been done to remove the weight contribution of each pattern  $H_j^{d_{C}}$  in its own  $j^{th}$  category. The Jensen-Shannon divergence is then calculated as follows:

$$\begin{aligned} O = d_{JS}(H_j^{d_{CNor}}, H_j^{d_{SIMNor}}) &= \frac{1}{2} \sum_j H_j^{d_{CNor}} \log \left( \frac{H_j^{d_{CNor}}}{H_j^{d_{SIMNor}}} \right) \\ &+ \frac{1}{2} \sum_j H_j^{d_{SIMNor}} \log \left( \frac{H_j^{d_{SIMNor}}}{H_j^{d_{CNor}}} \right) \end{aligned} \quad (5.11)$$

Starting from an initial set of  $x^0$  DeeSse parameters (vector of decision variables), the SA algorithm runs the DS algorithm to produce a realisation and evaluate the objective function  $O_1$ . Then, the parameters are iteratively and randomly perturbed using the Cauchy-Lorentz visiting distribution. A new value of the objective function  $O_{i+1}$  is computed and the new set of parameters  $x_{i+1}$  is accepted or rejected according to a probability acquired from the generalised Metropolis algorithm. Factors influencing the probability of acceptance comprise the previous and candidate objective function values ( $O_i, O_{i+1}$ ) as well as the number of iterations that have been made so far. During the initial iterations, the SA has a relatively high probability of accepting changes that may worsen the objective function. In

order to explore the function that we want to minimise later in the process, this probability is reduced allowing the identification of the optimum better. The SA algorithm runs until the convergence is achieved and the objective function  $O$  is minimised. The overall flowchart of the methodology is shown in Figure 5.3.

The methodology was mainly implemented using R statistics software (Team, 2017). The MPS simulations were performed by calling the DeeSse algorithm (which is coded in C) from R. For the optimisation part, we implemented the GSA using the GenSA package of R software (Xiang et al., 2013).

## 5.4 Results and Discussion

### 5.4.1 Implementation and Analysis of the Tuned Parameters on the Simulations

As explained in Section 5.3, the aim is to model the lateral variability of the ore/waste contact surface in a lateritic bauxite deposit using borehole and GPR datasets (Figure 5.1). For that purpose, we use a training dataset (bivariate TI) coming from an analogue site (Figure 5.1c,d). The simulation grid has a size of 97 and 214 nodes in the  $x$  and  $y$  directions, respectively. The TI, on the other hand, is  $180 \times 400$  in grid size. The spacing between the nodes of both the SG and the TI grids are 2.38 m. Since the data sets have apparent drifts, the types and the coefficients of these drifts were initially detected. First order trend surfaces for the TI and the second order trend surfaces for the simulation variables were found suitable and subtracted from the data sets to obtain the residuals. The residuals obtained were then used to perform the MPS simulations. Following the simulations, the trend surfaces were added back to each realisation.

Several combinations of DS parameters were initially tested to determine the appropriate simulation parameters by visually checking the simulation qualities. These initial attempts have shown that the simulation qualities were rather sensitive to changes in the parameter, making it difficult to identify optimal parameters through visually analysing the resulting simulations. The proposed methodology

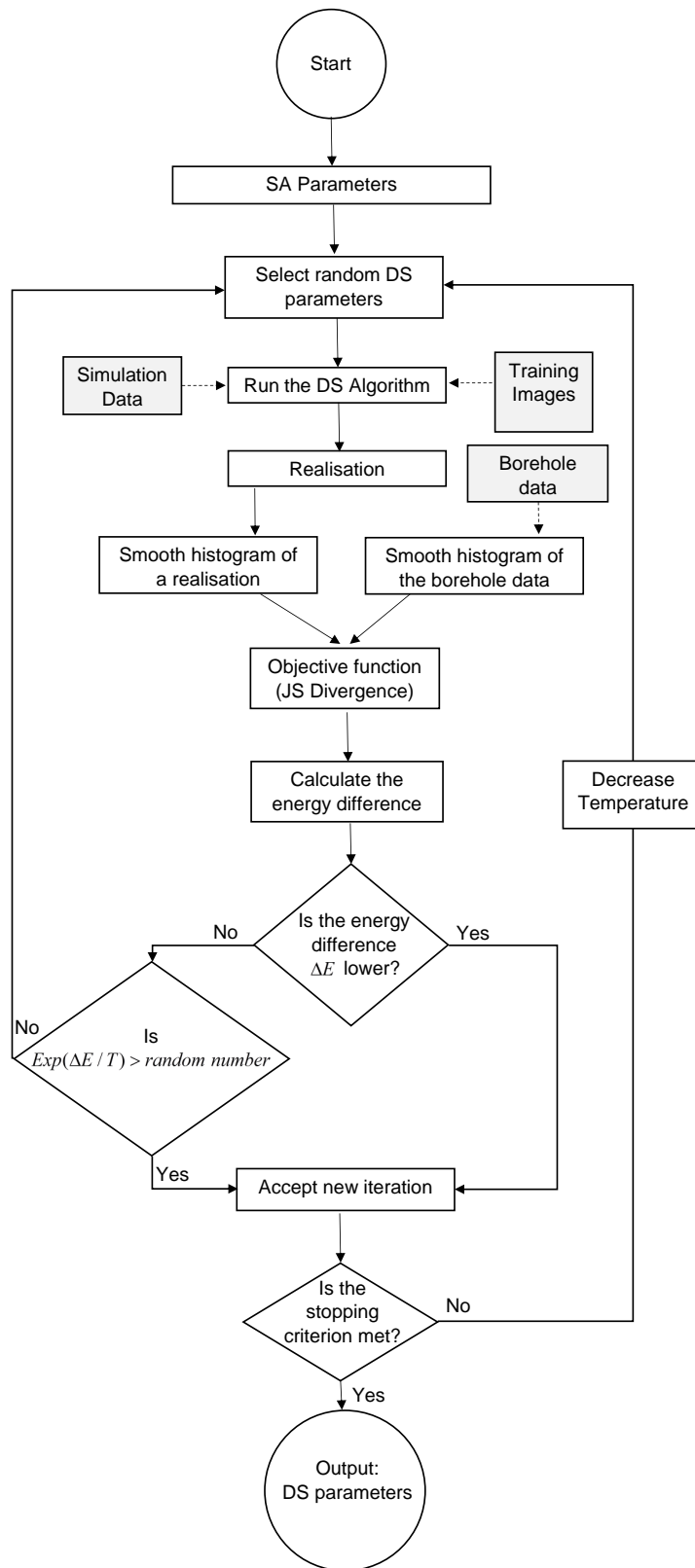


Figure 5.3: Flowchart of the methodology.

was then implemented. Five parameters that were considered during the optimisation procedure were  $w_{hd}$ ,  $n_{hd}$ ,  $n_{GPR}$ ,  $t_{hd}$  and  $t_{GPR}$ . The weighting factor  $w_{GPR}$  of the GPR data was not selected for the optimisation, as the simulation grid  $SG_{GPR}$  of the GPR data was already exhaustively informed and any change in  $w_{GPR}$  would not affect the simulations. One of the main parameters of the DS algorithm, the scan fraction  $f$ , was also not considered in the procedure. Based on our previous trials, it did not affect the simulation quality significantly for this example. This observation of ours is rather consistent with the results presented in Meerschman et al. (2013); they state that the parameter  $f$  has a small effect on the simulation qualities unless it is below 0.2 for continuous simulations. Therefore, the scan fraction  $f$  was kept equal to 0.5 throughout the optimisation procedure.

As the first step of our proposed algorithm, a search template is built to capture 4 points to construct the pseudo-histogram. Because the boreholes are drilled on a regular grid of  $19.05 \times 19.05$  metres and the grid spacing is 2.38 m, they are located in the SG at every 9<sup>th</sup> node in the  $x$  and  $y$  directions. Therefore, a  $9 \times 9$  search template was constructed to capture four borehole data at a time. This search template was then used to construct the pseudo-histogram  $H^{d,C}$  of the borehole patterns.

Having constructed the pseudo-histogram of the conditioning data, the SA algorithm was run with the following initial parameters:  $w_{hd} = 10$ ,  $n_{hd} = 5$ ,  $n_{GPR} = 20$ ,  $t_{hd} = 0.01$  and  $t_{GPR} = 0.1$ . These parameters were selected as the ones yielding visually good simulation quality during the manual sensitivity analyses. In addition, the optimisation is also constrained by an upper and lower boundary, as can be seen in Table 5.1. The resulting realisation is shown in Figure 5.4a.

The pseudo-histogram  $H^{d,SIM}$  of the initial realisation is then calculated to compute the  $d_{JS}$  mismatch. The JS divergence  $d_{JS}$  between the pseudo-histograms  $H^{d,C}$  and  $H^{d,SIM}$  was initially calculated as 0.102. Using 30,000 function calls as the stopping criteria, the SA algorithm converged and found the parameters for the DS algorithm yielding a local minimum divergence value. The resulting realisation can be seen in Figure 5.4b. The parameters along with their associated JS divergence values are provided in Table 5.1.

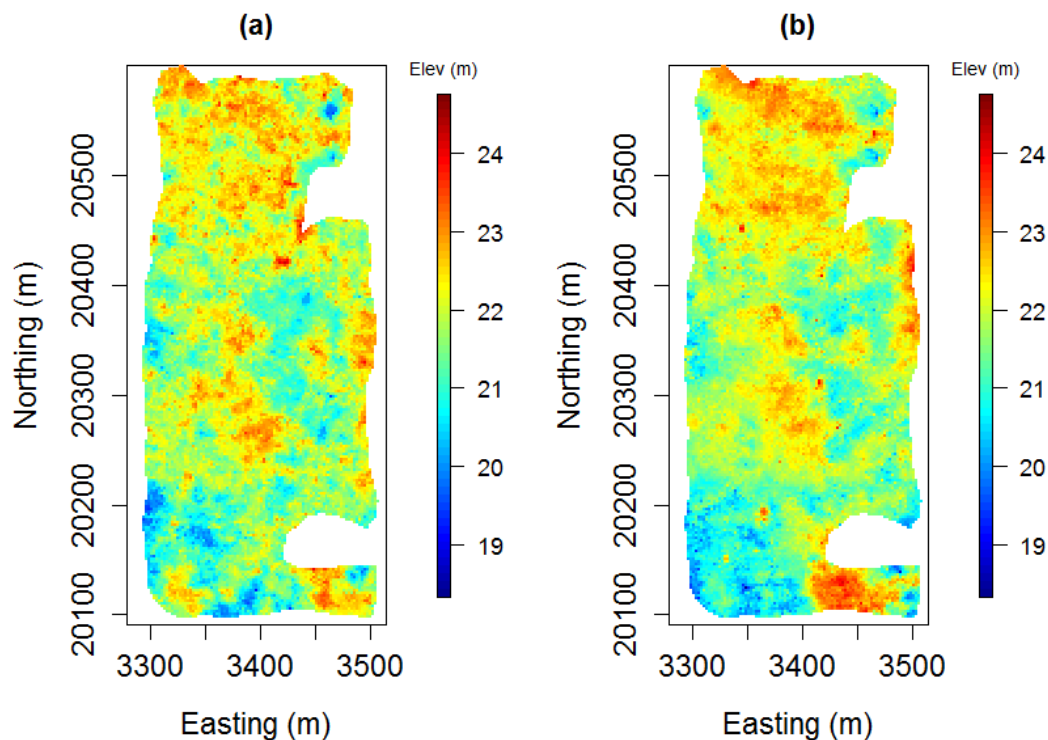


Figure 5.4: Realisations produced with (a) the initial parameters and (b) final parameters.

Table 5.1: Initial and optimum parameters together with the associated JS divergence values.

Parameters	$w_{hd}$	$n_{hd}$	$n_{GPR}$	$t_{hd}$	$t_{GPR}$	$d_{JS}$
Lower Boundaries	1	1	1	0.0001	0.0001	0.1520
Initial Parameters	10	5	20	0.010	0.10	0.1020
Automatic Tuning	13.07	14	45	0.009	0.305	0.0667
Upper Boundaries	15	50	50	0.5	0.5	0.1793

Although the initial and final realisations exhibit a noticeable difference, the average of 40 realisations generated using the initial and optimum parameters do not show an apparent dissimilarity, as can be seen in Figure 5.5. However, the interquartile ranges of the elevation values calculated at each grid node noticeably dropped after the parameter tuning procedure, as illustrated in Figure 5.5.

This can be seen more clearly in the distributions of IQR values shown in Figure

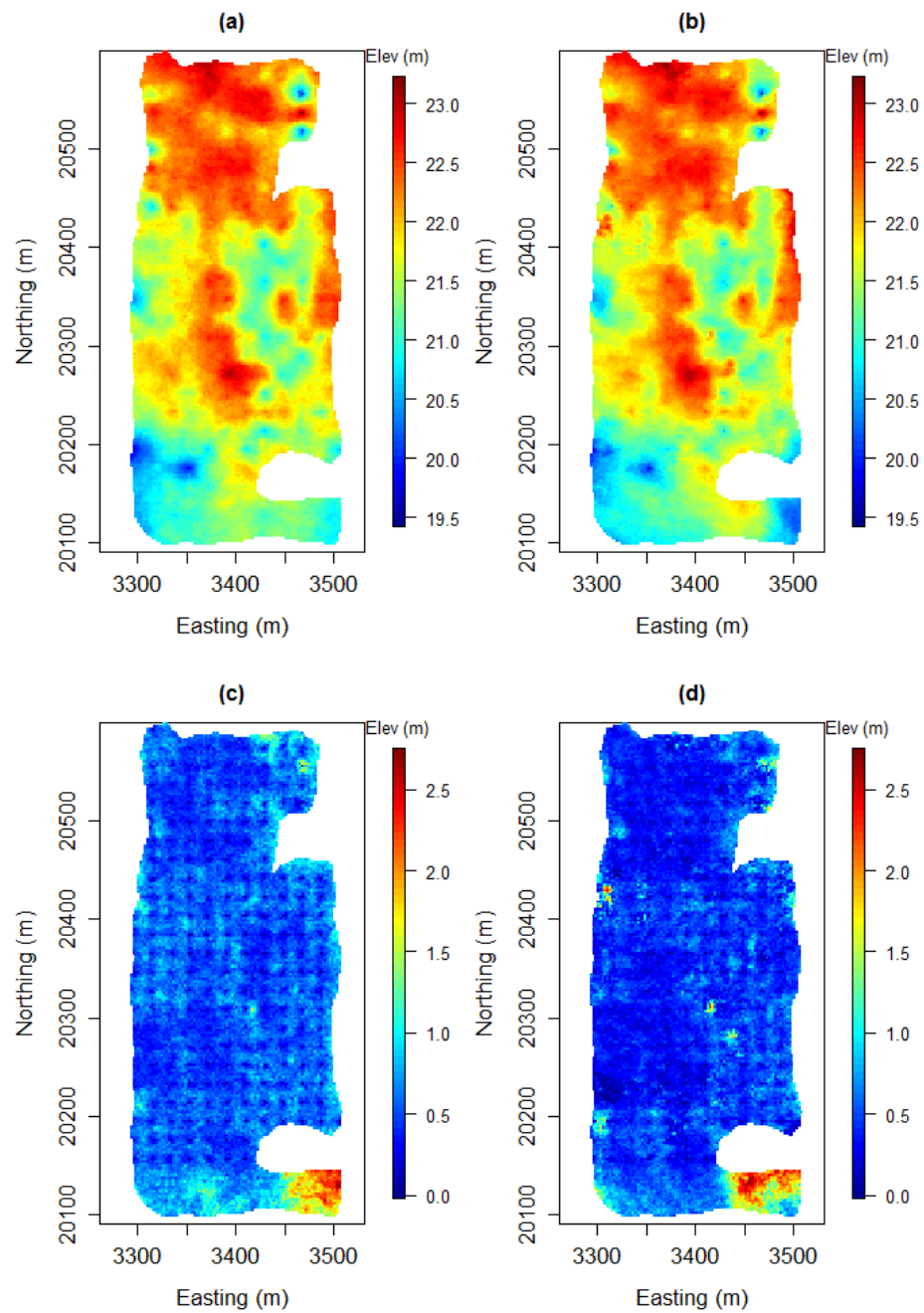


Figure 5.5: (a) Initial average of the realisations; (b) final average of the realisations. Interquartile range maps: (c) before the parameter optimisation and (d) after the parameter optimisation.

5.6. Similarly, the evolution of the variograms in Figure 5.7 also shows that the variability between the realisations has decreased. In other words, the statistical fluctuations are better centred around the experimental variogram of the borehole data. These results indicate that although the means of the realisations exhibit a considerable similarity, there is a reduction in the estimated uncertainty around the mean. Since the optimised realisations are richer in terms of conditioning data patterns, the model of uncertainty would represent the original data variability better. Therefore, we consider the resulting uncertainty reduction as an improvement.

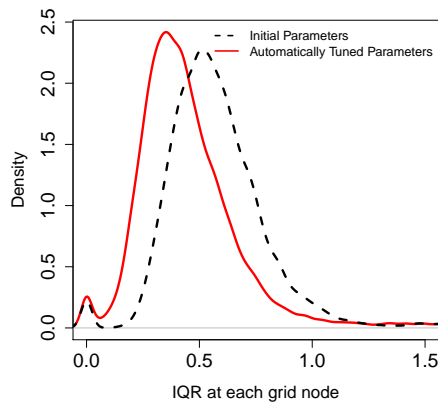


Figure 5.6: Evolution of the distributions of the IQR values calculated at each grid node. Mean of the IQR values dropped from 0.60 to 0.47 in automatically found parameters.

#### 5.4.2 Effect of the Optimisation Method and the Initial Parameters on the Final Results

In order to test the choice of the SA, we compared the performances of different optimisation techniques. The parameter tuning process has also been carried out using other optimisation methods. These optimisers include the L-BFGS (quasi-newton type) (Zhu et al., 1997) and BOBYQA (trust region) (Powell, 2009) methods. The optimisers were set to run with different initial parameters. These included a set of input parameters yielding a high JS divergence value, suggested



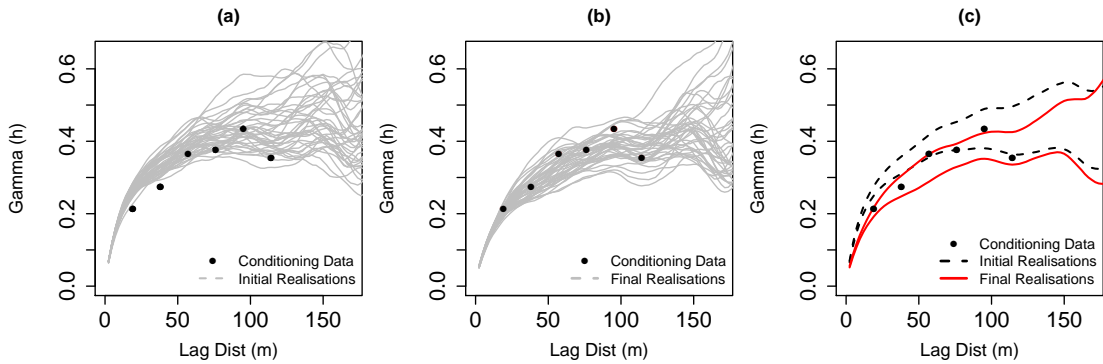


Figure 5.7: Variograms of the realisations and the experimental variogram of the borehole data: (a) Variograms before the optimisation; (b) variograms after the optimisation and (c) 99.9% confidence interval of the variograms before and after the optimisation.

parameters in Meerschman et al. (2013) and our initial parameters based on the manual sensitivity analysis.

The results in Figure 5.8 indicate that the final value of the JS divergence is highly dependent on the initial parameters chosen for the BOBYQA and LBFGS optimisers. The performance of the SA algorithm, on the other hand, is less sensitive to the choice of initial parameters for these cases. Therefore, it is more robust and requires less preliminary manual sensitivity analysis. As discussed earlier, such a manual analysis is rather time consuming as it involves testing different set of parameters and observing the results manually. Furthermore, in three out of four cases, the SA outperformed the other optimisation algorithms in terms of reaching the minimum JS divergence.

It should be noted that the optimizers used in this paper are reaching an optimum solution by utilising an iterative process. Hence, as the number of iterations increases, the probability of achieving a better solution increases as well. However, this would increase the total CPU time required for the optimisation. In the case study presented, 5 DS parameters were tuned to simulate the elevation values of the ore/waste interface. The simulation of 20,758 grid nodes took, on average, 4 s to perform one run using 8 threads. Using 30,000 simulations in the optimisation process, therefore, takes roughly 33.3 h to complete. If the method is applied for

a problem with larger grid size, CPU time may become an important concern. Therefore, the number of iterations should be adjusted depending on the nature and the complexity of the simulations. Based on our experience obtained from this study, even 100 function calls with the SA seem to be adequate to find a better set of parameters than the visual inspection method. This can also be seen in the convergence plots of the SA for 30,000 function calls (Figure 5.9). Taking the total reduction in 30,000 function calls as a reference (from 0.180 to 0.067), the graphs show that 75% of the reduction that is achieved actually takes place within 100 function calls. Whereas, 99% is achieved in 7000 function calls. Therefore, if feasible, the optimisation can be set to run around seven to eight thousand function calls. Otherwise, a hundred function calls would also yield satisfactory results.

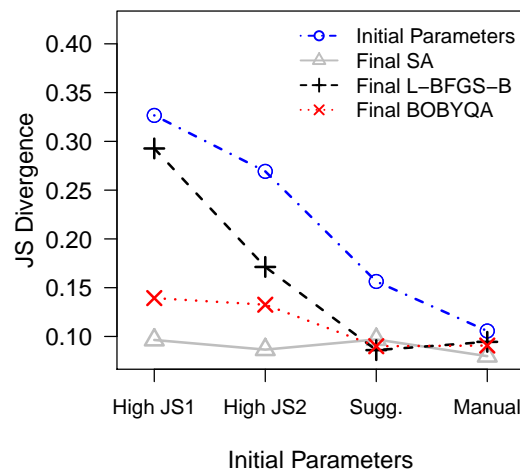


Figure 5.8: Convergence comparisons of three different optimisers with three different set of initial DS parameters.

### 5.4.3 Effect of the Chosen Parameters on the Pattern Reproduction

Since the JS divergence is a function of five DS parameters, the direct visualisation of its behaviour depending on the parameters is not straightforward. An

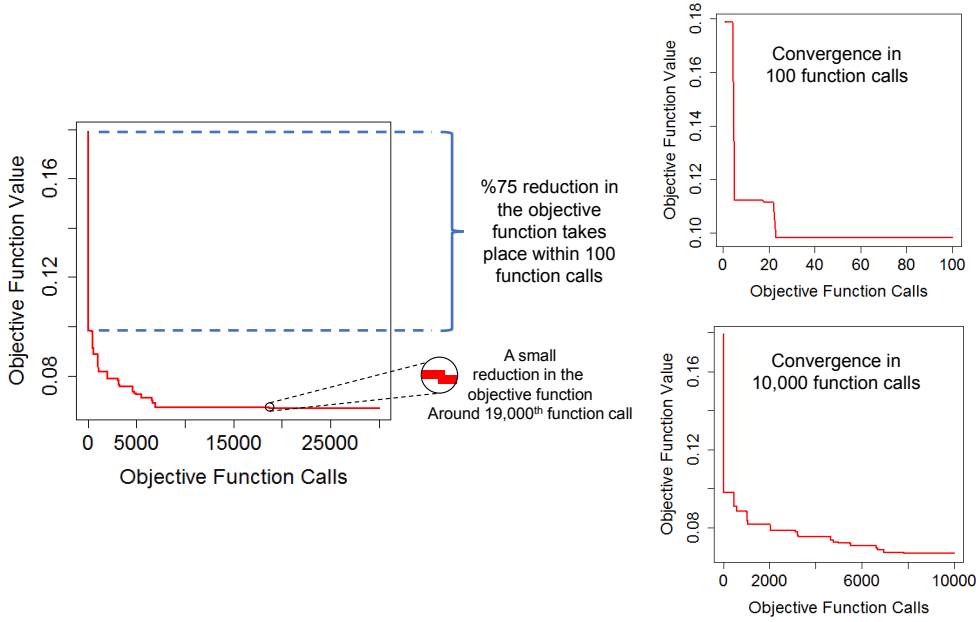


Figure 5.9: Convergence graphs obtained by running 30,000 function calls.

approach to visualise such multi-dimensional problems is to plot the lowest JS divergence values in 1-D against individual parameters. This technique is called profile-likelihood and is utilised to check visually the parameters of a statistical model obtained via maximum likelihood techniques. The idea is to plot the maximum likelihood function values as a function of all but one parameter. It can also be considered analogous to tracking the maximum values along the crests of a five dimensional function. Given a model with parameters  $(\alpha, \psi)$  and the likelihood  $L(\alpha, \psi)$ , the profile of the likelihood  $L_p(\alpha)$  for parameter  $\alpha$  can be denoted as follows (Diggle et al., 1998):

$$L_p(\alpha) = L(\alpha, \hat{\psi}(\alpha)) = \max_{\psi} (L(\alpha, \psi(\alpha))) \quad (5.12)$$

To create such 1-D profiles, 30,000 DS simulations were first performed. During these simulations, the tested parameters and their associated JS divergence values were recorded. In our case, we were interested in minimising the objective function. Therefore, we used the created data set to plot the individual parameters against the minimum JS divergence values.

The created profiles shown in Figure 5.10 help our understanding of the optimisation problem better and its sensitivity to the parameters. For instance, they show that as the parameter  $t_{hd}$  increases, the JS divergence increases as well. This is an expected outcome as a high  $t_{hd}$  would lead to poor reproduction of the conditioning patterns; hence, would yield a high JS divergence.

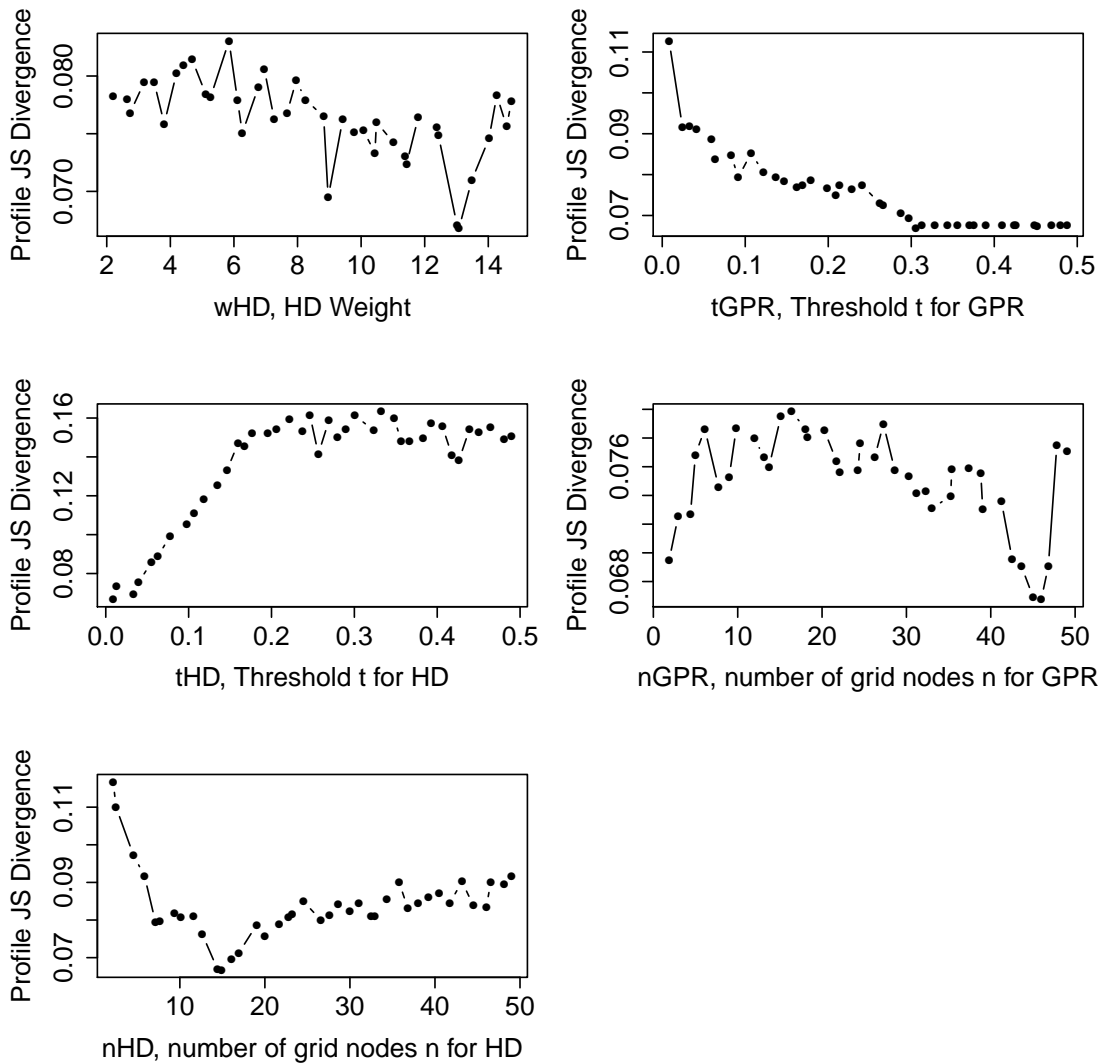


Figure 5.10: Profile JS divergences as a function of 5 DS input parameters.

For the  $n_{hd}$  parameter, the JS divergence decreases as expected with the maximum number of neighbours in the search template. But, the decrease stops at

around thirteen and then starts to slightly increase with increasing  $n_{hd}$ . It means that the conditioning data patterns are best reproduced with search templates consisting of approximately thirteen informed nodes. This is unusual since in most cases the patterns of the TI are better reproduced with more points (Meerschman et al., 2013). Here, the TI seems to contain features that are not present in the conditioning punctual data. One can see this as an indication that the TI is not fully compatible with the point data. But also, a very interesting feature is that the proposed automatic parameter identification is able to find automatically the best compromise between reproducing the patterns from the TI and those of the available point data. It is able to adapt the level of reproduction of the TI to ensure compatibility with the conditioning data.

Similarly, there is an apparent trend in the profile JS divergence for  $t_{gpr}$ . As the  $t_{gpr}$  increases, the JS divergence decreases and yields better simulation results. This is an unusual behaviour and can be better understood by being reminded of how the DS works. Considering a bivariate pattern collected from the SG, the grid nodes of the TI are visited to find a compatible match. During this search, a compatible pattern is considered to be the one with computed distances lower than the specified  $t_{hd}$  and  $t_{gpr}$ . An increase in  $t_{gpr}$  increases the probability of accepting a GPR component of the multivariate pattern as compatible. Therefore, the effect of the GPR variable becomes less pronounced in the simulations. Secondary information is normally considered to enhance the quality of the simulations. Here, the TI could be imperfectly compatible with the simulated area, as the multiple-point dependence between the reference mine and the simulation area could be different.

For the  $n_{GPR}$  and  $w_{HD}$  parameters, the plots are noisy and do not exhibit a simple systematic decrease or increase. Considering that these plots were created using 30,000 simulation data, these noisy plots should not be due to insufficient data. A more likely cause could be the existence of many local minima. This could explain why the SA performed more efficiently than the other optimization methods.

Another observation is related to the sensitivity of the JS divergence to the parameters. When the range of the JS divergence values are considered,  $t_{hd}$  seems

to be the most sensitive parameter as it has a high range of JS divergence values. This parameter is followed by  $n_{hd}$  and  $t_{gpr}$ . The parameters  $w_{hd}$  and  $n_{gpr}$  seem to have less effect on the resulting JS divergences. Another interesting observation is related to the  $t_{hd}$  and  $t_{gpr}$  values. Both threshold values lose their significance on the JS divergence values above  $t = 0.3$ . Therefore, it can be concluded that the most influential range for these parameters for this case is between 0 and 0.3.

#### 5.4.4 Analysis of the Automatically Tuned Parameters

The DS algorithm we used in this study was DeeSse. This version of the DS incorporates some modifications such as the introduction of a specific threshold for each variable rather than a global one as in the original version of Mariethoz et al. (2010). Therefore, a direct comparison of the parameters we found with the parameters suggested by Meerschman et al. (2013) (based on the original DS) cannot be made. However, we can still get an insight into the parameters from their study. For instance, they have documented an increase in the pattern consistency by altering the conditioning data weight from 1 to 5. In our case, the optimal value was found to be 13.07. This means that high importance was given to the conditioning data points. Hence, better reproduction of the conditioning data pattern was attained. Also, the threshold value for a good quality simulation is suggested to be lower than 0.1 for the continuous variables. In addition,  $t \geq 0.2$  is stated to produce noisy images. Our automatically found threshold values for the conditioning data, which is 0.009 is also consistent with their findings. However, GPR threshold  $t_{GPR}$  was determined as 0.305 and it is higher than their suggested values. The fact that  $t_{GPR}$  value is higher than  $t_{HD}$  could be due to the objective function chosen. As the aim was to reproduce the conditioning data better in the simulations, the focus has automatically become more on better pattern consistency of the reference mined-out surface patterns.

## 5.5 Summary and Conclusions

In this paper, we presented a methodology to improve bauxite footwall contact simulations through an automatic parameter tuning process. The main idea was to identify the parameters best reproducing the patterns existing in the borehole/conditioning data. There are several benefits of relying on the borehole data patterns to tune the parameters. First, although the borehole data is a partial representation of the ground truth, they serve as precise and rather reliable information. Second, in lateritic bauxite deposits, the data set can be dense enough to compute the pattern statistics of the conditioning data as a reference to be targeted in the simulations. Third, getting a TI may sometimes be difficult. Considering the mined-out topography exposed after a bauxite extraction, the resulting surface might not be entirely compatible with the conditioning data of another area. Therefore, putting more emphasis on the conditioning data and borrowing the patterns from a TI can be considered as a merge between two sets of information. As a result, the adapted MPS parameters lead to generating realisations which contain the correct patterns observed at the location where the forecast is being made.

The proposed approach mainly provides three significant advantages. The first advantage is related to the uncertainty estimation. The evolution of the calculated IQR values at each grid node reveals a fair drop in the uncertainty after the parameter optimisation. Also, the observed convergence of the simulation variograms towards the conditioning data variogram indicates that the uncertainty estimation is more reliable after optimisation. Secondly, the parameter selection is carried out more objectively than with the usual trial and error approach. Identification of the appropriate parameters is performed based on an objective and quantitative criterion. Therefore, it naturally alleviates any subjective decision that can be made as a result of the visual inspection method. Third, the approach reduces the time and effort spent for the sensitivity analysis. Contrary to a traditional parameter tuning procedure in which a set of different parameters are tested and the results are analysed manually, the proposed approach selects the parameters automatically. This reduces the time and effort spent for the manual sensitivity

analysis. Only some initial tests are required to set up the optimisation process.

In addition to the above-mentioned advantages, the selection of the pseudo-histograms of Melnikova et al. (2015) as pattern statistics provides: (1) the ability to calculate the pattern statistics of continuous images; (2) computation of the pattern statistics of regularly spaced borehole data and (3) an extensive flexibility when constructing the histogram categories. For instance, apart from the patterns of the borehole data, one can also include the patterns from the other sources to update the pattern categories. Examples of such sources may include the training images or simulations. All the patterns coming from different sources can then be used to construct the categories of the pseudo-histograms.

Visualisation of the behaviour of the JS divergence values based on the parameters has been performed through the profile-likelihood method. It allowed the analysis and better understanding of the behaviour of the objective function. The results have shown that the most influential parameters were  $t_{hd}$ ,  $t_{gpr}$  and  $n_{hd}$ . An interesting observation was made on  $t_{gpr}$ . Contrary to the expectation, the JS divergence value increased with decreasing  $t_{gpr}$ . Such a behaviour could be due to a possible TI compatibility issue and should be further investigated. In general, we believe that the observed objective function behaviours are related to both the choice of objective function as well as the case study. For instance, if the chosen objective function included the better reproduction of the TI patterns rather than the conditioning data, a different result could have been observed. Therefore, a different case study and the choice of the objective function could lead to a different behaviour.

One of the drawbacks of the proposed approach is the requirement of regularly spaced conditioning data to calculate the pattern statistics. We use the patterns with a particular spatial configuration in the pseudo-histogram construction. Therefore, the proposed method can only be used if the conditioning data is regularly spaced. Considering the stratified deposits such as lateritic bauxite, nickel or coal seams, a regularly spaced exploration borehole configuration is often implemented. Hence, the proposed approach can be used to aid parameter tuning for the MPS algorithms for such deposits. However, the conditioning data



sets may sometimes be sparse and it may not be possible to accurately infer the multiple-point statistics of the conditioning data and make a comparison with the simulations. In such a case, the objective function could focus more on the reproduction of the multiple-point statistics of the training image and can easily be integrated into the objective function in the manner described in this paper.

The general workflow proposed in the paper could be adapted to irregularly spaced conditioning data, but this would require using a different quantification of pattern statistics. One idea could be to use cumulants (Dimitrakopoulos et al., 2010), because of the tolerances it offers for irregular grids. Another option would be to define a very different objective function focusing on a cross-validation approach where the quality criteria would be based on evaluating the MPS model performances. Most likely, the simulated annealing would still remain the most efficient algorithm to obtain the parameters since previous experiences have shown that the objective functions may contain local minima (Meerschman et al., 2013).

Lastly, as this approach identifies the parameters iteratively, it requires computational time. This is a well-known feature of the SA technique. However, the strength of the approach is that it allows obtaining a global minimum even when the objective function contains many local minima. This has also been observed by the comparison made using different optimisers with different initial DS parameters. The SA does not get stuck in local minimum values as compared to the other methods. Moreover, when different the optimisations are initiated with initial input parameters, the SA appears to be the least sensitive method among the ones tested. Therefore, it might require less preliminary sensitivity analysis to define the initial parameters for the optimisations. These facts favour the use of the SA to tune the parameters automatically.

## References

- Arpat, G. B. and Caers, J. (2007). Conditional simulation with patterns. *Mathematical Geology*, 39(2):177–203.
- Boisvert, J. B., Pyrcz, M. J., and Deutsch, C. V. (2010). Multiple point metrics to assess categorical variable models. *Natural resources research*, 19(3):165–175.
- Boucher, A. (2007). Algorithm-driven and representation-driven random function: A new formalism for applied geostatistics.
- Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- Dagasan, Y., Erten, O., and Topal, E. (2018a). Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study. *The Journal of South African Institute of Mining and Metallurgy*, 118(2):173–184.
- Dagasan, Y., Renard, P., Straubhaar, J., Erten, O., and Topal, E. (2018b). Pilot point optimization of mining boundaries for lateritic metal deposits: Finding the trade-off between dilution and ore loss. *Natural Resources Research*, pages 1–19.
- De Iaco, S. and Maggio, S. (2011). Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. *Mathematical Geosciences*, 43(4):483–500.
- Diggle, P. J., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.
- Dimitrakopoulos, R., Mustapha, H., and Gloaguen, E. (2010). High-order statistics of spatial random fields: exploring spatial cumulants for modeling complex non-gaussian and non-linear phenomena. *Mathematical Geosciences*, 42(1):65–99.
- Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860.

- Erten, O. (2012). Profiling and mining control to mitigate dilution effect from sio<sub>2</sub> at the base of a bauxite deposit.
- Gloaguen, E. and Dimitrakopoulos, R. (2009). Two-dimensional conditional simulations based on the wavelet decomposition of training images. *Mathematical Geosciences*, 41(6):679–701.
- Guardiano, F. and Srivastava, R. (1992). Borrowing complex geometries from training images: The extended normal equations algorithm. *Stanford Center for Reservoir Forecasting Report, Stanford University*.
- Honarkhah, M. and Caers, J. (2010). Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences*, 42(5):487–517.
- Huysmans, M. and Dassargues, A. (2011). Direct multiple-point geostatistical simulation of edge properties for modeling thin irregularly shaped surfaces. *Mathematical Geosciences*, 43(5):521.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.
- Liu, Y. (2006). Using the snesim program for multiple-point statistical simulation. *Computers & Geosciences*, 32(10):1544–1563.
- Mariethoz, G. and Caers, J. (2014). *Multiple-point geostatistics: stochastic modeling with training images*. John Wiley & Sons.
- Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11).
- Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., and Renard, P. (2013). A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. *Computers & Geosciences*, 52:307–324.

- Melnikova, Y., Zunino, A., Lange, K., Cordua, K. S., and Mosegaard, K. (2015). History matching through a smooth formulation of multiple-point statistics. *Mathematical Geosciences*, 47(4):397–416.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.
- Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G. (2014). Simulation of rainfall time series from different climatic regions using the direct sampling technique. *Hydrology and Earth System Sciences*, 18(8):3015.
- Pérez, C., Mariethoz, G., and Ortiz, J. M. (2014). Verifying the high-order consistency of training images with data for multiple-point geostatistics. *Computers & Geosciences*, 70:190–205.
- Pirot, G., Straubhaar, J., and Renard, P. (2014). Simulation of braided river elevation model time series with multiple-point statistics. *Geomorphology*, 214:148–156.
- Powell, M. J. (2009). The bobyqa algorithm for bound constrained optimization without derivatives. *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge, pages 26–46.
- Reeves, C. R. (1993). *Modern heuristic techniques for combinatorial problems*. John Wiley & Sons, Inc.
- Renard, P. and Allard, D. (2013). Connectivity metrics for subsurface flow and transport. *Advances in Water Resources*, 51:168–196.
- Rezaee, H., Mariethoz, G., Koneshloo, M., and Asghari, O. (2013). Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. *Computers & geosciences*, 54:293–308.
- Straubhaar, J. (2016). Deesse users guide. *The Centre for Hydrogeology and Geothermics (CHYN)*, University of Neuchâtel.

- Straubhaar, J., Renard, P., and Mariethoz, G. (2016). Conditioning multiple-point statistics simulations to block data. *Spatial Statistics*, 16:53–71.
- Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.
- Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical geology*, 34(1):1–21.
- Sun, N.-Z. and Sun, A. (2015). *Model Calibration and Parameter Estimation: For Environmental and Water Resource Systems*. Springer.
- Szu, H. and Hartley, R. (1987). Fast simulated annealing. *Physics letters A*, 122(3-4):157–162.
- Tahmasebi, P., Hezarkhani, A., and Sahimi, M. (2012). Multiple-point geostatistical modeling based on the cross-correlation functions. *Computational Geosciences*, 16(3):779–797.
- Tan, X., Tahmasebi, P., and Caers, J. (2014). Comparing training-image based algorithms using an analysis of distance. *Mathematical Geosciences*, 46(2):149–169.
- Team, R. C. (2017). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2016.
- Tsallis, C. and Stariolo, D. A. (1996). Generalized simulated annealing. *Physica A: Statistical Mechanics and its Applications*, 233(1-2):395–406.
- Xiang, Y., Gubian, S., and Martin, F. (2017). Generalized simulated annealing. In *Computational Optimization in Engineering-Paradigms and Applications*. In-Tech.
- Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for global optimization: The gensa package. *R Journal*, 5(1).

- Zhang, T., Switzer, P., and Journel, A. (2006). Filter-based classification of training image patterns for spatial simulation. *Mathematical Geology*, 38(1):63–80.
- Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560.

## Chapter 6

# Pilot-point optimisation of mining boundaries for lateritic metal deposits: finding the trade-off between dilution and ore loss<sup>1</sup>

---

<sup>1</sup>This chapter has been published in *Natural Resources Research* as:

**Dagasan, Y.**, Renard, P., Straubhaar, J., Erten, O., Topal, E. (2018) Pilot-point optimisation of mining boundaries for lateritic bauxite deposits: finding the trade-off between dilution and ore loss. *Natural Resources Research*, DOI: 10.1007/s11053-018-9380-9 and reprinted with permission in Appendix I

**Abstract**

Geological contacts in lateritic metal deposits (footwall topographies) often delineate the orebody boundaries. Spatial variations seen in such contacts are frequently higher than those for the metal grades of the deposit. Therefore, borehole spacing chosen based on the grade variations cannot adequately capture the geological contact variability. Consequently, models created using such boreholes cause high volumetric uncertainties in the actual and targeted ore extraction volumes, which, in turn, lead to high unplanned dilution and ore losses. In this paper, a method to design optimum ore/mining boundaries for lateritic metal deposits is presented. The proposed approach minimizes the dilution/ore losses and comprises two main steps. First, the uncertainty on the orebody boundary is represented using a set of stochastic realizations generated with a multiple-point statistics algorithm. Then, the optimal orebody boundary is determined using an optimization technique inspired by a model calibration method called *Pilot Points*. The pilot points represent synthetic elevation values and they are used to construct smooth mining boundaries using the multilevel B-Spline technique. The performance of a generated surface is evaluated using the expected sum of losses in each of the stochastic realizations. The Simulated Annealing algorithm is used to iteratively determine the pilot point values which minimize the expected losses. The results show a significant reduction in the dilution volume as compared to those obtained from the actual mining operation.

**6.1 Introduction**

Given a laterite-type bauxite deposit formed from tropically weathered mafic-ultramafic complexes, the bauxite mineral exists in the soil horizons (Erten, 2012). Therefore, the deposit can be mined easily by a front-end loader due to the free-flow characteristics of the loose soil. Being an underlying geological unit, ferricrete is very likely to dilute the bauxite ore during mining operations due to poorly defined geological interface between bauxite and ferricrete units. Although this dilution can partly be alleviated by the front-end loader operator, who subjectively



discriminates the bauxite ore from the ferricrete based on the hardness and color differences of the geological units at the time of mining, it still cannot be avoided entirely.

Ferricrete dilution is the major cause of high silica content in the bauxite ore, as it is tremendously rich in silica-bearing minerals such as kaolinite and quartz (Morgan, 1995). The contact topography between the bauxite and ferricrete units is rather undulating and cannot be modeled satisfactorily by using an economically viable drilling program, as the drill spacing is usually determined based on the continuity and variation in the aluminum grade (Singh, 2007; Hartman and Mutmamsky, 2002). In other words, since the peaks and troughs cannot be sampled adequately, they cannot be inferred from the geostatistical estimates either (Philip and Watson, 1986). This situation is also illustrated in Figure 6.1. Failing to model the contact surface accurately introduces a major uncertainty, which may then lead to the following: (1) inaccurate calculations of the ore volume/tonnage and the quantity of the caustic soda being consumed; and (2) subjective ore extraction strategies by the front-end loader operators.

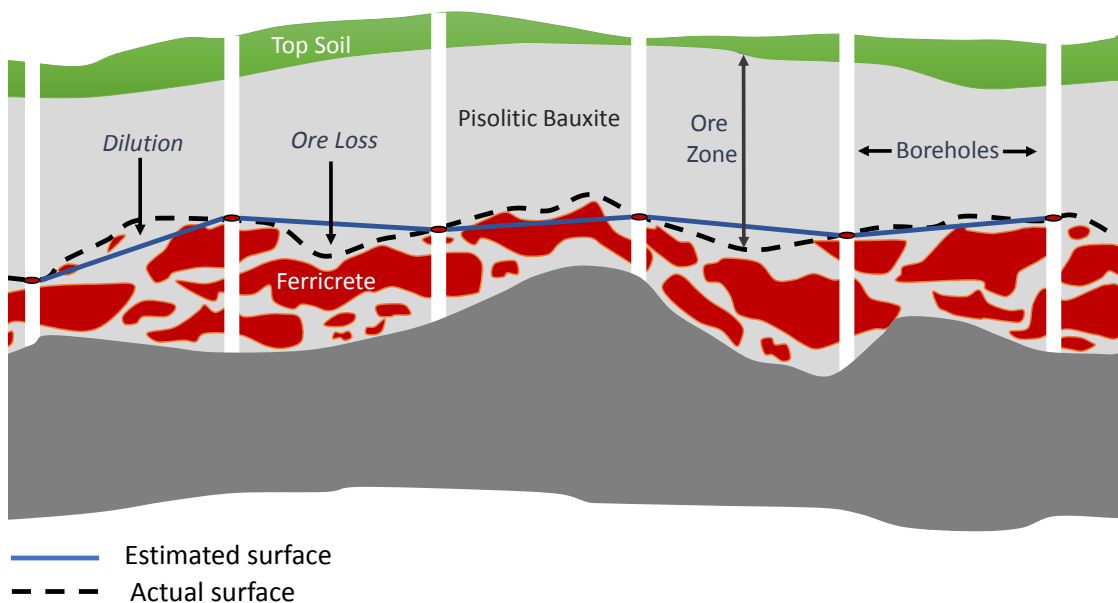


Figure 6.1: The peaks and troughs of the actual ore/waste interface cannot be detected by an economically viable drilling spacing. This results in an inaccurate estimation of the ore/waste contact (after Erten (2012))

There are several ways to reduce the uncertainty in the contact surface and its possible consequences. One of the easiest ones would be to conduct a dense drilling program to capture the peaks and troughs of the contact surface. However, this would dramatically increase the associated costs making the operation less profitable and even not feasible at all. Another way to reduce the uncertainty in the contact surface models is the use of geophysical methods to contribute to the orebody delineation (Campbell, 1994; Fallon et al., 1997). Among the geophysical methods, ground penetrating radar (GPR) has been efficiently used to improve the delineation of ore/waste boundaries in lateritic ore bodies (Francke, 2010, 2012b,a; Francké and Nobes, 2000; Francke and Parkinson, 2000; Francke and Utsi, 2009; Francké and Yelf, 2003; Barsottelli-Botelho and Luiz, 2011; Daganan et al., 2018). However, GPR surveys alone cannot replace the traditional drilling due to their lack of accuracy. They are most efficiently used as secondary information to complement the borehole data through geostatistical data integration techniques (Erten, 2012). Applications of such data integration presented by Erten (2012); Erten et al. (2013, 2015) demonstrate the benefits of using the secondary information on the model precision. However, even though a better model representing the ore/waste contact surface is attained, the large spatial variations inherent to the ore/waste interface limits the mining equipment to track down a given contact surface accurately.

Due to large spatial variations and the uncertainties inherent to the geological contact, any excavation surface inevitably causes dilution and ore losses. Although this problem shows a similarity with the dig-limit problems in open-pit mining, which has been covered by several studies such as Norrena and Deutsch (2001, 2002); Richmond (2004); Richmond and Beasley (2004); Isaaks et al. (2014); Ruiseco et al. (2016); Ruiseco and Kumral (2017); Sari and Kumral (2018), the problem with lateritic deposits is rather specific due to the nature of free-digging mining method. Research on finding the optimum elevation values for a lateritic nickel mine has been carried out by McLennan et al. (2006), but the focus was to optimize the dilution and ore losses. The approach did not put a strong emphasis on the equipment selectivity due to low dilution/ore loss ratio and good equipment

selectivity.

The aim of this research is to design optimum extraction boundaries for lateritic metal deposits based on the simulated ore/waste interface. The proposed approach can be used to generate mining boundaries minimizing the unplanned dilution and ore loss as well as increasing the mining equipment flexibility. It is inspired by a model calibration technique, which is frequently used in hydrogeology, called pilot points. In this technique, several pilot points are first placed in the area to be mined out. These pilot points represent synthetic elevation values and act as points controlling the shape of proposed extraction boundaries. The elevation values at the pilot points are iteratively modified in order to find a smooth excavation surface minimizing the possible dilution and ore losses. Multilevel B-spline method (MBS) (Lee et al., 1997) was used to create a smooth surface by interpolating the values at the pilot point locations with a predefined smoothness parameter. The losses associated with a decision surface are calculated using several hundreds of equiprobable realizations generated using the direct sampling (DS) (Mariethoz et al., 2010) multiple-point statistics algorithm. This makes the generated excavation surfaces account for the uncertainties in the ore/waste interface. The elevation values at the pilot point locations were iteratively optimized using the simulated annealing (SA) algorithm (Kirkpatrick et al., 1983), which uses the sum of the losses in all the realizations as the objective function. The pilot point values yielding minimum losses were then employed to construct the suggested extraction surface.

## 6.2 Review of underlying methods

The following subsections provide the required background information to comprehend these methods, which form the foundations of the methodology described in Section 6.3.

### 6.2.1 The Direct Sampling MPS Algorithm

The direct sampling (DS) is a pixel-based MPS algorithm used to simulate a random function  $Z(x)$  on a simulation grid (SG) (Mariethoz et al., 2010). It stochastically reproduces the spatial or temporal patterns in the simulation domain by integrating the datasets from analogue sites through training images (TI) (Oriani et al., 2014; Pirot et al., 2014). A TI serves as a conceptual geological model and contains spatial structures that are thought to exist in the simulation area (Guardiano and Srivastava, 1992). The DS uses the spatial patterns in the TI to stochastically simulate a random function  $Z(x)$ . The steps to perform the simulations are as follows:

1. Migrate any available conditioning data to the SG.
2. Visit a non-informed grid node at  $x$  following a predefined random or regular path.
3. Determine  $n$  number of closest informed nodes at  $\{x_1, x_2, \dots, x_n\}$ .
4. Define the lag vectors  $L = \{h_1, h_2, \dots, h_n\}$ , where  $h_i = x_i - x$ , to construct the data event  $d_n(x, L) = \{Z(x + h_1), \dots, Z(x + h_n)\}$ .
5. Randomly scan the TI at  $y$  locations and calculate the distance between  $d_n(x, L)$  and  $d_n(y, L) = \{Z(y + h_1), \dots, Z(y + h_n)\}$  until it falls below a threshold  $t$  or a maximum scan fraction  $f$  is reached.
6. Take the pattern as the best match and paste the central node  $Z(y)$  to the grid node at  $x$  location.
7. Repeat the steps 2-6 until all the grid nodes are informed.

The DS algorithm also makes it possible for multivariate simulations of  $m$  variables, which are spatially dependent by an unknown function (Mariethoz et al., 2010). This is basically carried out by computing the distances between the joint data events  $d_n(x)$  and  $d_n(y)$  of  $m$  variables in both the SG and the TI, respectively. In this research, the MPS simulations were carried out by calling the DS algorithm,

which is coded in C, from R software (Team, 2017). The version of the DS used is called DeeSse (Straubhaar, 2016). Detailed information on the algorithm can be found in Mariethoz et al. (2010); Meerschman et al. (2013); Straubhaar (2016).

### 6.2.2 Pilot Point Method (PPM)

PPM is an inverse modeling technique that is commonly used to calibrate groundwater models (Jung, 2008). It was first suggested by de Marsily et al. (1984) and later modified by several researchers Certes and de Marsily (1991); LaVenue et al. (1995); RamaRao et al. (1995); Oliver et al. (1996); Cooley (2000); Alcolea et al. (2006). The primary motivation of the PPM was to overcome the non-uniqueness and instability problems of the previous inverse techniques using a reduced parameter space. In this method, several calibration points are first chosen from the model domain where there are no conductivity measurements taken. These points are called pilot points and represent synthetic conductivity values to be iteratively calibrated by minimizing the squared errors between the actual and observed head values. At every step, the pilot point values are used to generate the conductivity field using a geostatistical interpolation technique with a particular prescribed spatial structure inferred from the measurements.

In this research, the PPM was tailored for a mining application. Rather than calibrating the conductivity field, the method was used to create optimum mining boundaries. The pilot points located within the modelling domain control the shape of the mining boundaries and were iterated to seek the pilot point values yielding minimised dilution and ore losses. The details of our proposed mining application are explained in the following sections.

### 6.2.3 Multilevel B-Splines

The MBS method was used to interpolate or approximate a scattered dataset (Lee et al., 1997). Given a scattered dataset  $P = \{(x_c, y_c, z_c)\}$  on a  $\Omega$  domain, the method uses  $z_c$  values at  $(x_c, y_c)$  locations to carry out the approximations. A function  $f(x, y)$  approximating the values  $z_c$  at  $(x_c, y_c)$  locations were sought

to interpolate the  $Z$  field. To carry out this, the method utilizes a hierarchy of control lattices  $\Phi_0, \Phi_1, \dots, \Phi_h$  overlain the domain  $\Omega$ . Each of the control lattices  $\Phi_k$  contains a different number of control points with varying spacing. The spacing between the control points of a  $\Phi_k$  is always halved for the subsequent control lattice  $\Phi_{k+1}$ . Therefore, the  $0^{th}$  control lattice  $\Phi_0$  becomes the coarsest and  $\Phi_h$  as the finest. Approximation with the coarsest control lattice  $\Phi_0$  comprises the first step of the MBS method yielding  $f_0$  function. Being an initial smooth approximation,  $f_0$  results in a deviation  $\Delta^1 z_c = z_c - f_0(x_c, y_c)$  for each point  $(x_c, y_c, z_c)$ . The algorithm proceeds by using the next control lattice  $\Phi_1$  to generate a function  $f_1$  which approximates the preceding deviation  $P_1 = \{(x_c, y_c, \Delta^1 z_c)\}$ . A better approximation with a less departure from the original data points  $P$  would then be obtained by the sum of  $f_0 + f_1$ . This would result in the deviation  $\Delta^2 z_c = z_c - f_0(x_c, y_c) - f_1(x_c, y_c)$ . Therefore, the deviation for a level  $k$  can be calculated as  $\Delta^k z_c = z_c - \sum_{i=0}^{k-1} f_i(x_c, y_c)$ . Since the origin of the approach creates a surface approximating the points  $P$ , the interpolation is achieved through a sufficiently small finest control lattice  $\Phi_h$ . The introduction of the adaptive control lattice hierarchy helps to achieve finer lattices with a reasonable memory requirement. More information regarding the theory can be found in Lee et al. (1997). The MBS method in this research is implemented using the MBA package created for the R statistics software (Finley and Banerjee, 2010).

#### 6.2.4 Simulated Annealing (SA) Algorithm

Simulated annealing (SA) is one of the stochastic optimization techniques used to solve global optimization problems (Kirkpatrick et al., 1983; Xiang et al., 2013). The method finds the global minimum of an objective function by mimicking the annealing process of a molten metal. The artificial temperatures used in the algorithm allows to regulate the cooling schedule and to introduce stochasticity. This stochasticity is basically used to avoid the solution from trapping inside a local minimum by changing the probability of acceptance throughout the cooling schedule.

Given an objective function  $f(x)$  with the decision variables  $x = (x_1, x_2, \dots, x_n)$ ,

the SA algorithm utilizes the following to attain a global minimum (Sun and Sun, 2015):

1. Set a high initial temperature value  $T_0$  and an initial solution  $x^0$  to evaluate the objective function  $E^0 = f(x^0)$ .
2. Propose a new candidate solution  $x^{i+1}$ :
  - Propose a candidate solution  $x^{i+1}$  based on the current one ( $x^i$ ) through a predefined visiting distribution.
  - Evaluate the energy difference  $\Delta E = f(x^{i+1}) - f(x^i)$  to observe the change in the objective function for the candidate solution.
  - Accept the iteration if the candidate solution reduces the objective function,  $\Delta E < 0$ .
  - If the new candidate yields a greater objective function value, accept or reject the solution based on a probability of acceptance criterion.
3. Repeat step 2 for  $L$  number of iterations holding  $T$  constant.
4. Reduce the temperature to  $T_{n+1}$  based on a cooling function.
5. Repeat steps 2-4 until the convergence is achieved.

In this research, Generalized Simulated Annealing (GSA) method (Tsallis and Stariolo, 1996) was used to optimize the pilot point values. It makes use of the distorted Cauchy-Lorentz visiting distribution to seek for an optimum solution (Tsallis and Stariolo, 1996). The GSA offers different options for the stopping criteria such as maximum running time, maximum function calls, maximum iteration number or a threshold value for the objective function. The implementation of the GSA was performed using the GenSA package of the R statistics software (Xiang et al., 2013). The default SA parameters of the package were set to solve complex optimization problems (Xiang et al., 2017). Therefore, these values were used to optimize the pilot point values in this research.

## 6.3 Methodology

The methodology of the proposed approach includes several steps to generate an optimum ore/waste boundary. First one of these is to create an ensemble of equiprobable realizations representing the uncertainty on the position of the ore/waste interface. This step is followed by locating some pilot points in the simulation grid and fitting a smooth surface to them. The elevation values of the pilot points are then iterated and updated using the SA to seek the combination of the pilot point values minimizing the total losses in each of the realizations. These steps are illustrated in Figure 6.2. More information about the steps is given in the following subsections.

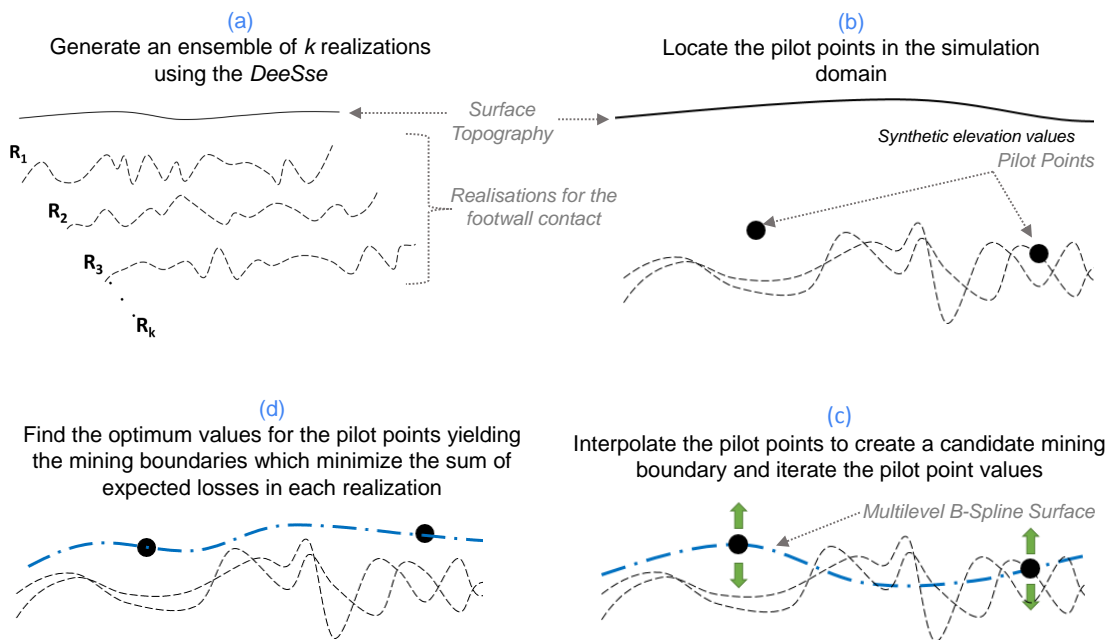


Figure 6.2: The main steps of the proposed methodology. See text for detailed explanations.

### 6.3.1 Simulations of the bauxite/ferricrete contact

The proposed methodology requires an ensemble of  $k$  conditional realizations  $R = \{R_j | j = 1, 2, \dots, k\}$  representing the ore/waste interface generated as a first



step. In order to perform the simulations, the borehole elevations of the geological contact were used as the conditioning data. Available GPR survey of the area, on the other hand, was used as the secondary information to guide the simulations. Creating such realizations rather than a single estimation plays an important role in integrating the uncertainty in the designed excavation surface.

The required simulations in this research were obtained using the DS MPS algorithm due to the benefits it provides in modeling the ore/waste boundaries of lateritic metal deposits. For instance, it utilizes a TI as a structural model, rather than a variogram. Therefore, knowledge on the spatial structures can be inferred from the previously mined-out areas through the TI concept. Since the mined-out topographies represent a complete picture of the geological variations inherent in the contact, they can provide rich structural information. In classical geostatistics, such information is derived from the sparse borehole data, which only offer a partial knowledge of the ground truth. An additional benefit of using a TI is that the resulting modeling framework is rather non-expert friendly as variogram modeling is not needed. Furthermore, the DS allows performing multivariate simulations by utilizing the multiple point dependence between multiple images. If geophysical data are available, as in our case, this can be incorporated easily in the modeling.

Although the use of MPS offers some benefits, the requirement of a TI to perform the simulations might sometimes limit its application. For instance, after the extraction of a bauxite deposit, a mined-out surface is exposed and this can be used as a TI through a topographic survey. However, such a survey data may not always be readily available. In such cases, the contact simulations can be performed using standard geostatistical simulation techniques as well. The DS MPS algorithm has been used in this research since a mined-out floor survey (TI) was already available.

### 6.3.2 Locating pilot points

Once a set of realizations for the footwall topography are generated, the next step involves locating several pilot points  $\theta = \{\theta(x_l) | l = 1, 2, \dots, m\}$  in the mining area. These pilot points function as synthetic elevation values, which are used to create

an optimum ore/mining boundary through interpolation.

The process required to set-up the pilot points can be explained in four steps, as illustrated in Figure 6.3. The first one of these is to create a grid to store their values and locations. The resolution of this grid can be chosen to be the same as the SG. Once this is done, the next step comprises locating the pilot points based on a predefined spacing. In order to better observe the effects of the chosen spacing on the results in this study, pilot points were regularly spaced. That is, if the spacing is chosen as five grid nodes, the pilot points are located at every 5<sup>th</sup> node of the pilot points grid. After this step, the initial values of the pilot points (synthetic elevation values) need to be assigned. This can either be performed by drawing numbers from a random number generator or using the simulations. The random values for the pilot points can be generated within a defined upper and lower boundary. Such boundaries can be determined using the maximum elevation value of the surface topography and the minimum elevation value of the contact realizations. Getting the initial values from the simulations can simply be achieved by copying the elevation values from the nodes of a realization which are co-located with the pilot points. The final pilot point values, on the other hand, are decided by the SA algorithm iteratively and lead to minimized dilution and ore losses. If the boundaries of the grid do not have at least one pilot point, additional pilot points are also placed at the boundaries. These additional points are required to make the interpolation cover the whole modeling domain. For example, when locating the pilot points in Figure 6.3, no points were placed in the right, left and bottom boundaries initially. Therefore, three random locations in each boundary were chosen to place an additional three pilot points.

It should be noted that the spacing chosen between the pilot points affect the smoothness of the created mining boundaries as well as the dilution/ore loss amounts. If the spacing between the pilot points is small, the resulting surface becomes more detailed. Therefore, it is advised to determine this number based on the equipment flexibility as well. A surface created using dense pilot points would yield an uneven surface which would increase the time and fuel consumption required to perform the excavation task.

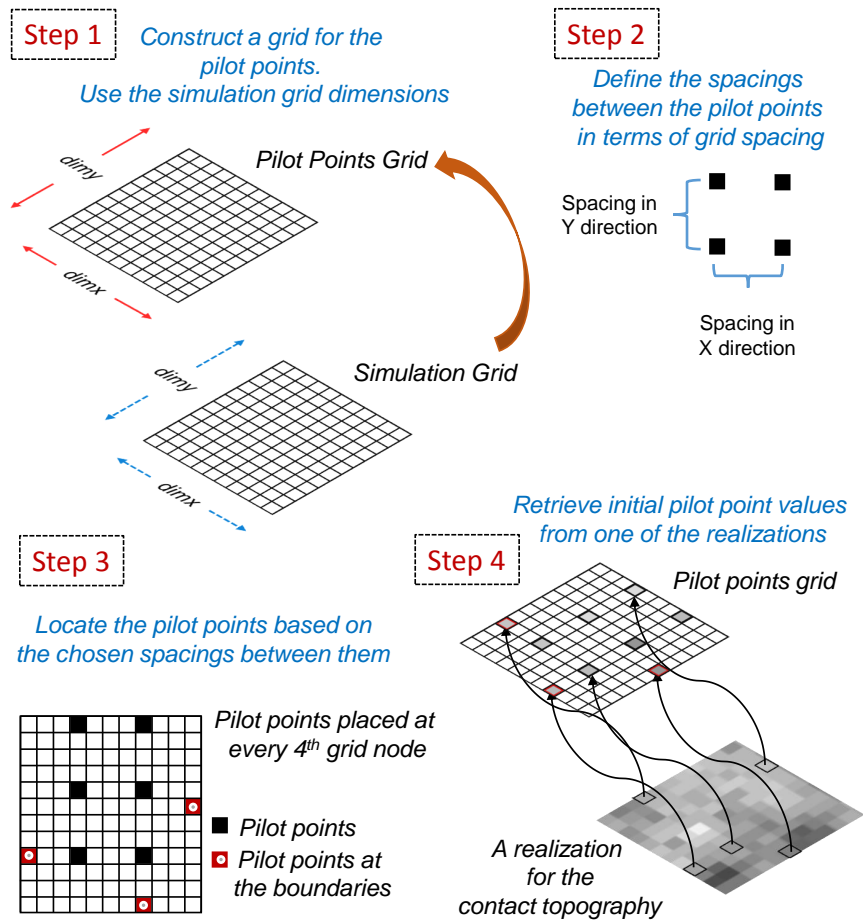


Figure 6.3: Steps followed to locate the pilot points and assign their initial values in the mining area

### 6.3.3 Smooth excavation surface design

Multilevel B-spline is used as an interpolation technique to construct the smooth excavation surface. The construction of the surface is mainly accomplished by interpolating the  $Z$  field at each grid node of the mine area using a number  $\theta$  of pilot points. The degree of fluctuations that the resulting surface exhibits is primarily influenced by two factors. The first one is the number of  $h$  levels used in the MBS interpolation. As this number increases, the fluctuations of the constructed surface also increases due to better approximations made in the finer levels. The second factor is related to the spacing between the pilot points. Small spacing

values between pilot points result in an increased number of pilot points. This would then lead to greater variations in the interpolated surface. We consider the fluctuations of the resulting surface as an essential factor for the equipment flexibility. Being able to adjust this allows one to integrate the equipment flexibility in the designed excavation surface.

### 6.3.4 Loss calculation-Objective Function

The objective of the optimization is to find the  $\theta = \{\theta(x_l) | l = 1, 2, \dots, m\}$  pilot point values, which lead to the decision surface  $S^d(\theta)$  that minimizes the sum of expected economical losses in an ensemble of  $k$  realizations:

$$\min_{\theta} \sum_{j=1}^k L_j(\theta) \quad (6.1)$$

where  $L_j(\theta)$  is the loss incurred in the realization  $R_j$  due to the decision surface  $S^d(\theta)$ . It can be calculated as follows:

$$L_j(\theta) = p_j^{max} - p_j^{act}(\theta) \quad (6.2)$$

where  $p_j^{max}$  represents the maximum profit that can potentially be made if all the ore between the surface topography and the ore/waste contact of the  $j^{th}$  realization were extracted. It can be calculated by multiplying the unit profit  $P$  by the extracted volume:

$$p_j^{max}(\theta) = P \sum_{i=1}^n T_{i,j}^{max}(\theta) \cdot A \quad (6.3)$$

where  $T_{i,j}^{max}$  represents the maximum bauxite thickness at the  $i^{th}$  grid node of the  $j^{th}$  realization (see Figure 6.4),  $A$  represents the area of a grid cell and  $n$  represents the number of informed grid nodes in the simulation area.  $T_{i,j}^{max}$  can simply be calculated by subtracting the elevation of the footwall topography realisations from

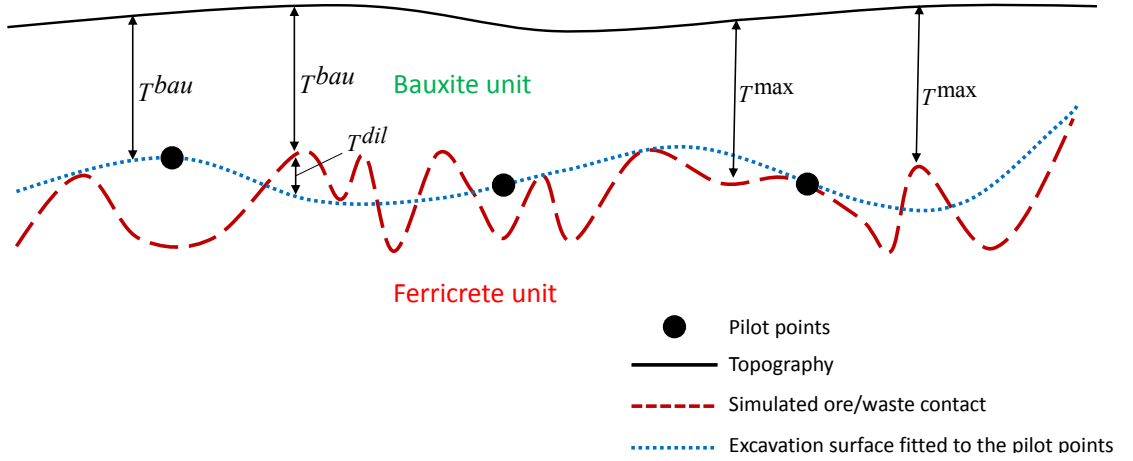


Figure 6.4: Thicknesses used to calculate the losses due to an excavation surface

those of the surface topography  $Z_i^{topo}$ :

$$T_{i,j}^{max} = Z_i^{topo} - R_{i,j} \quad (6.4)$$

$p_j^{act}(\theta)$ , on the other hand, represents the actual profit that can be made out of  $R_j$  if the mining is carried out following the boundaries defined by the decision surface  $S^d(\theta)$ . Its calculation is performed by subtracting the cost of dilution from the profit made out of extracting the ore at each grid node:

$$p_j^{act}(\theta) = P \sum_{i=1}^n T_{i,j}^{bau}(\theta) \cdot A - C \sum_{i=1}^n T_{i,j}^{dil}(\theta) \cdot A \quad (6.5)$$

where  $T_{i,j}^{bau}$  represents the mined bauxite (ore) thickness using the decision surface  $S^d(\theta)$ ,  $T_{i,j}^{dil}$  represents the ferricrete (waste) thickness overlying the decision surface and  $C$  represents the unit cost of dilution. These thicknesses can be calculated as follows:

$$T_{i,j}^{bau}(\theta) = Z_i^{topo} - \max(R_{i,j}, \min(Z_i^{topo}, S_i^d(\theta))) \quad (6.6)$$

$$T_{i,j}^{dil}(\theta) = R_{i,j} - \min(R_{i,j}, \min(Z_i^{topo}, S_i^d(\theta))) \quad (6.7)$$

To sum up, the objective function used for the optimization was evaluated

based on the expected losses incurred due to a decision surface  $S_d(\theta)$ . Its calculation was performed in four steps: (1) generation of an ensemble of realizations (only for once), (2) generating a set of pilot points, (3) fitting an MBS surface to the pilot points and (4) evaluating the loss due to the constructed smooth surface in each of the stochastic realizations.

### 6.3.5 Determination of optimum values at the pilot points

The pilot point values yielding an excavation surface that minimizes the expected loss were determined by the optimization framework of the SA algorithm. Steps of the optimization framework to design the optimal mining boundaries can be seen in Figure 6.5.

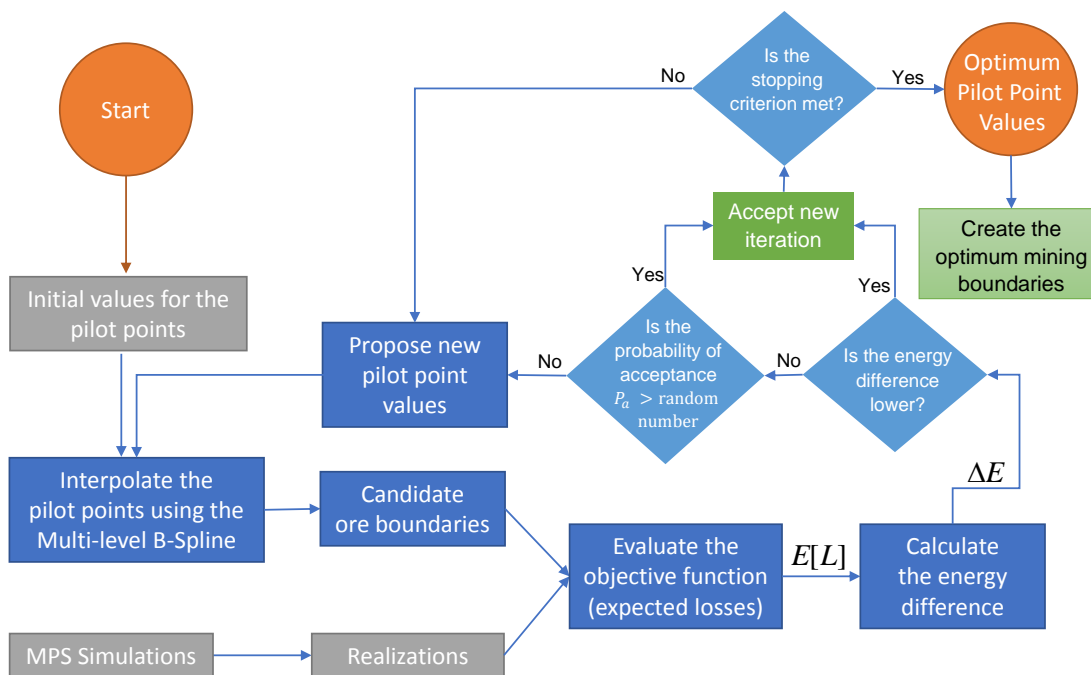


Figure 6.5: Steps used to determine the pilot points yielding optimum mining boundaries

Given a set of initial  $\theta$  pilot points (vector of decision variables), an MBS surface is first fitted to them, and the sum of losses in all the realizations is calculated.

The SA algorithm then perturbs the pilot point values using the Cauchy-Lorentz visiting distribution to evaluate the performance of a new solution (losses caused by the updated pilot point values). A change in the pilot point values yielding an improvement in the objective function (reduction in the losses) is always accepted. On the other hand, any change in the pilot point values resulting in a worse solution (increase in the loss) can be accepted or rejected based on the probability calculated using the generalized Metropolis algorithm. The acceptance probability depends on the artificial temperature parameter of the SA. As the temperature set is high in the initial stages of the optimization process, the probability of accepting worse solutions is high as well. Therefore, the solution space is well explored in the beginning. As the iterations progress, the probability of accepting a worse solution goes down since the artificial temperature approaches to zero. After several thousands of iterations, the SA converges and finds the  $\theta_{opt}$  pilot point values minimizing the losses. Once the optimum pilot point values are found, they are then used to design an optimum excavation surface through the MBS method.

## 6.4 Results and Discussion

The proposed approach was implemented to generate optimum mining boundaries for a laterite-type bauxite deposit. Being an initial step of the proposed methodology, simulations of the bauxite/ferricrete interface were first performed. In order to achieve this, the elevation variable of the interface was used as the attribute to be simulated. Due to the existence both of boreholes and ground penetrating radar data (GPR), the simulations were performed in the form of bivariate simulations. This was carried out by utilizing the borehole data as the primary variable to condition the simulations and GPR data, which is exhaustively sampled throughout the simulation domain, as the secondary information to guide the simulations.

Both the borehole and the GPR data contain the elevation variable of the bauxite/ferricrete interface. The borehole elevations were obtained by observing the elevation values at which the lithology changes from bauxite to ferricrete. The GPR elevations, on the other hand, were obtained indirectly from the original raw

GPR measurements. In the first place, the raw GPR data were acquired in two-way travel time. Therefore, it initially allowed the determination of the depth from the surface to the bauxite/ferricrete interface. After subtracting the depths to the interface from the surface elevations, the GPR elevations for the bauxite/ferricrete were obtained. These elevations were used as the secondary variable to guide the simulations. The conditioning data used in the simulations are shown in Figure 6.6.

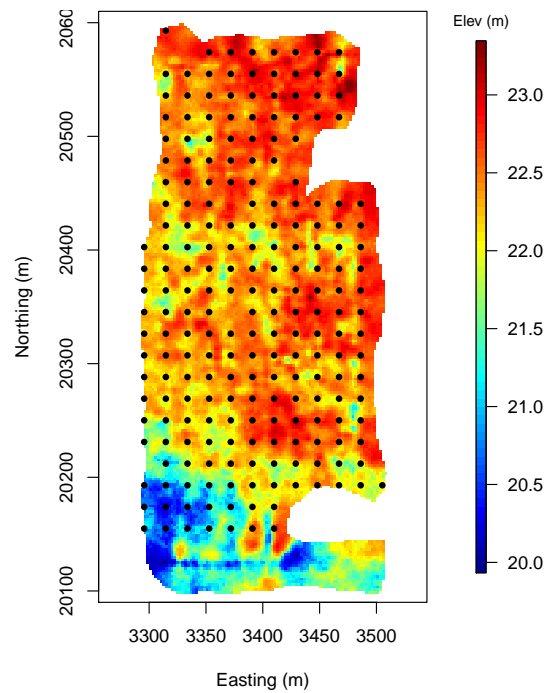


Figure 6.6: Conditioning data used in the simulations. Black dots represent the borehole data locations (primary variable) and the underlying image represents the GPR data (secondary variable)

A bivariate TI was then constructed to infer the multiple-point dependence between the borehole and the GPR data. Its variables comprise an exposed mined-out surface of a previously extracted mining area and an extensive GPR survey conducted before mining. The variables of the constructed TI can be seen in Figure 6.7.

The grid used to store the TI dataset consists of 180 nodes in easting (X) and 400 nodes in northing (Y) directions. On the other hand, the SG is comprised of 97



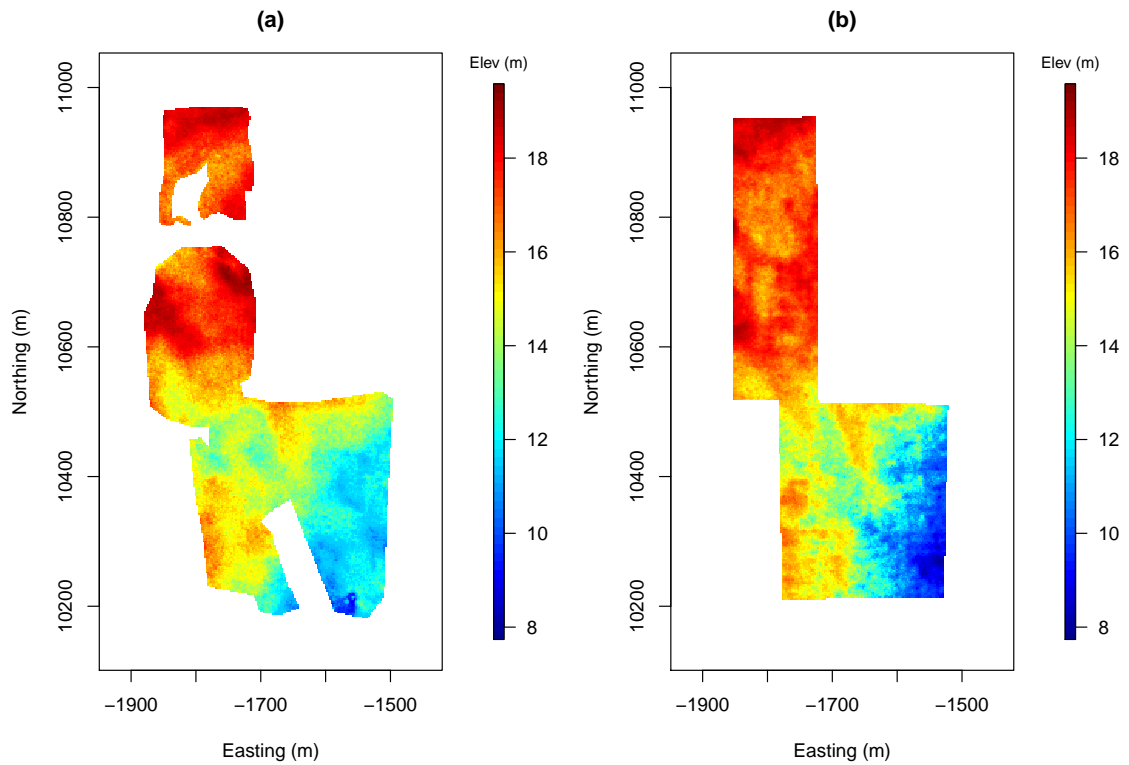


Figure 6.7: The constructed bivariate TI: (a) extraction surface of a previously mined out area and (b) extensive GPR survey carried out prior to the excavation of the mining area

and 214 nodes in both easting and northing directions, respectively. The single grid size is defined as  $2.38 \times 2.38m$  for both the TI and the SG grids. Since the original data for the GPR and the mined-out floor surface were in the form of point data, they were migrated to the TI grid node locations. This was performed through the conditional sequential Gaussian simulation (sGs) technique so as to preserve the original statistical properties as well as avoiding any smoothing effect. Using the constructed TI, the DS was used to generate an ensemble of 200 realizations. The average of the resulting simulations can be seen in Figure 6.8.

The pilot points were placed in the simulation domain based on a defined grid spacing between them. In order to analyze the effect of the spacing on the losses and the fluctuations of the decision surface, pilot points spaced 8, 16 and 24 number of grids were tested. The defined spacings yielded 251, 60 and 26 pilot

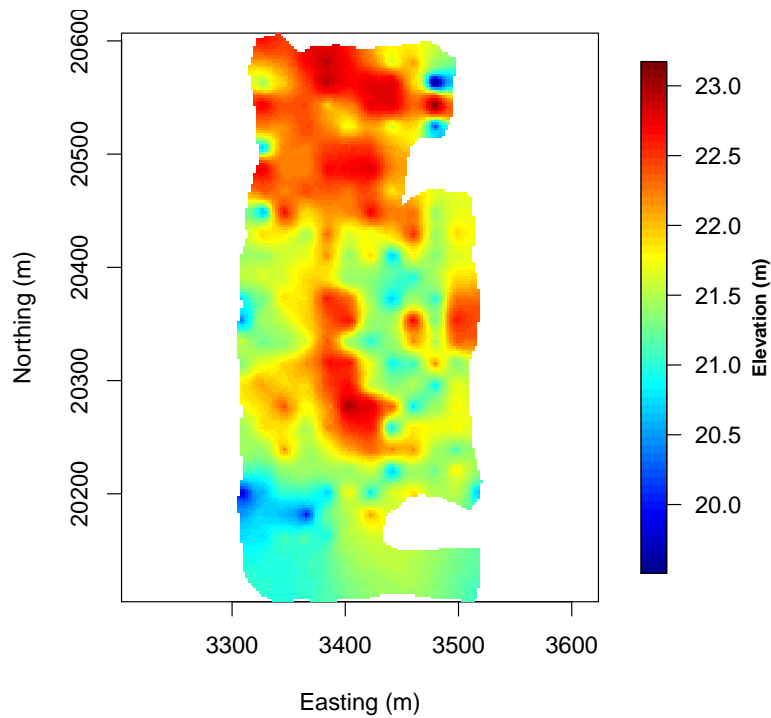


Figure 6.8: Average of 200 realizations representing the bauxite/ferricrete interface

points, respectively. Plan views of the pilot point locations in the mining area can be seen in Figure 6.9.

In addition to the spacings between the pilot points, the number of  $h$  levels used in the MBS method also affects the smoothness of the decision surface. This parameter was chosen as 10 in this study based on visually inspecting the smoothness of the resulting surface. Although our choice in this research was mainly due to the visual inspection, we suggest that a suitable value of this parameter be determined in the future to yield a design surface mimicking the front-end loader equipment selectivity.

The loss calculations require some unit costs and profits be defined. These include the profit  $P$  of mining a unit volume of bauxite ore and the cost  $C$  incurring in the case of a unit volume of dilution. Since the grade distribution does not show a significant variability throughout the deposit, we simply assume that the  $\text{Al}_2\text{O}_3\%$  grades overlying the ore/waste interface are the same everywhere. A similar assumption was also made for the  $\text{SiO}_2\%$  grades within the ferricrete unit.

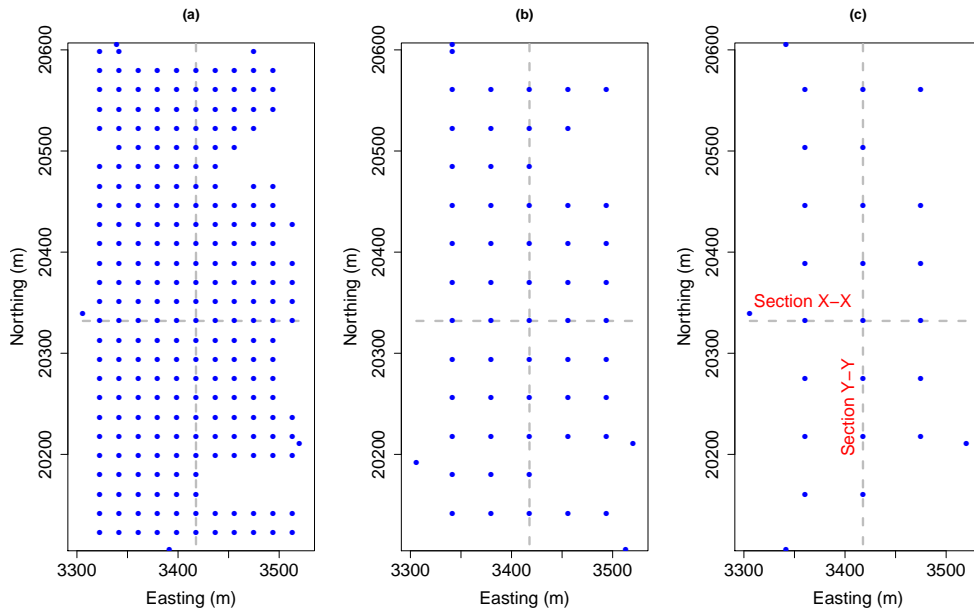


Figure 6.9: Locations of the pilot points in the mine area. Blue dots represent the pilot points located at every (a) 8th grid node, (b) 16th grid node and (c) 24th grid node. The dashed lines represent the coordinates where the cross-sections were taken as presented in Figures 6.10 and 6.11

Therefore, given that the dilution is approximately 60 times costlier than ore loss in the mining of such deposits (Erten, 2012), we simply assumed that a unit loss occurring due to dilution basically costs \$60. We also assumed that the profit made when one unit of ore is mined is \$1. It should be noted that these prices are hypothetical and might not reflect the reality.

Optimization process begins with assigning an initial set of values for the pilot points to be optimized. These values can either be randomly chosen or pre-specified before the optimization. In our case, the values were taken from the elevation values corresponding to the average of 200 realizations at the pilot point locations. We also defined a lower and an upper boundary in which the optimum values of the pilot points are sought. These boundaries function as the constraints of the optimization. The lower boundary was calculated based on the minimum elevation value of the bauxite/ferricrete realizations. Maximum elevation constraint, on the other hand, was the maximum elevation value of the topography. We defined the

maximum iteration number of the SA as 50,000 and the temperature as 5,000. An objective function call during the optimization process leads to the loss calculation for 200 images, which were of size  $97 \times 214$ , to calculate the losses in each of the realizations. Having ran the SA using the defined setup, it converged and yielded the optimum pilot point values. The resulting cross-sections of the deposit for a different number of pilot points are shown in Figures 6.10 and 6.11. Plan views of the generated smooth surfaces are shown in Figure 6.12.

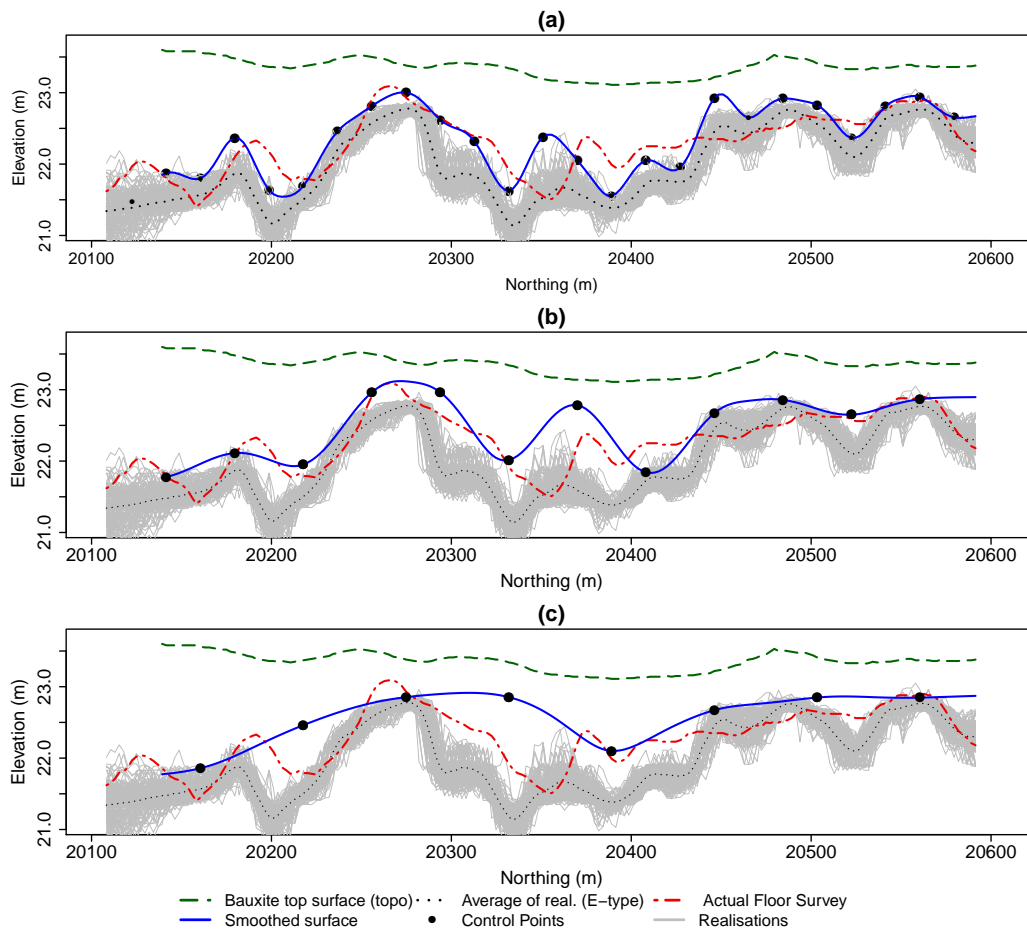


Figure 6.10: Y-Y cross-sections of the optimum surfaces found by using (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points

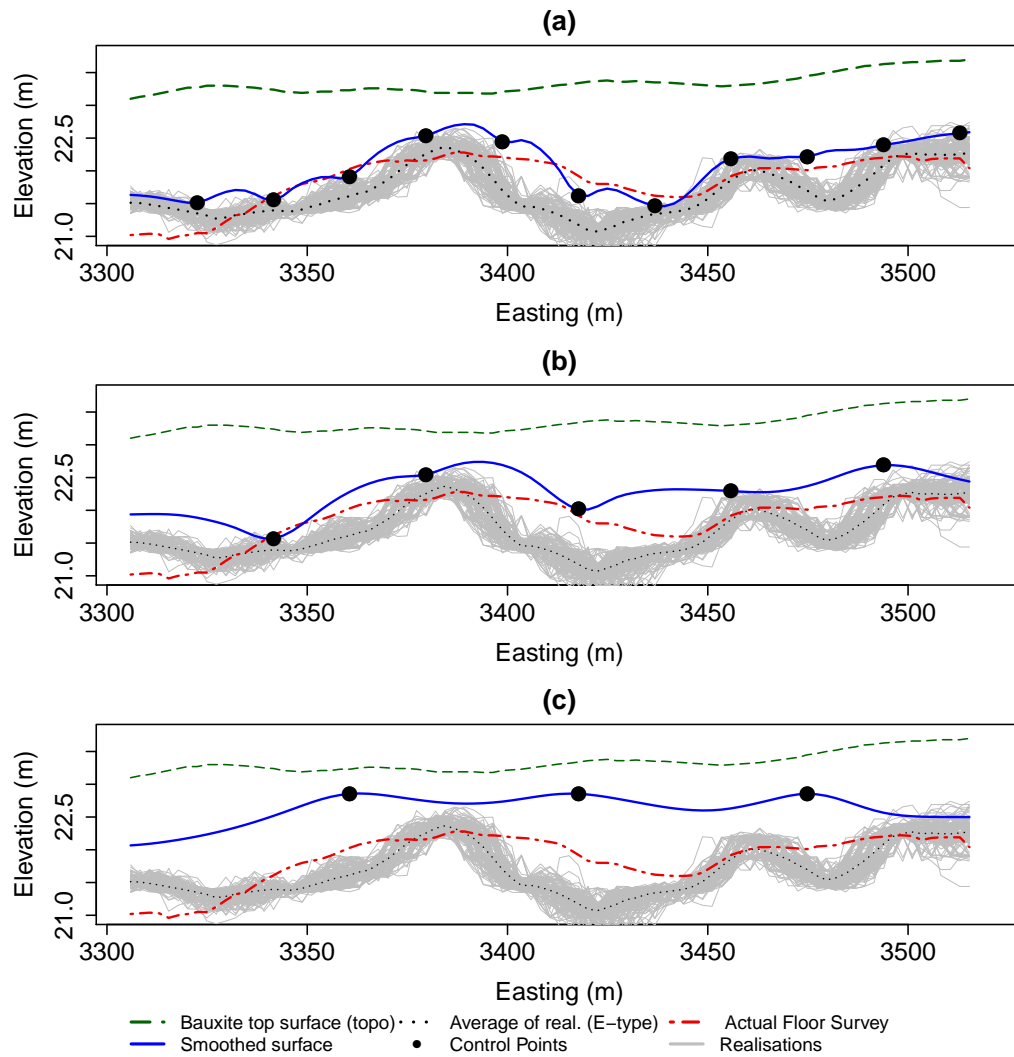


Figure 6.11: X-X cross-sections of the optimum surfaces found by using (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points

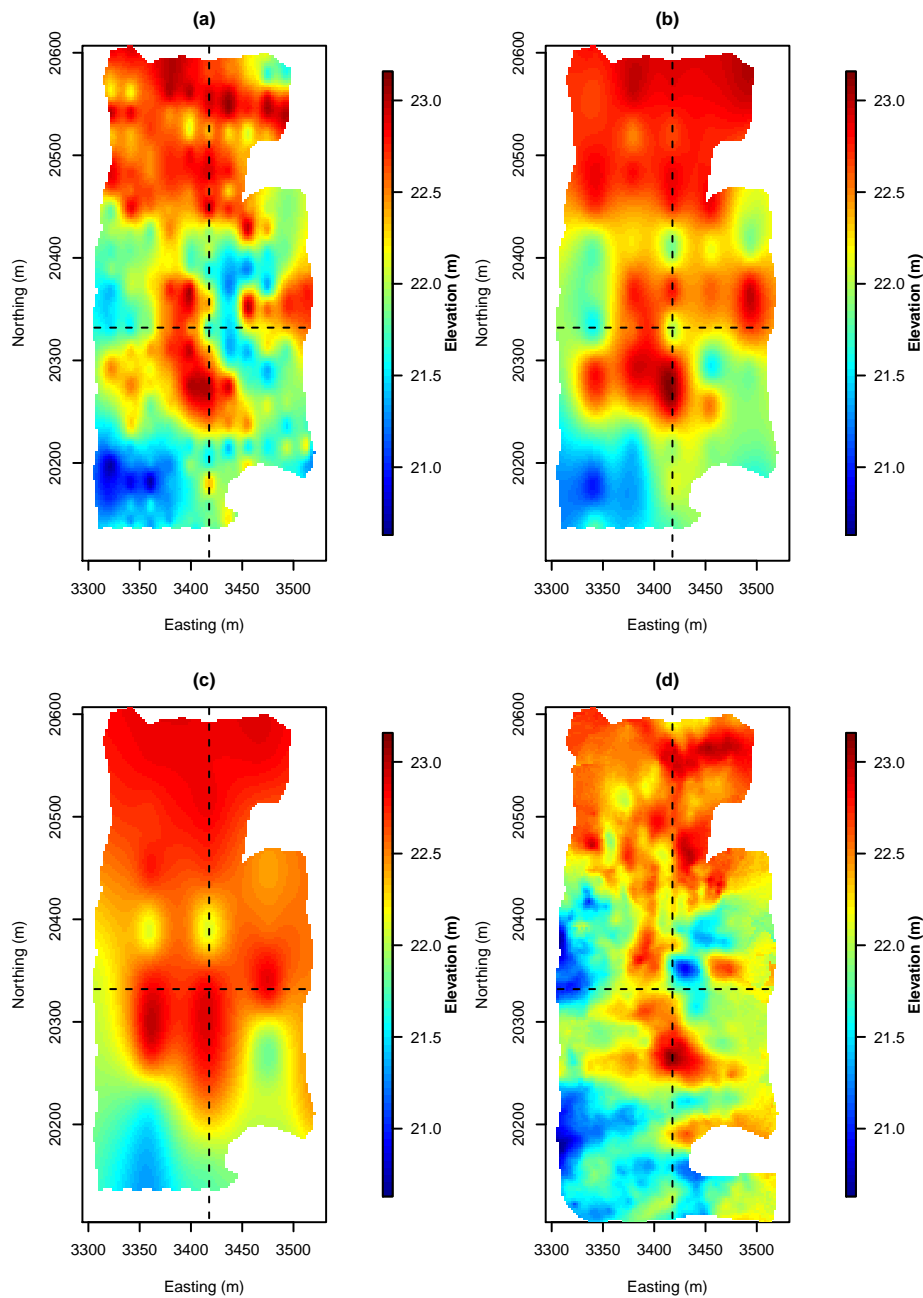


Figure 6.12: Plan view of the excavation surface designed using (a) 251 pilot points, (b) 60 pilot points, (c) 26 pilot points and (d) the actual mined-out surface. The dashed lines show the sections where the cross-sections in Figures 6.10 and 6.11 were constructed.

The cross-sections demonstrate that the optimum surfaces constructed lie above most of the realizations. This is mainly due to the introduction of a higher dilution cost compared to that of ore loss. The proposed approach automatically avoids generating a surface that causes dilution, as it leads to greater losses in the objective function. Note that the position of the decision surface may seem to be high above the simulations in certain cases, but this can be since we are only looking at a section while there are fluctuations in the perpendicular direction that the decision surface needs to consider to remain optimal.

The use of a different number of pilot points has two main consequences. The first one is about the fluctuations seen in the decision surface generated. When the number of pilot points used increases, the decision surface exhibits more fluctuations. Similarly, the use of sparser pilot points yields decision surfaces that exhibit less fluctuations, as can be seen in Figure 6.12. The second consequence is that the calculated losses decrease when the number of pilot points increases as shown in Figure 6.13. This indicates that there is a trade-off between the fluctuations and the resulting losses. More pilot points allow defining a design surface that will be rougher and more difficult to excavate but will produce higher revenue.

Following the collection of the borehole and GPR data, the area was mined out by the front-end loader operator utilizing the hardness difference between the ore/waste to track the actual geological interface. The surface exposed was mapped through a topographic survey, and the collected survey points were then used to create the complete image of the mined-out surface. Point to grid data conversion has been achieved by the conditional sGs, as in the construction of the TI. The main idea was again to prevent any smoothing effect.

In order to make a comparison, the expected volumes for the bauxite reserve, mined portion of the reserve, dilution and ore losses were calculated using the three optimized boundaries. The expected reserve volume was calculated by taking the average of the volume between the surface topography and 200 contact realizations. For the mined reserve calculations, the bauxite volume overlying the optimized surfaces in 200 realizations were averaged. If the proposed surfaces were below a realization at a grid node, the elevation differences were multiplied by the area to

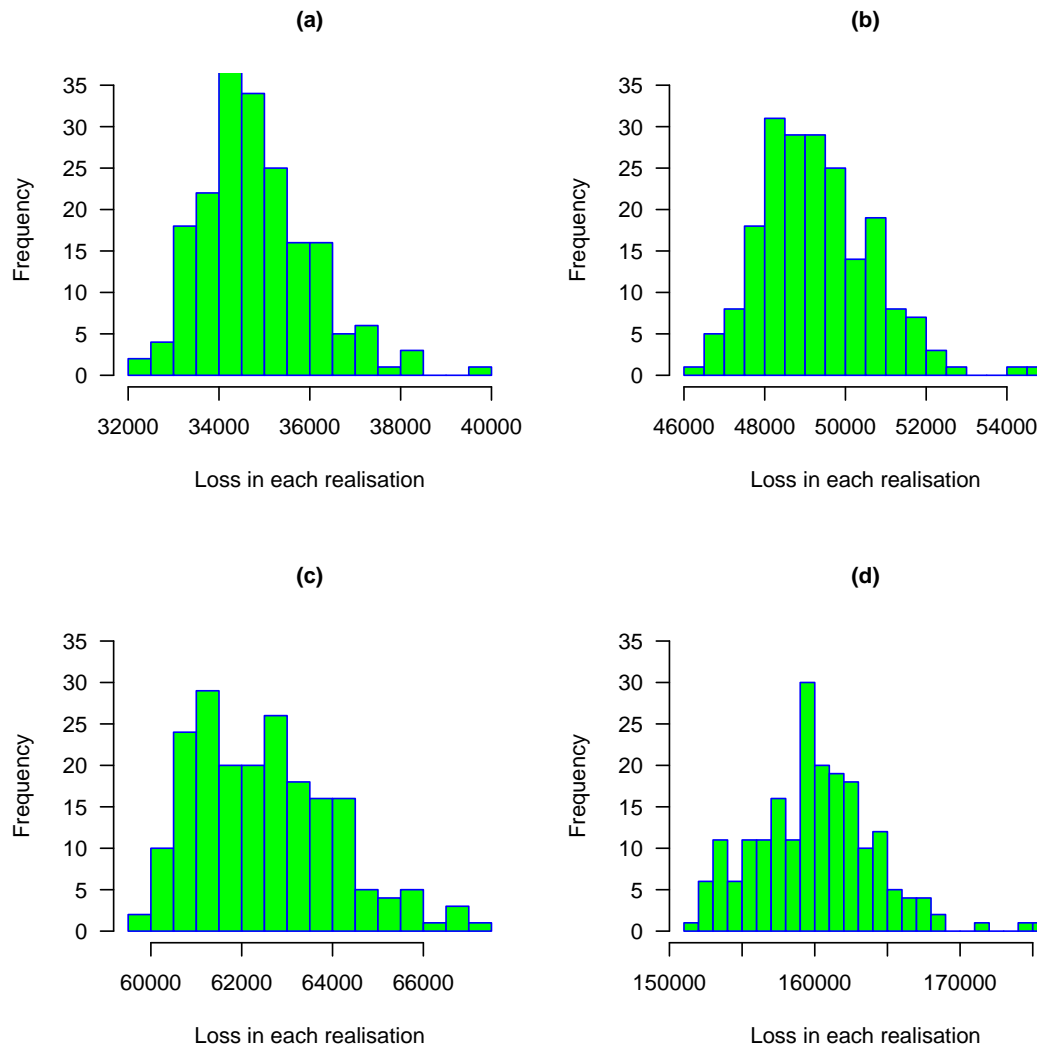


Figure 6.13: Histograms of the losses calculated using the (a) 251 pilot points, (b) 60 pilot points and (c) 26 pilot points and (d) the actual mined out surface

calculate the dilution volume. The sum of all the dilution at each grid node yielded the total dilution amount for a realization. Similarly, if the optimized surfaces were above a realization, they were considered to cause an ore loss and the associated volumes were calculated for each of the realizations. The dilution and ore losses calculated for 200 realizations were then averaged to find the expected dilution and ore losses for a given decision surface. In addition, the results of these optimized



surfaces were also compared with the mined out surface using the same calculation logic. The summary of the volume calculations given in Table 6.1 demonstrates the benefit of the proposed method. Concerning the volume calculations, the bauxite mined using 251 pilot points comprises 77% of the expected bauxite reserve. This is very similar to the amount of bauxite mined by the operator, which was 76% of the deposit. However, although both of the surfaces result in obtaining the similar amount of bauxite deposit, the dilution amount resulted using 251 pilot point surface was 0.076% of the total mined volume and significantly lower than the dilution amount of mined out surface, which was 2.7%. The percentages described are also illustrated in Figure 6.14 in terms of barplots.

Table 6.1: Statistics of the proposed and mined out surfaces

Expected Stats	251 Pilot Points	60 Pilot Points	26 Pilot Points	Mined Out
Reserve Volume ( $m^3$ )	130,448	130,448	130,448	130,448
Mined Reserve ( $m^3$ )	100,506	86,938	75,324	99,316
Ore Loss ( $m^3$ )	29,942	43,510	55,124	31,131
Ferricrete Dilution ( $m^3$ )	76	86	113	2791
Economical Losses	\$34,502	\$48,713	\$61,946	\$159,104

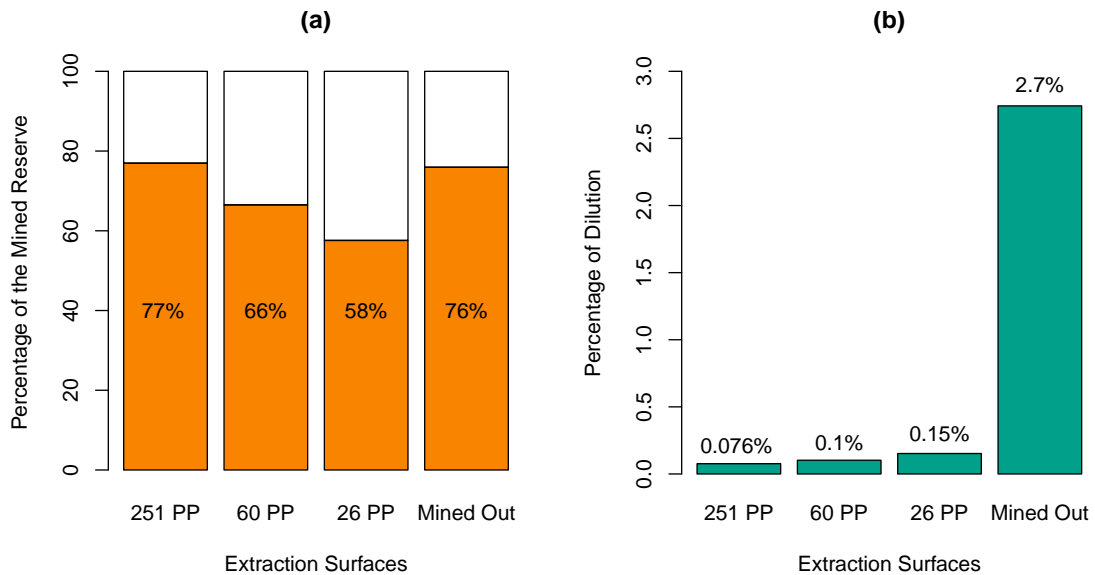


Figure 6.14: Barplots representing (a) the mined bauxite ore percentage of the expected reserve and (b) dilution percentages of the total mined volumes

## 6.5 Summary and conclusions

In this paper, we presented a new grade control technique to minimize the risk of operational dilution and ore losses in lateritic metal deposits. Although the development was performed on a lateritic metal deposit, the method can be applied to any stratified deposit to create an ore/waste boundary with a certain degree of smoothness. The proposed approach benefits from a parameter calibration technique called "pilot points" to create a design surface with multi-level B-spline method. The optimized pilot point values are iteratively obtained within the simulated annealing algorithm to create a new ore/waste boundary minimizing the risk of dilution and ore losses. Possible losses of a constructed surface are calculated using several scenarios of the ore/waste boundaries generated by the multiple-point statistical simulations. We have implemented the proposed approach on a lateritic bauxite deposit and compared the resulting losses with the ones calculated using the actual mining operation.

The major advantage of the proposed method is the reduction in the economical losses. Implementation of the method on bauxite has demonstrated much less losses compared to the actual mining operation that took place. It was also observed that the losses resulting from our proposed approach are affected by the spacing between the pilot points. Densely spaced pilot points give smaller losses but increase the fluctuations in the resulting surface. Another benefit of the proposed approach comprises the integration of the uncertainties in the ore/waste contact. The losses of a decision surface are calculated using the equiprobable realizations representing the ore/waste interface. Therefore, the uncertainty in the ore/waste boundary is accounted for by the design surface. Although the realizations were generated using multiple-point statistics, the methodology could also work well with standard geostatistical simulations, such as Turning Bands. One should, however, be cautious of the quality of the simulations used as it significantly affects the designed surface. High variability in the ore/waste contact simulations, for instance, tends to result in a design surface deviating away from the average of the realizations due to the high penalty associated with the dilution. This can lead to the underestimation of the mineable reserve volume.

The last benefit is about the adjustable smoothness of the generated surface. The number of  $h$  levels of the multilevel B-spline method allows constructing surfaces with a varying degree of smoothness. This can help designing surfaces which are capable of reflecting the mining equipment selectivity. Although  $h$  parameter of the multilevel B-spline fundamentally controls the smoothness of the surface, it needs to be calibrated in conjunction with the pilot point spacing used, as it also plays a crucial role on the fluctuations seen in the resulting surface.

The proposed approach reveals several points to be studied in the future as an improvement. The first one is the subjectivity introduced when placing the pilot points. The number and the locations of the pilot points are chosen based on personal preference in this study. Therefore, automatic determination of the number and the location of the pilot points as in Jiménez et al. (2016) may eliminate the subjectivity introduced. The second point that can be improved is related to the degree of smoothness that the decision surface exhibits. Since there is a trade-off between the losses and the degree of smoothness of the decision surface, the optimum degree of smoothness yielding minimum losses needs to be specified. This can be achieved by establishing a relationship between the surface smoothness and the mining equipment-related losses. Once such relationship is formulated as a function of the fluctuations of a given surface, this can then be used in the objective function as an additional term to calculate the total losses. Lastly, although the computation time required to perform the optimization was reasonable (6-8 hours) using the R software, it can significantly be reduced by utilizing parallel computing and coding in C language.

Instead of using the multilevel B-Spline, alternative interpolation techniques, such as kriging, can be used to generate the excavation surface. The benefit of kriging would be the possibility to adjust the smoothness of the surface with the range parameter of the variogram model. In addition, it could be possible to infer this range from previously excavated surfaces. Therefore, the equipment flexibility can automatically be integrated into the designed surface with the prior knowledge from the mined-out areas. The implementation of multilevel B-splines in this research was due to its high-speed computation (Saveliev et al., 2005). Future

research could investigate the use of kriging for the interpolation and explore its possible advantages.

Due to the fairly continuous nature of  $\text{Al}_2\text{O}_3\%$  and  $\text{SiO}_2\%$  grades throughout the deposit and also for the sake of simplicity, the grades are considered constant in this study. Therefore, in order to squeeze more performance out of the approach and also to better reflect the reality, a block model of these attributes can also be constructed to calculate the losses. Use of such a model would then involve the loss calculations based on the partial or complete mining of a specific block.

## References

- Alcolea, A., Carrera, J., and Medina, A. (2006). Pilot points method incorporating prior information for solving the groundwater flow inverse problem. *Advances in Water Resources*, 29(11):1678–1689.
- Barsottelli-Botelho, M. and Luiz, J. (2011). Using gpr to detect bauxite horizons in laterite deposits of amazon basin, brazil. In *73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011*.
- Campbell, G. (1994). Geophysical contributions to mine-development planning: A risk reduction approach. In *15th CMMI Congress, S. African Inst. Min. Metall.*, volume 3, pages 283–325.
- Certes, C. and de Marsily, G. (1991). Application of the pilot point method to the identification of aquifer transmissivities. *Advances in Water Resources*, 14(5):284–300.
- Cooley, R. L. (2000). An analysis of the pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. *Water Resources Research*, 36(4):1159–1163.
- Dagasan, Y., Erten, O., and Topal, E. (2018). Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study. *The Journal of South African Institute of Mining and Metallurgy*, 118(2):173–184.
- de Marsily, G., Lavedan, G., Boucher, M., and Fasanino, G. (1984). Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model. *Geostatistics for natural resources characterization, Part, 2*:831–849.
- Erten, O. (2012). Profiling and mining control to mitigate dilution effect from sio<sub>2</sub> at the base of a bauxite deposit.

- Erten, O., Kizil, M. S., Topal, E., and McAndrew, L. (2013). Spatial prediction of lateral variability of a laterite-type bauxite horizon using ancillary ground-penetrating radar data. *Natural resources research*, 22(3):207–227.
- Erten, O., McAndrew, L., Kizil, M. S., and Topal, E. (2015). Incorporating fine-scale ground-penetrating radar data into the mapping of lateral variability of a laterite-type bauxite horizon. *Mining Technology*, 124(1):1–15.
- Fallon, G., Fullagar, P., and Sheard, S. (1997). Application of geophysics in metalliferous mines. *Australian Journal of Earth Sciences*, 44(4):391–409.
- Finley, A. and Banerjee, S. (2010). Mba: multilevel b-spline approximation. *R package version 0.0-7*, URL <http://CRAN.R-project.org/package=MBA>.
- Francke, J. (2010). Applications of gpr in mineral resource evaluations. In *Ground Penetrating Radar (GPR), 2010 13th International Conference on*, pages 1–5. IEEE.
- Francke, J. (2012a). A review of selected ground penetrating radar applications to mineral resource evaluations. *Journal of Applied Geophysics*, 81:29–37.
- Francke, J. (2012b). The role of ground penetrating radar in bauxite resource evaluations. In *Ground Penetrating Radar (GPR), 2012 14th International Conference on*, pages 459–463. IEEE.
- Francke, J. and Parkinson, G. (2000). The new role of geophysics in nickel laterite exploitation and development. In *Mining Millennium/PDAC 2000 Conference Proceedings, Toronto*.
- Francke, J. and Utsi, V. (2009). Advances in long-range gpr systems and their applications to mineral exploration, geotechnical and static correction problems. *first break*, 27(7).
- Francké, J. C. and Nobes, D. C. (2000). Preliminary evaluation of gpr for nickel laterite exploration. In *Eighth International Conference on Ground Penetrating Radar*, volume 4084, pages 7–13. International Society for Optics and Photonics.

- Francké, J. C. and Yelf, R. (2003). Applications of gpr for surface mining. In *Advanced Ground Penetrating Radar, 2003. Proceedings of the 2nd International Workshop on*, pages 115–119. IEEE.
- Guardiano, F. and Srivastava, R. (1992). Borrowing complex geometries from training images: The extended normal equations algorithm. *Stanford Center for Reservoir Forecasting Report, Stanford University*.
- Hartman, H. L. and Mutmansky, J. M. (2002). *Introductory mining engineering*. John Wiley & Sons.
- Isaaks, E., Treloar, I., and Elenbaas, T. (2014). Optimum dig lines for open pit grade control. In *9th International Mining Geology Conference, Adelaide, South Australia*.
- Jiménez, S., Mariethoz, G., Brauchler, R., and Bayer, P. (2016). Smart pilot points using reversible-jump markov-chain monte carlo. *Water Resources Research*, 52(5):3966–3983.
- Jung, Y. (2008). *Evolutionary algorithm based pilot point methods for subsurface characterization*. PhD thesis, North Carolina State University.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *science*, 220(4598):671–680.
- LaVenue, A. M., RamaRao, B. S., De Marsily, G., and Marietta, M. G. (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 2. application. *Water Resources Research*, 31(3):495–516.
- Lee, S., Wolberg, G., and Shin, S. Y. (1997). Scattered data interpolation with multilevel b-splines. *IEEE transactions on visualization and computer graphics*, 3(3):228–244.

- Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11).
- McLennan, J., Ortiz, J., and Deutsch, C. (2006). Geostatistical simulation of optimum mining elevations for nickel laterite deposits. *CIM Magazine*, 1(6).
- Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., and Renard, P. (2013). A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. *Computers & Geosciences*, 52:307–324.
- Morgan, C. (1995). Geology of the spheres at weipa. *Trav. Com. Int. Etude Bauxites, Alumine Alum. (Yugoslavia)*, 22(24):61–74.
- Norrena, K. and Deutsch, C. (2001). Determination of dig limits subject to geostatistical, economic, and equipment constraints. In *2001 SME Annual Conference and Exhibition*, pages 133–148.
- Norrena, K. and Deutsch, C. (2002). Optimal determination of dig limits for improved grade control. In *APCOM 2002: 30 th International Symposium on the Application of Computers and Operations Research in the Mineral Industry*, pages 329–339.
- Oliver, D. S., He, N., and Reynolds, A. C. (1996). Conditioning permeability fields to pressure data. In *ECMOR V-5th European Conference on the Mathematics of Oil Recovery*.
- Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G. (2014). Simulation of rainfall time series from different climatic regions using the direct sampling technique. *Hydrology and Earth System Sciences*, 18(8):3015.
- Philip, G. and Watson, D. (1986). Matheronian geostatistics quo vadis? *Mathematical Geology*, 18(1):93–117.



- Pirot, G., Straubhaar, J., and Renard, P. (2014). Simulation of braided river elevation model time series with multiple-point statistics. *Geomorphology*, 214:148–156.
- RamaRao, B. S., LaVenue, A. M., De Marsily, G., and Marietta, M. G. (1995). Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. theory and computational experiments. *Water Resources Research*, 31(3):475–493.
- Richmond, A. (2004). Integrating multiple simulations and mining dilution in open pit optimisation algorithms. In *Orebody Modelling and Strategic Mine Planning Conference*.
- Richmond, A. and Beasley, J. (2004). Financially efficient dig-line delineation incorporating equipment constraints and grade uncertainty. *International Journal of Surface Mining*, 18(2):99–121.
- Ruiseco, J. R. and Kumral, M. (2017). A practical approach to mine equipment sizing in relation to dig-limit optimization in complex orebodies: multi-rock type, multi-process, and multi-metal case. *Natural Resources Research*, 26(1):23–35.
- Ruiseco, J. R., Williams, J., and Kumral, M. (2016). Optimizing ore–waste dig-limits as part of operational mine planning through genetic algorithms. *Natural Resources Research*, 25(4):473–485.
- Sari, Y. A. and Kumral, M. (2018). Dig-limits optimization through mixed-integer linear programming in open-pit mines. *Journal of the Operational Research Society*, 69(2):171–182.
- Saveliev, A. A., Romanov, A. V., and Mukharamova, S. S. (2005). Automated mapping using multilevel b-splines. *Applied GIS*, 1(2):17–01.
- Singh, N. (2007). Ground penetrating radar (gpr) in mineral base profiling and orebody optimization. In *6th International Heavy Minerals Conference*, pages 185–194.

- Straubhaar, J. (2016). Deesse users guide. *The Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel*.
- Sun, N.-Z. and Sun, A. (2015). *Model Calibration and Parameter Estimation: For Environmental and Water Resource Systems*. Springer.
- Team, R. C. (2017). R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2016.
- Tsallis, C. and Stariolo, D. A. (1996). Generalized simulated annealing. *Physica A: Statistical Mechanics and its Applications*, 233(1-2):395–406.
- Xiang, Y., Gubian, S., and Martin, F. (2017). Generalized simulated annealing. In *Computational Optimization in Engineering-Paradigms and Applications*. In-Tech.
- Xiang, Y., Gubian, S., Suomela, B., and Hoeng, J. (2013). Generalized simulated annealing for global optimization: The gensa package. *R Journal*, 5(1).

## **Chapter 7**

### **Conclusions and Future Work**

## 7.1 Conclusions

The main objective of this thesis was to develop approaches to improve the resource estimation of laterite-type bauxite deposits. In particular, the study focused, directly and indirectly, on the management of unplanned dilution and ore losses. To achieve the above-mentioned goal of the thesis, a series of techniques was developed and examined. The first technique utilised a framework to model the geological contact between the bauxite and ferricrete units using multiple-point statistics. The second technique was developed to automate the input parameter tuning process of the multiple-point statistical simulations. In the last technique, optimum ore boundaries were generated based on the equiprobable realisations created for the ore boundaries.

The methodology to model the geological contact was developed and investigated using the Direct Sampling algorithm. Topographies obtained in mined-out areas after the extraction of the bauxite ore, could be deemed representative of the geological contact to be modelled for a future mining area in the same mine site. Therefore, the proposed methodology utilised previously mined-out areas as analogous sites to infer the structural information through the training image concept. The simulation results have shown that the multiple-point statistics concept has successfully simulated the ore boundaries and has proven to be a promising tool to be incorporated in modelling the ore boundaries for lateritic bauxite deposits. Contrary to expectations, however, it did not significantly improve the simulations when compared to the turning bands simulations. This was thought to be due to the incompatibility observed between the training images and the data collected from the simulation areas. Nevertheless, the investigation on the incorporation of multiple-point statistics has provided a variogram-free means of modelling the ore boundaries in lateritic bauxite deposits.

The automatic parameter tuning method was developed by adopting an optimisation framework. The primary criterion for scoring the parameters was the level of success achieved in reproducing the conditioning data patterns in the simulations. Therefore, the objective function of the optimisation comprised the dissimilarity between the pattern statistics of the conditioning data and the simulated

image. Parameters minimising this dissimilarity were determined iteratively using the simulated annealing optimisation algorithm. The developed technique provided automatic determination of the appropriate simulation parameters. Hence, it allowed a reduction in time spent in the labour-intensive manual tuning of the parameters. Furthermore, it enabled a more objective determination of the parameters, as the parameters were determined based on a quantitative metric.

A method to define the optimum ore boundaries to be extracted was also developed using an optimisation framework. The primary goal set for this purpose was to find the optimum mining boundaries minimising the expected dilution and ore losses. The objective function used in the optimisation comprised the losses due to a candidate mining boundary in an ensemble of stochastic realisations of the ore boundaries. The pilot point technique, which is typically used in hydrogeology to calibrate conductivity fields, was tailored to control the shape of the candidate ore boundaries. The synthetic elevation values of the pilot points were used as the decision variables of the simulated annealing algorithm and iteratively optimised. The optimum values for the pilot points, which were determined after the optimisation process, were then used to construct optimum mining boundaries. The approach resulted in a significant reduction in the economic losses. Similarly, a reduction was observed in the ferricrete dilution in the mined bauxite volume. Another benefit of the developed method was the adjustable smoothness of the created mining boundaries. The number of  $h$  levels of the multi-level B-spline and the number of pilot points allow constructing mining boundaries with a varying degree of smoothness. Therefore, the optimum smoothness for the mining boundaries can be chosen to reflect better the capabilities of the mining machinery used to extract the deposit.

### 7.1.1 Contributions to the Current State of Knowledge

In summary, the contributions of this research to the current state of knowledge are as follows:

1. An approach to variogram-free modelling of the ore boundaries for lateritic bauxite deposits was developed.

2. A method to automatically tune the parameters of multiple-point statistical simulations for lateritic bauxite deposits was developed.
3. A method to generate optimum mining boundaries minimising the dilution and ore loss was developed.

## 7.2 Future Work

The recommendations for the future work are as follows:

### Derivation of a more compatible TI

One of the important factors affecting the MPS simulations is the TI used. In this study, two historical mine areas were used to perform the simulations. The mined-out topography and the GPR data of one mine area were used as a bivariate TI to simulate the other mine area and vice versa. Therefore, there was only one TI available to simulate each mine area. The computed experimental variograms illustrated that the two mine areas exhibit different two-point statistics. Despite this fact, the TIs were used to perform the simulations. (This was justified by the fact that the two mine areas were within the same mine site sharing the same geology.) Nevertheless, the MPS simulations reproduced the variograms of the conditioning data. If these imperfectly-compatible TIs allow the reproduction of the variograms, a better performance can be achieved by using more suitable TIs.

One way to derive a better suited TI is to form a geological catalogue. Considering the mining practices in such lateritic bauxite deposits, a mined-out topography is always obtained following the extraction of the bauxite unit. All these mined-out topographies can be collected to form a geological database. This can provide the modeller with a rich variety of TIs to choose from. Therefore, a better suited TI can be readily found.

Another approach to deriving a compatible TI could be to modify the TI. Such an approach can be performed if the available TI has spatial statistics that are different to the simulation data set, and is believed to contain relevant structural

information. This could probably be achieved by stretching or shrinking the TI until it has the desired spatial statistics.

One important aspect that should be taken into consideration is the selectivity of the mining equipment. The selectivity of the mining equipment used, which is a front-end loader, is not adequate to precisely track down the short scale fluctuations of the ore boundaries. Therefore, the produced TIs can also reflect this smoothing effect. To create better quality TIs, a pilot test area next to the future mining areas can be excavated with smaller mining machinery. This way, the operator can better track the fluctuations of the geological contact; hence, a high quality TI reflecting the geology better can be acquired.

### **Calculation of the reference pattern statistics for irregularly spaced or sparse conditioning data**

Automatic tuning of the parameters utilises the conditioning data patterns and requires that the boreholes are on a regular grid. Due to this reason, the parameter tuning process cannot be used if the conditioning data is irregularly spaced. Due to the fairly continuous grade distributions in lateritic bauxite deposits, preferential drilling is not a common practice. Hence, boreholes are often drilled on a regular grid. However, if this is not the case and the proposed approach is used for a different type of deposit, one way to acquire the pattern statistics could be the use of Cumulants. Cumulants allow computing the pattern statistics of irregularly spaced data using the lag tolerances. Another approach to solve such an irregularity problem could be to define a distinct objective function and incorporate a cross validation approach to assess the performance of the parameters.

Having a large spacing between the exploration boreholes is rather appealing due to its associated economical benefits. Therefore, there may not always be an abundance of conditioning data available in the simulation domain. When the boreholes are sparse, the pattern statistics of the conditioning data cannot be reliably computed. In such cases, rather than relying on the pattern statistics of the conditioning data, one can also incorporate the TI patterns as reference pattern statistics.

### **CPU time required to tune the parameters**

One of the drawbacks of the automatic parameter tuning method is the high CPU time demand. Although 100 iterations of the SA can provide a reasonable tuning of the parameters, a fine tuning requires more than 7,000 function calls/simulations. Given a simulation grid consisting of millions of grid nodes, that many simulations might not be feasible. Therefore, future work should also consider techniques which might not require performing simulations. A possible research direction could be the direct analysis of the TI-conditioning data to infer the appropriate DS parameters.

### **Determination of the optimum number and locations of the pilot points**

The pilot points in this study were automatically located in the simulation domain using a predefined spacing between them. In order to observe the effect of the number of pilot points, three different pilot point spacings were used to create the mining boundaries. The results have shown that as the number of pilot points increase, the possible losses decrease. This is an intuitive result since as the number of points increase, the resulting interpolated surface becomes more detailed. However, such a detailed surface would make it difficult to track the boundaries using the current mining equipment used. In other words, as the dilution and ore losses increase, the extraction duration would decrease. Longer excavation durations would increase the excavator related costs such as depreciation, fuel cost, operation cost and maintenance. Therefore, the number of pilot points chosen should also focus on this important trade-off. It should be noted that the smoothness of the created surface is also dependent on the smoothness parameter of the multilevel B-spline interpolation technique (number of levels used).

In addition, future studies should also investigate the potential benefit of using irregularly spaced pilot points (the pilot points used in this study are on a regular grid). Recent applications of automatic determination of the locations and numbers of pilot points show that reverse jump Markov Chain Monte Carlo can be used to achieve this.



### **Investigation on the benefits of other interpolation techniques for constructing the mining boundaries**

The pilot points are used as synthetic elevations to interpolate a surface. In this study, the multi level B-spline technique was used to perform the interpolations between the pilot points. The choice was due to its high speed computation. Instead of using the multi level B-spline method, a different interpolation technique such as Kriging can be used. The benefit of using Kriging would be that the smoothness of the surface could be adjusted using the range parameter of the variograms. The appropriate range parameter can be inferred from the previously mined-out areas.

### **Incorporation of real mining and processing costs through a block model**

Due to the unavailability of real mining and processing costs, several assumptions were made when calculating the losses. One of these assumptions is related to the  $Al_2O_3$  grades. The bauxite unit does not have a wide range of  $Al_2O_3$  grade values. Therefore, for the sake of simplicity, the grades were considered constant when calculating the losses. The same assumption was also made for the ferricrete unit. The volume of dilution was directly multiplied by its constant penalty. In other words, the optimisation has been made using the 1:60 ratio between the dilution and ore loss. To get better performance out of the proposed approach, future studies should focus on incorporating real costs and profits. This can be achieved by constructing the block model and assigning the associated costs and revenues depending on the block attributes. Although it would increase the CPU demand significantly, the results would reflect the reality better.

### **CPU time required to optimise the boundaries**

As in the parameter tuning process, the optimisation for the mining boundaries utilises an iterative process. The objective function calculated on each iteration is the losses generated in each realisation due to a decision surface. Therefore, the iterations of the ore boundary optimisation do not require as much computational

resource as the parameter tuning process. Due to this reason, the CPU demand is not as significant for the pilot point optimisation. To speed up the optimisation, future studies can focus on parallel computing or deterministic design of the ore boundaries utilising the model of uncertainty of the geological contact.

# Appendix A

## Simulation and TI Variables

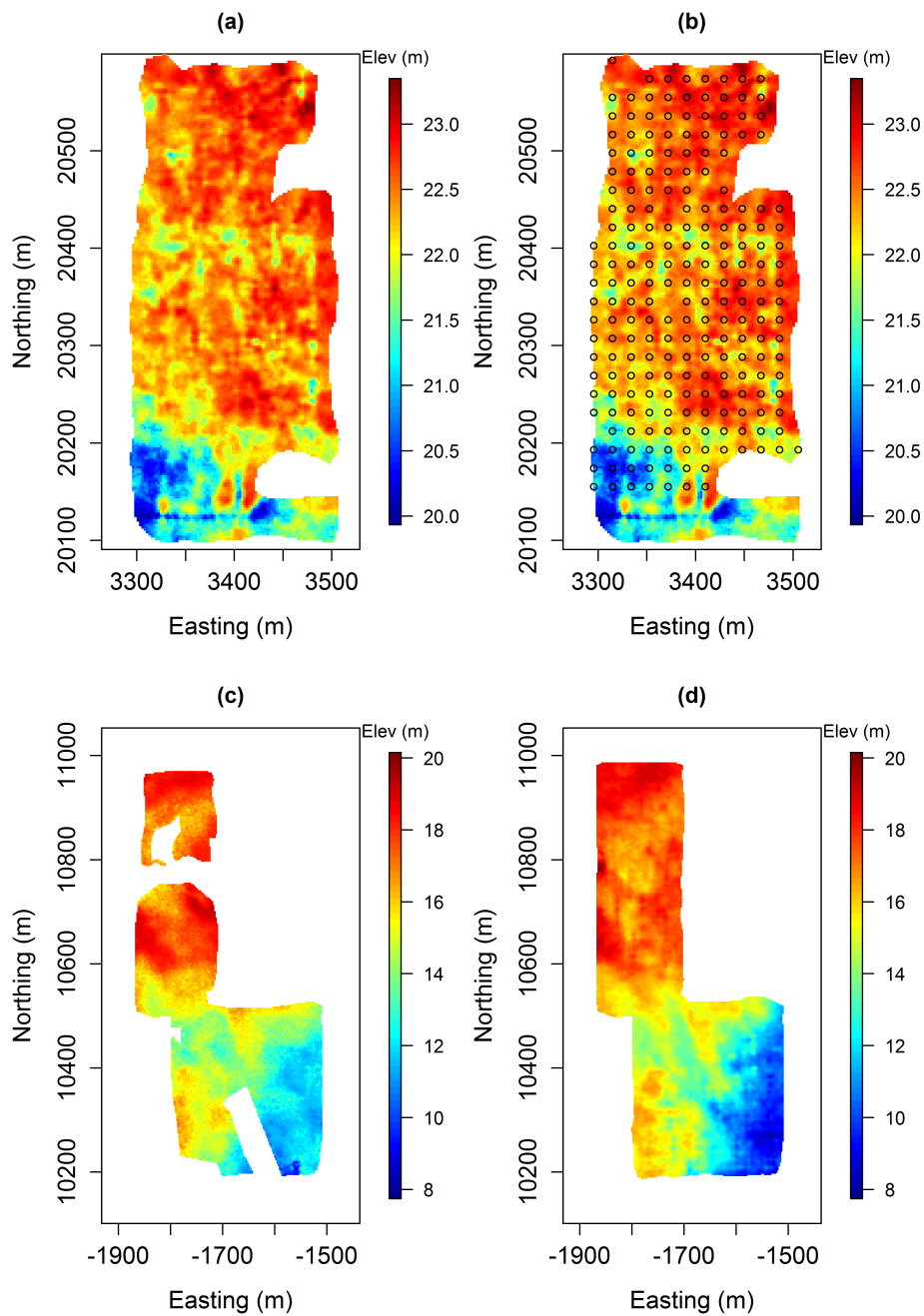


Figure A.1: Variables used to simulate the Kumbur mine area: (a) Kumbur GPR survey used to guide the simulations as a secondary variable, (b) Kumbur PCD used to condition the simulations as a primary variable (circles overlain the GPR map), (c) Oak mined-out floor topography (primary variable of the TI) and (d) Oak GPR survey (secondary variable of the TI)

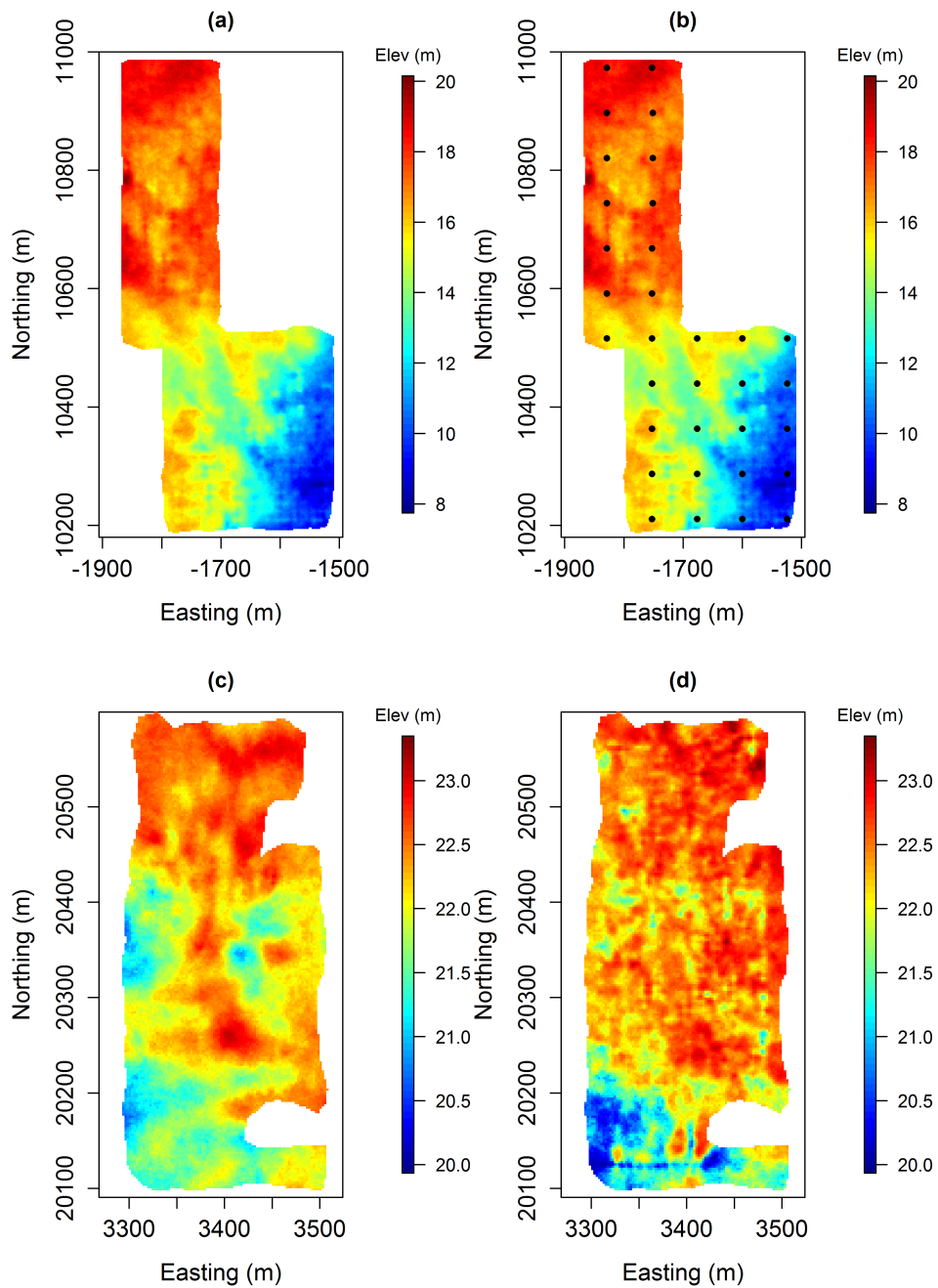
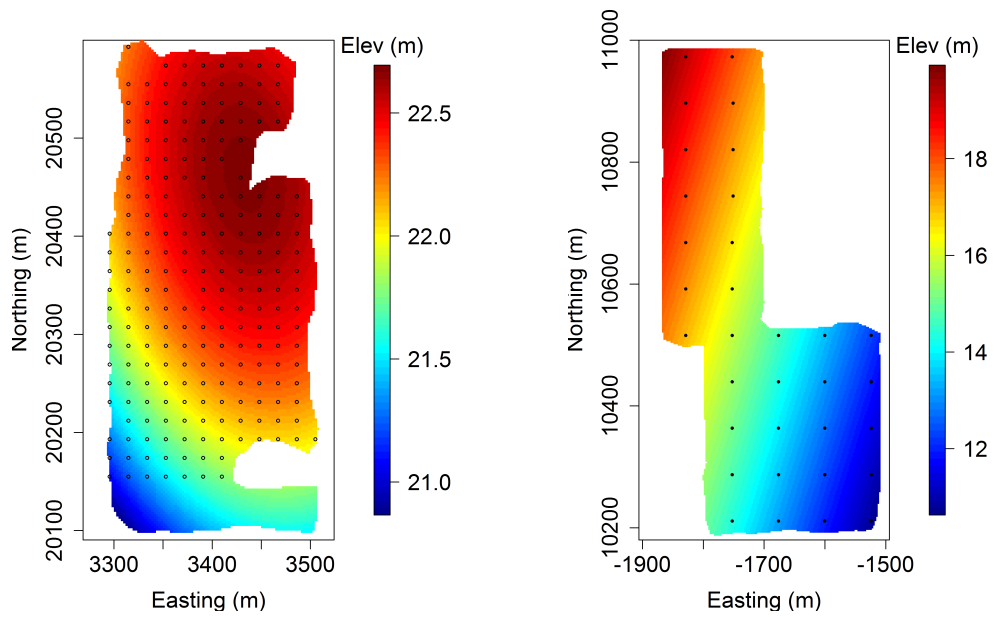


Figure A.2: Variables used to simulate the contact topography at Oak mine area: (a) Oak GPR data, (b) Oak borehole data overlain the GPR map (c) floor survey variable of the TI (Kumbur Mine) and (d) GPR component of the TI (Kumbur mine)

## Appendix B

### Trend Surfaces and the Residuals



(a) Kumbur PCD trend (Second Order)      (b) Oak Borehole trend (First Order)

Figure B.1: Trends inferred using the conditioning data available for both Oak and Kumbur mine areas

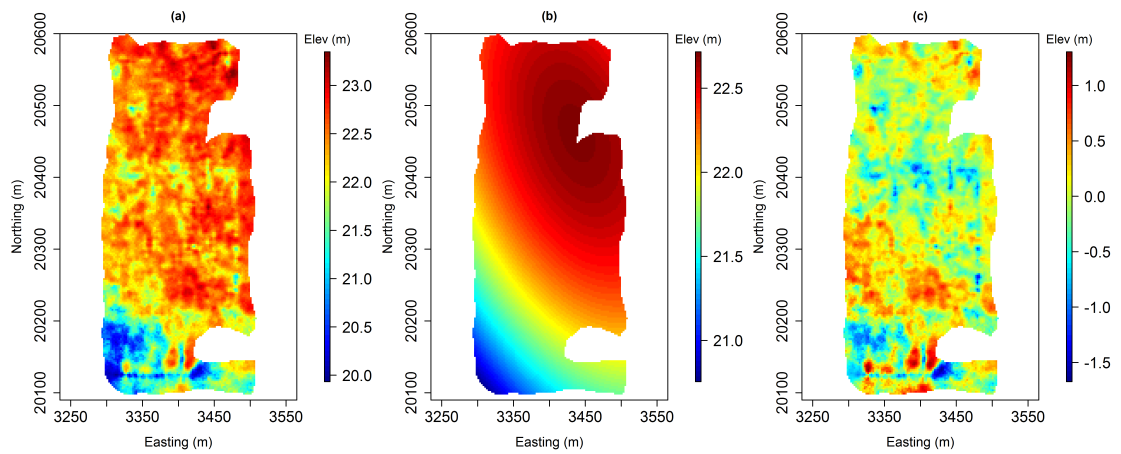


Figure B.2: Kumbur GPR trend and residuals: (1) Kumbur GPR data, (2) second order trend for the Kumbur GPR and (3) Kumbur GPR residuals obtained by subtracting the trend from the raw data

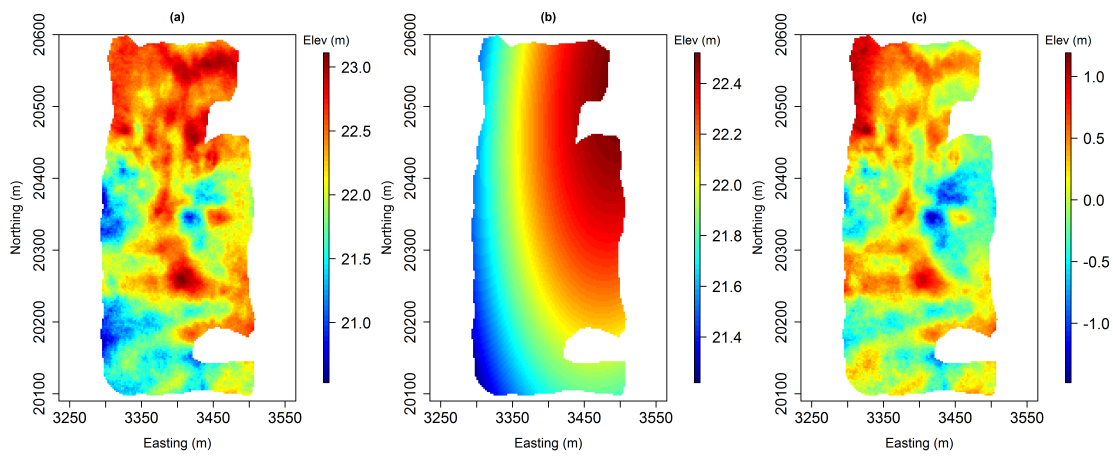


Figure B.3: Kumber floor trend and residuals: (a) Kumber mined-out floor survey, (b) second order trend fitted to the floor survey and (c) residuals of the floor survey

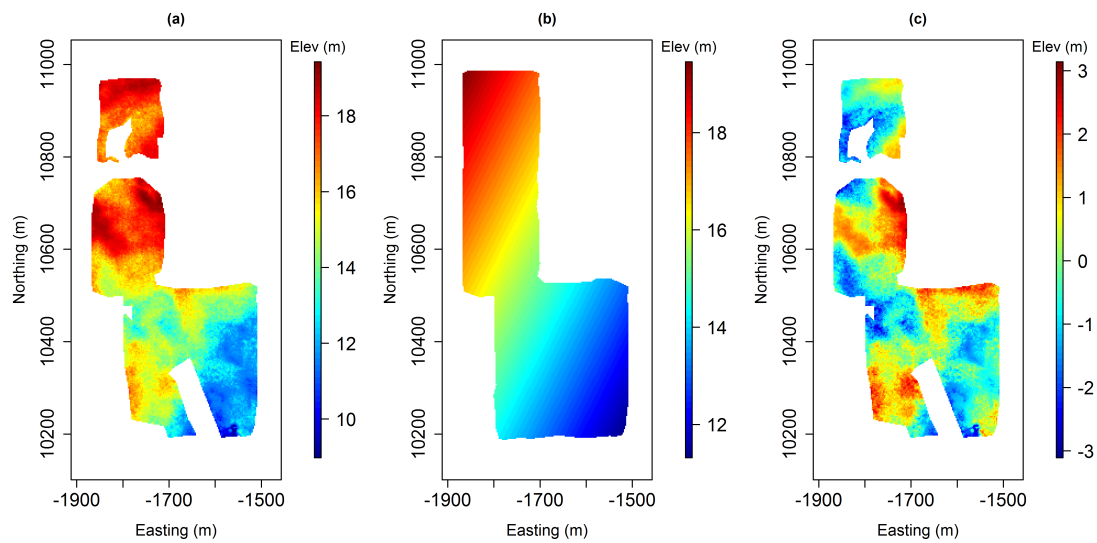


Figure B.4: Oak floor trend and residuals: (a) Oak mined-out floor survey, (b) first order trend fitted to the floor survey and (c) residuals of the floor survey



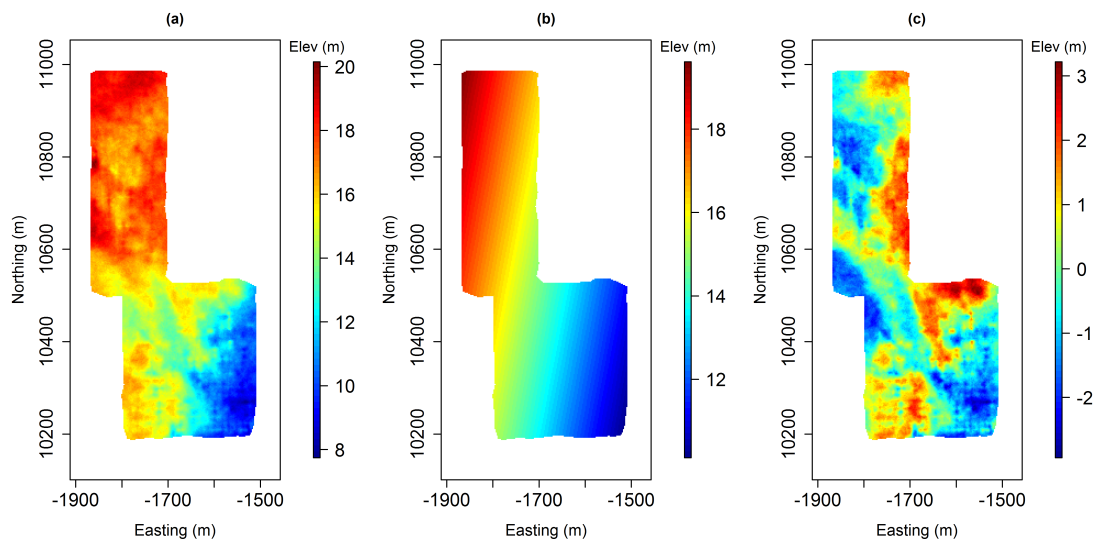


Figure B.5: Oak GPR trend and residuals: (a) Oak GPR data, (b) first order trend fitted to the Oak GPR data and (c) the residuals of the GPR data

# Appendix C

## Variogram Models Fitted in Trend Free Directions

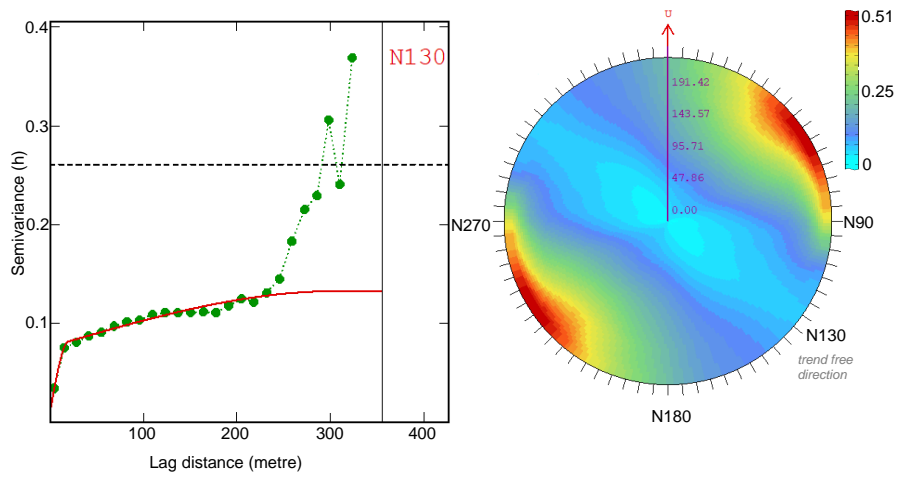


Figure C.1: Experimental variogram calculated along the trend-free direction (N130) and the fitted model for the Kumbur GPR variable

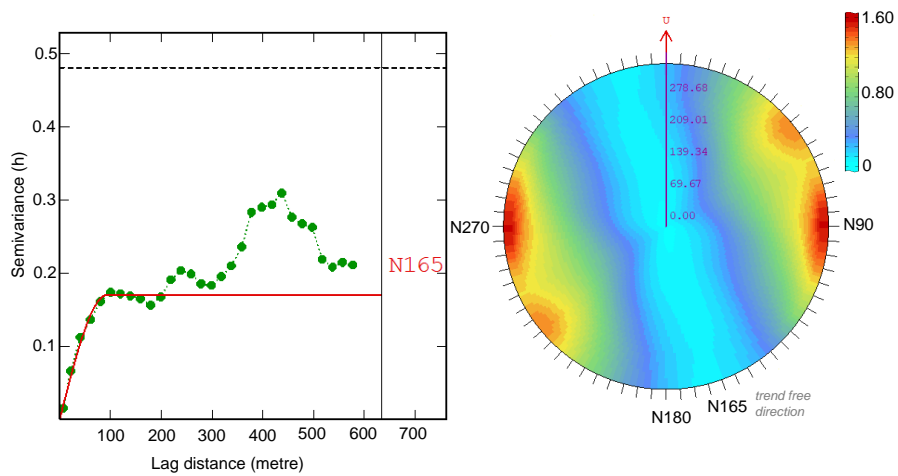


Figure C.2: Experimental variogram calculated along the trend-free direction (N165) and the fitted model for the Kumbur floor survey variable

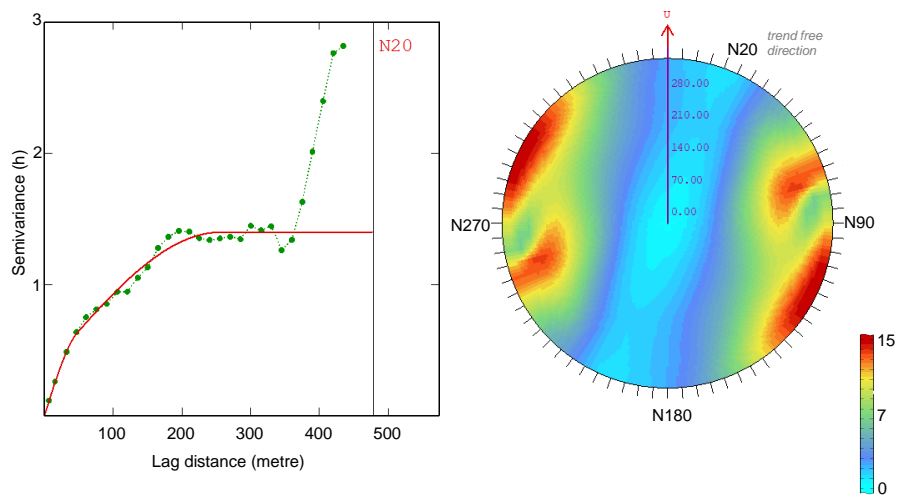


Figure C.3: Experimental variogram calculated along the trend-free direction (N20) and the fitted model for the Oak GPR variable

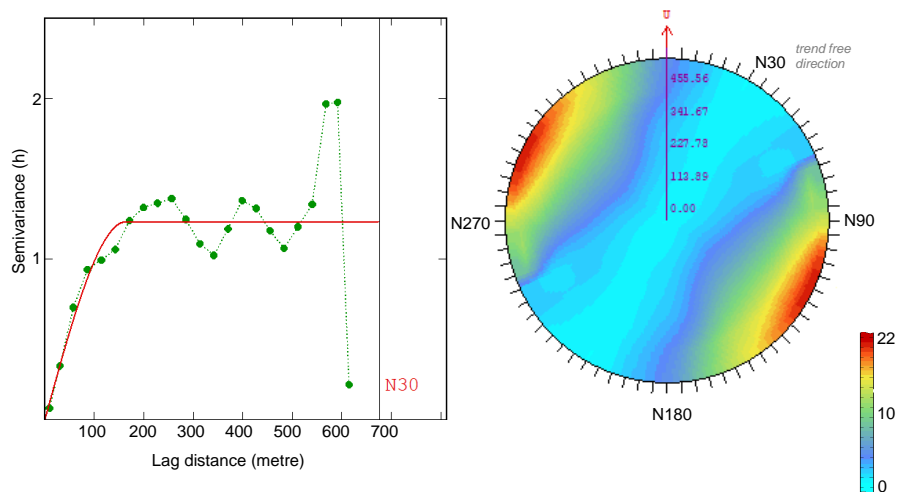


Figure C.4: Experimental variogram calculated along the trend-free direction (N30) and the fitted model for the Oak floor survey variable

Table C.1: Variogram models used to migrate the GPR and floor surveys into the grid nodes

Variable	Mine Area	Structure	Nugget	Range	Sill
GPR Survey	Oak	Spherical	-	47.85	0.338
		Spherical	-	255.54	1.058
	Kumbur	Spherical	0.001	19.33	0.065
		Spherical	-	306.05	0.056
Floor Survey	Oak	Spherical	-	164	1.232
	Kumbur	Spherical	-	73.32	0.134

# Appendix D

## Residual Variograms and Models

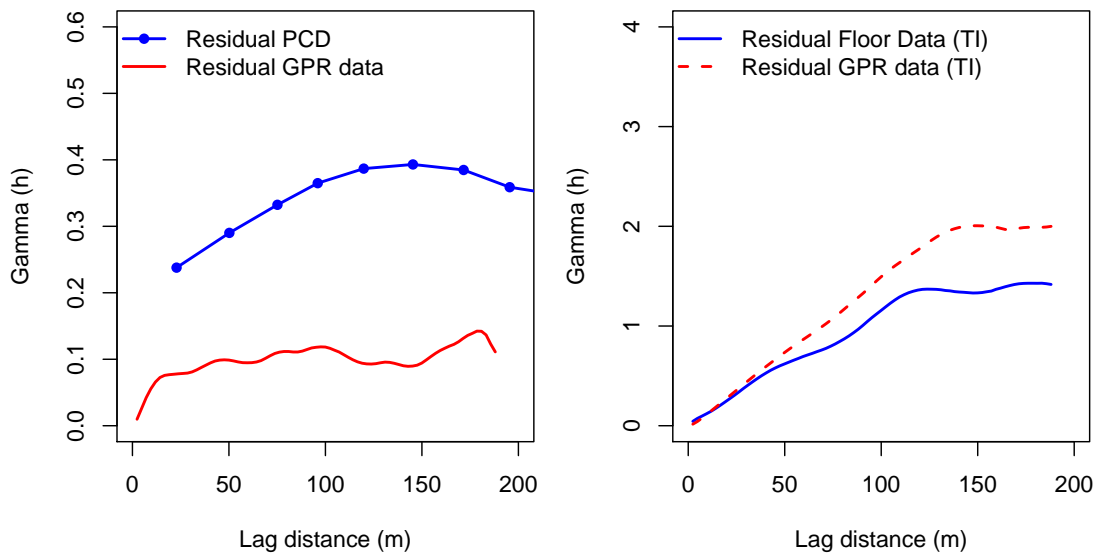


Figure D.1: Variograms of the variables used to simulate the Kumbur mine (residuals)

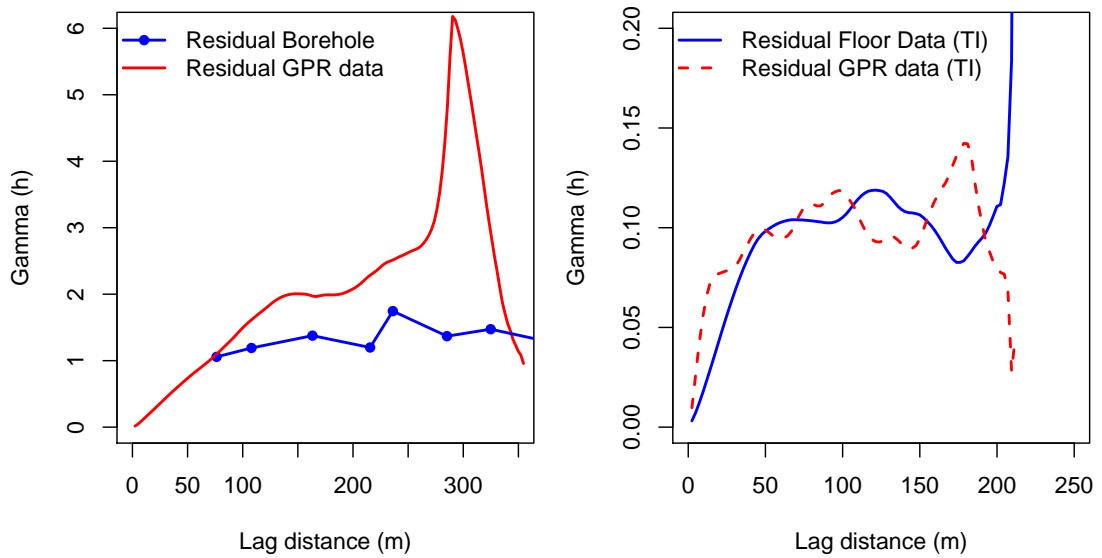


Figure D.2: Variograms of the variables used to simulate the Oak mine (residuals)

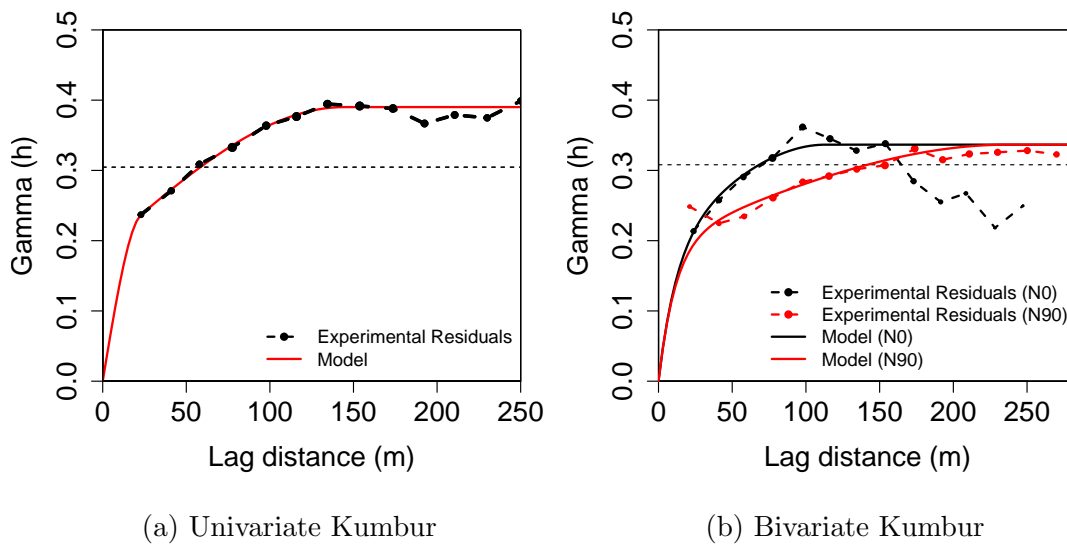


Figure D.3: Variogram models fitted to the PCD residuals for the univariate and bivariate Kumbur TB simulations. The difference in the experimental variograms is due to the differences in trends inferred using the borehole data for the univariate simulations and the GPR data (external drift) for the bivariate simulations. No anisotropy was detected for the univariate case. Therefore, only the omnidirectional variogram was modelled



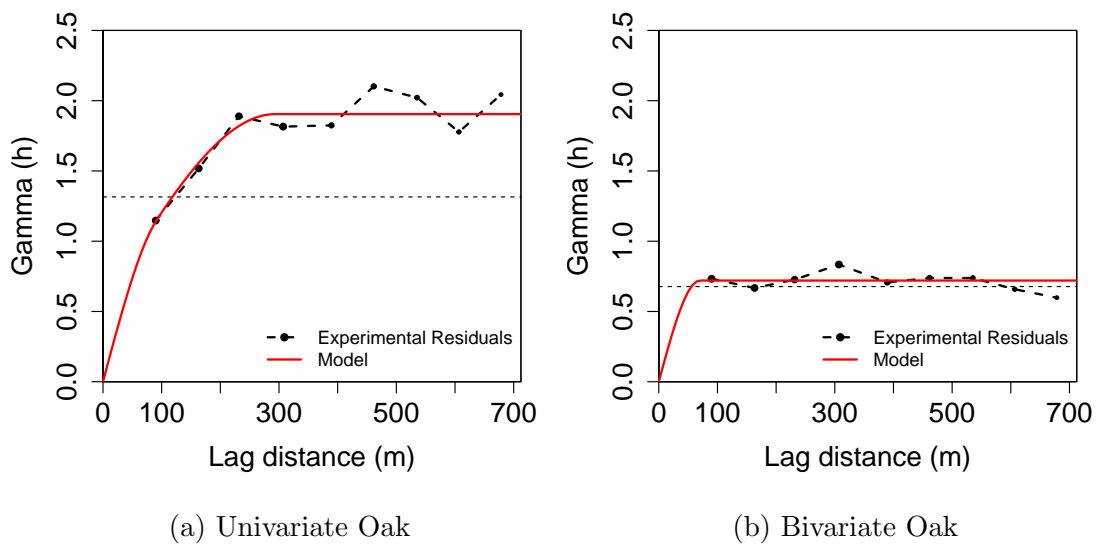


Figure D.4: Variogram models fitted to the borehole residuals for the univariate and bivariate Oak TB simulations. As in Kumbur mine variables, the trend for the univariate simulations was inferred from the borehole and the trend for the bivariate simulations was inferred from the GPR as an external drift

# Appendix E

## Kumbur Simulation Results

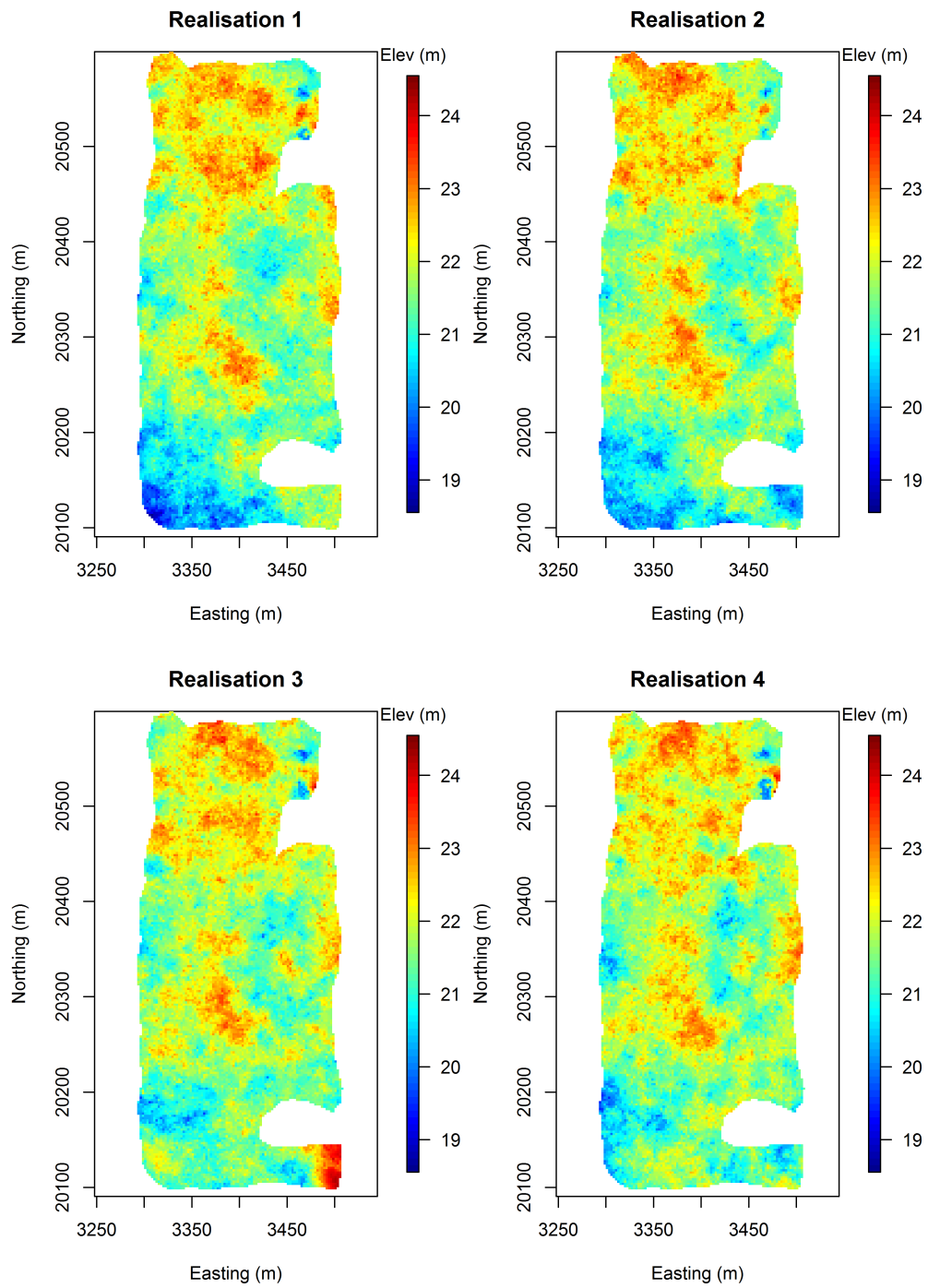


Figure E.1: First four DS realisations of the Kumbur univariate simulations

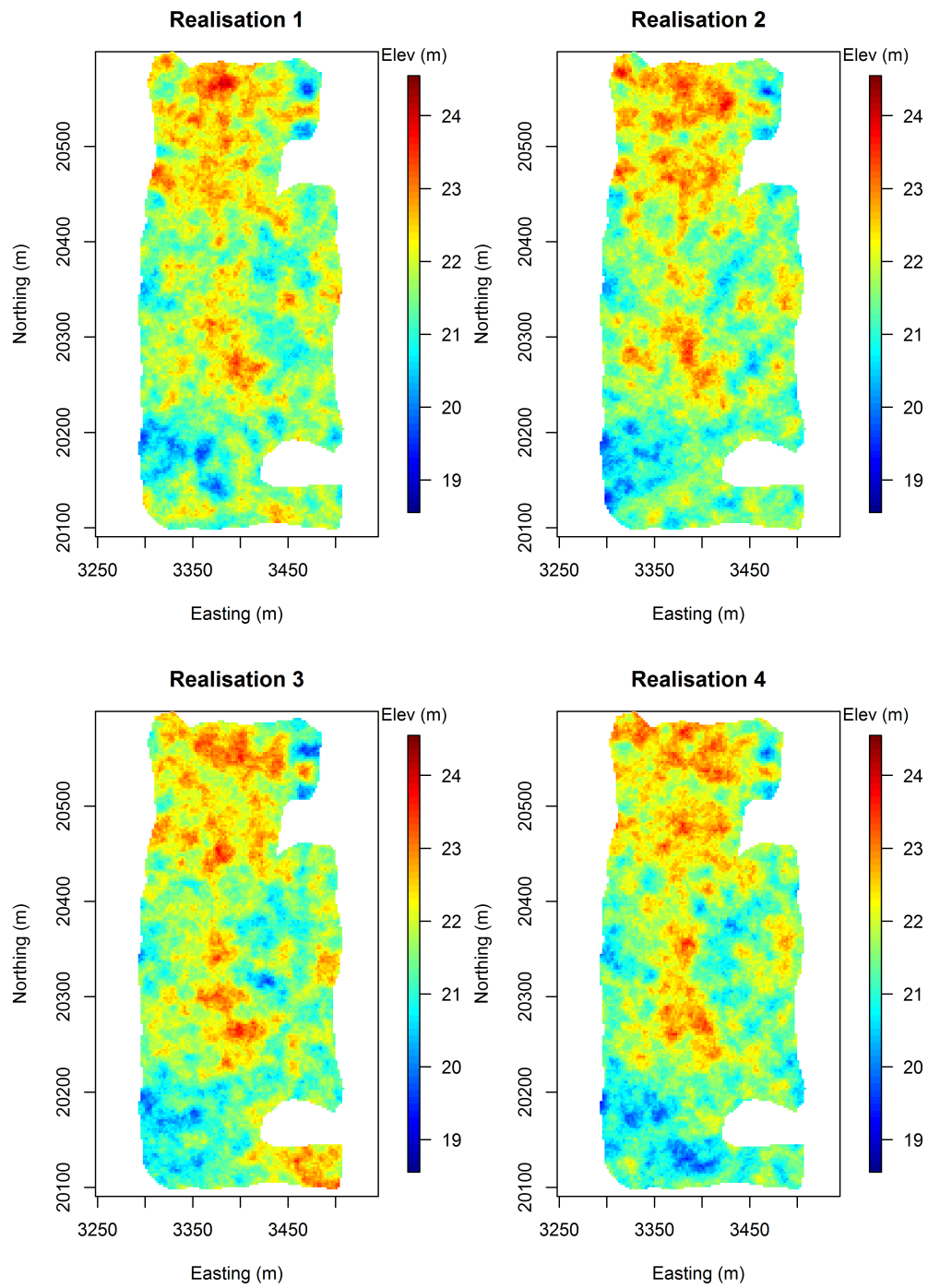


Figure E.2: First four TB realisations of the Kumber univariate simulations

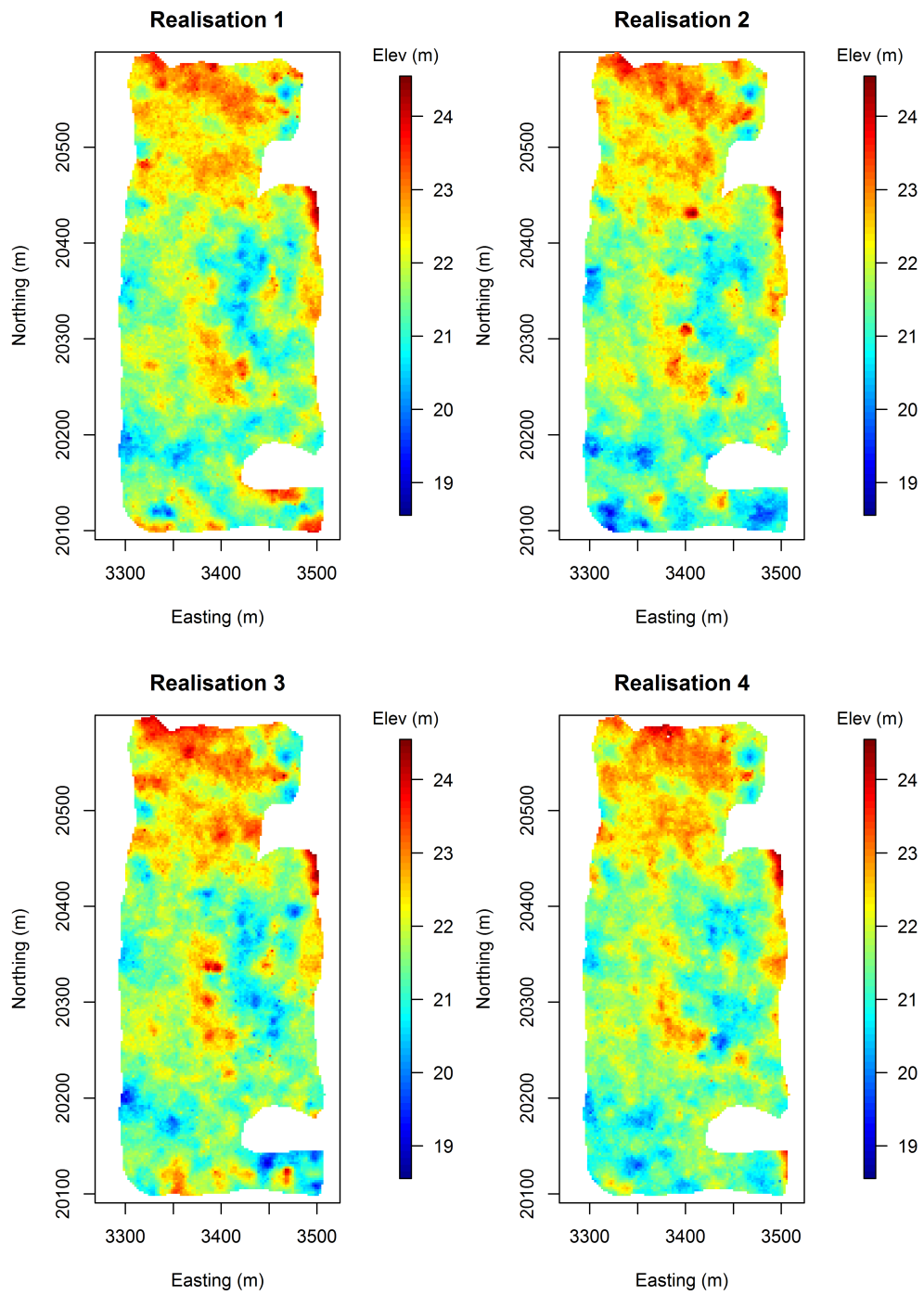


Figure E.3: First four DS realisations of the Kumbur bivariate simulations

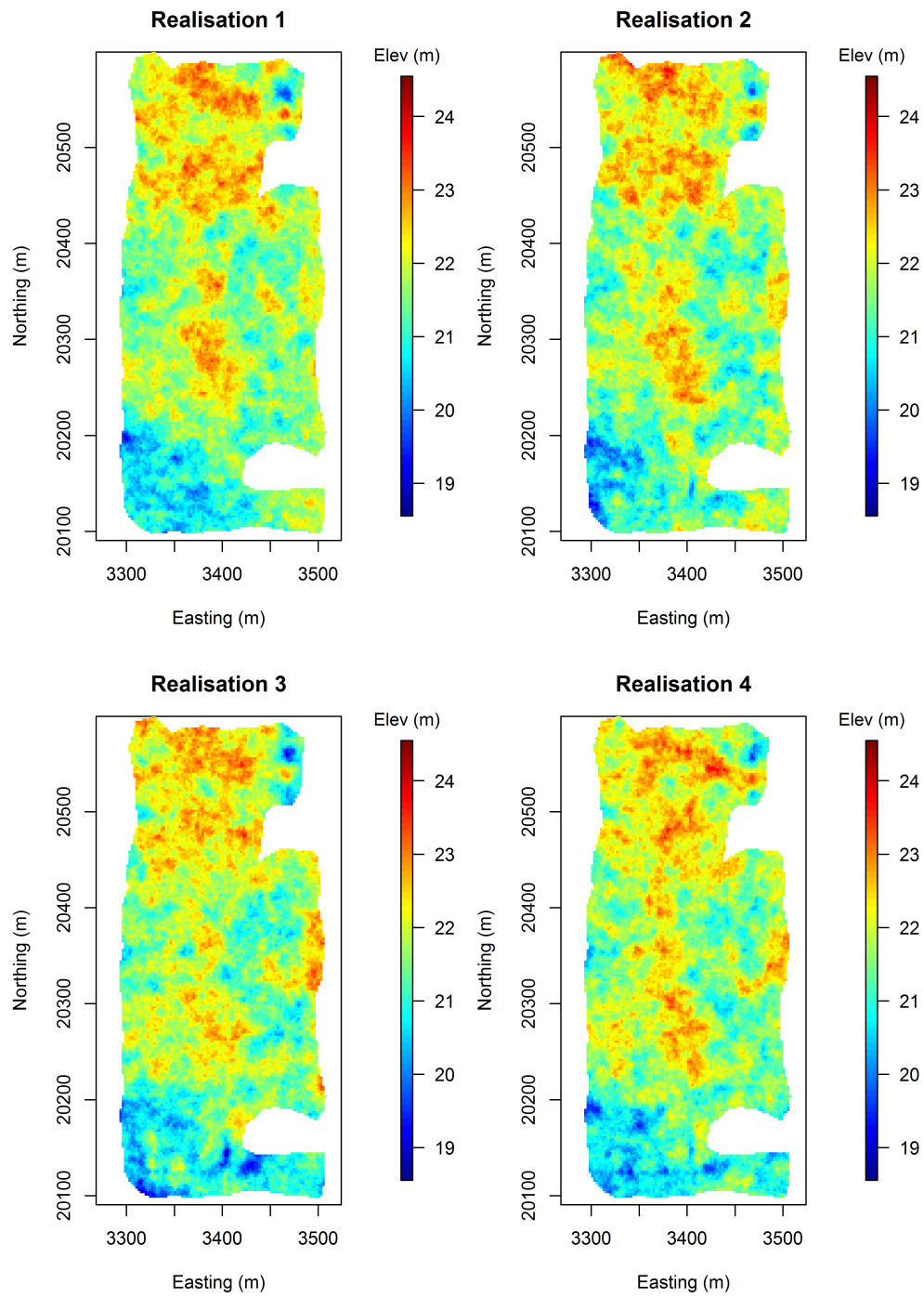


Figure E.4: First four TB realisations of the Kumbur bivariate simulations

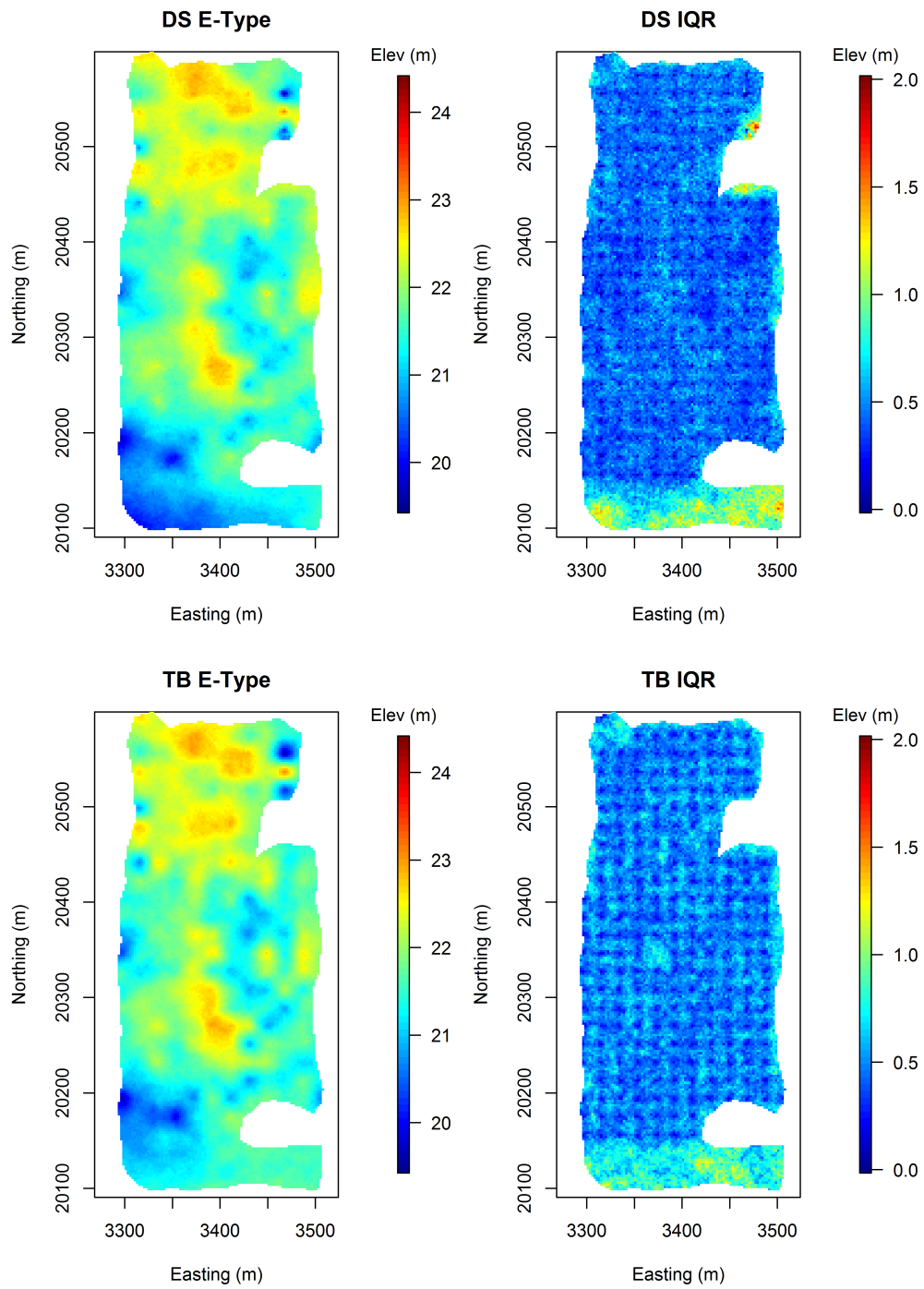


Figure E.5: Average of 40 univariate Kumbur realisations and the IQR maps

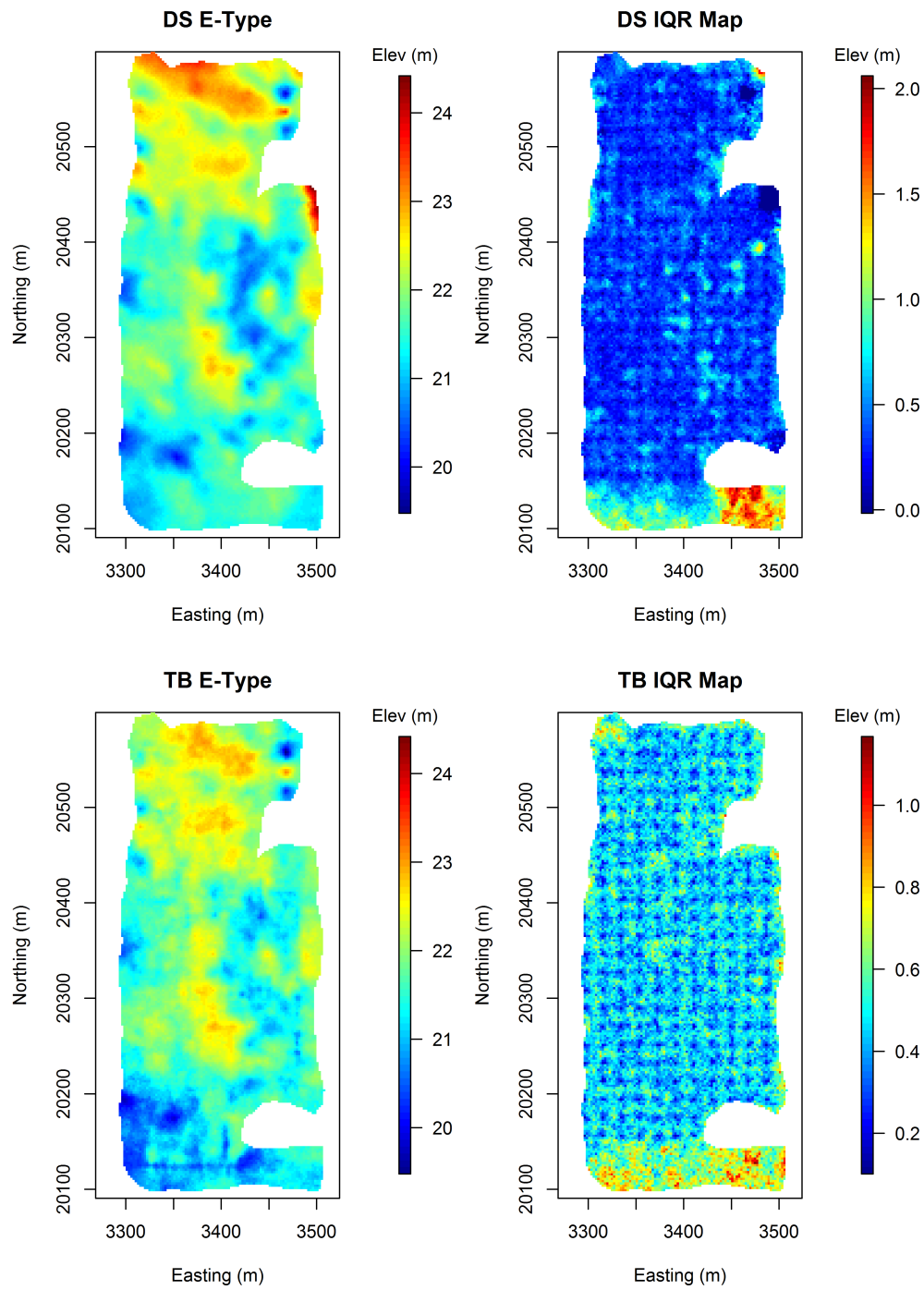


Figure E.6: Average of 40 bivariate Kumbur realisations and the IQR maps



# Appendix F

## Oak Simulation Results

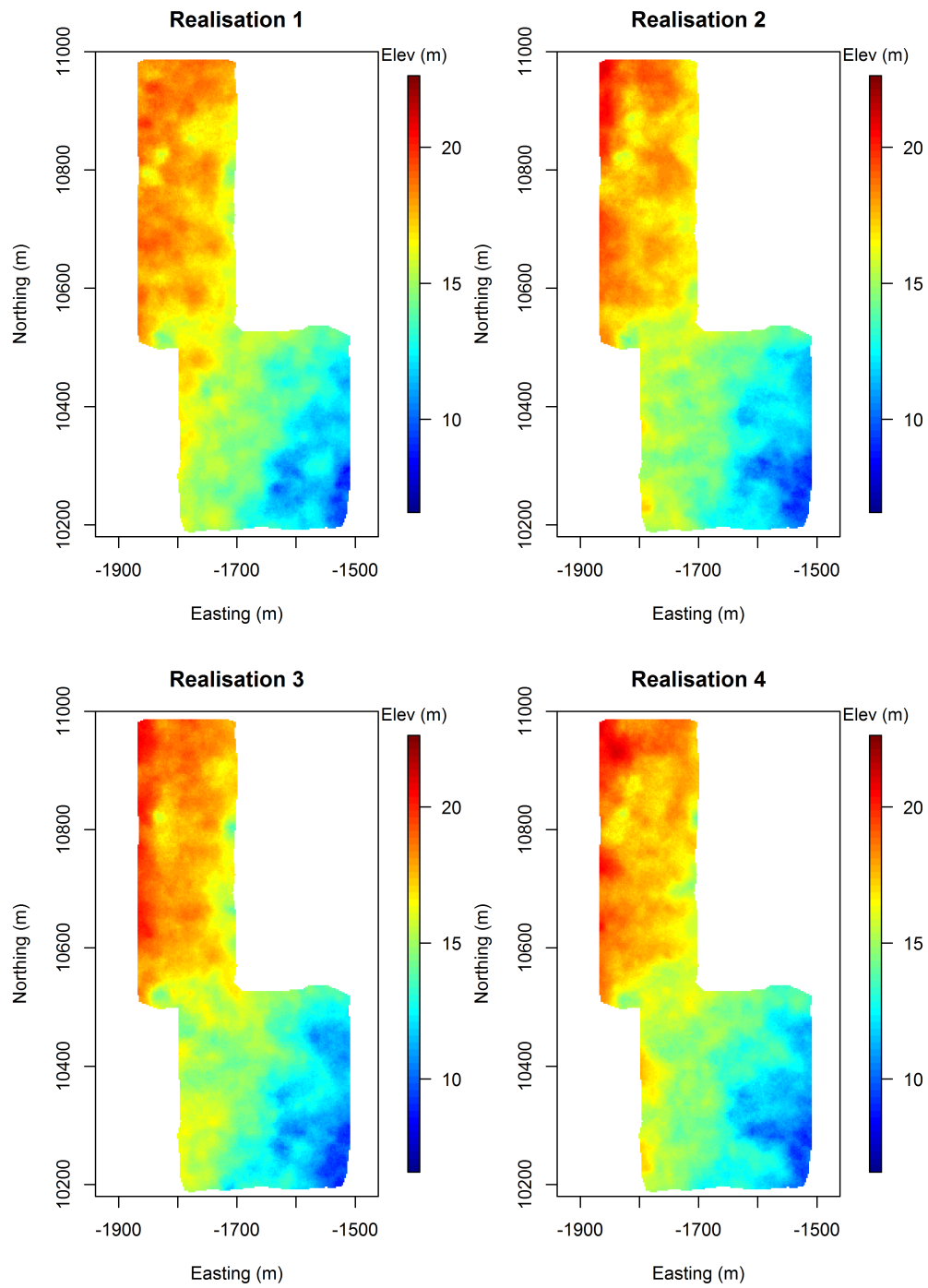


Figure F.1: First four DS realisations of the Oak univariate simulations

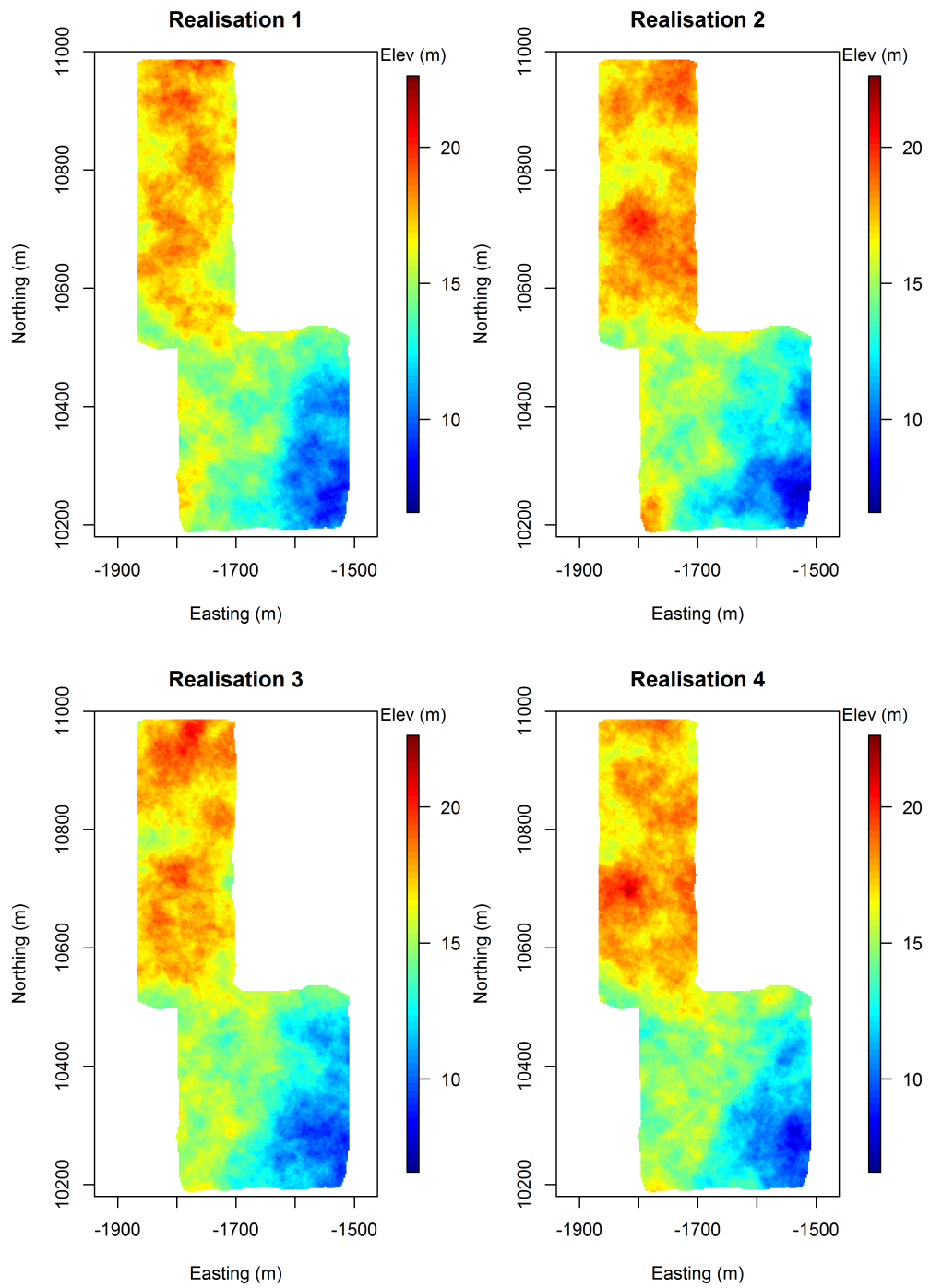


Figure F.2: First four TB realisations of the Oak univariate simulations

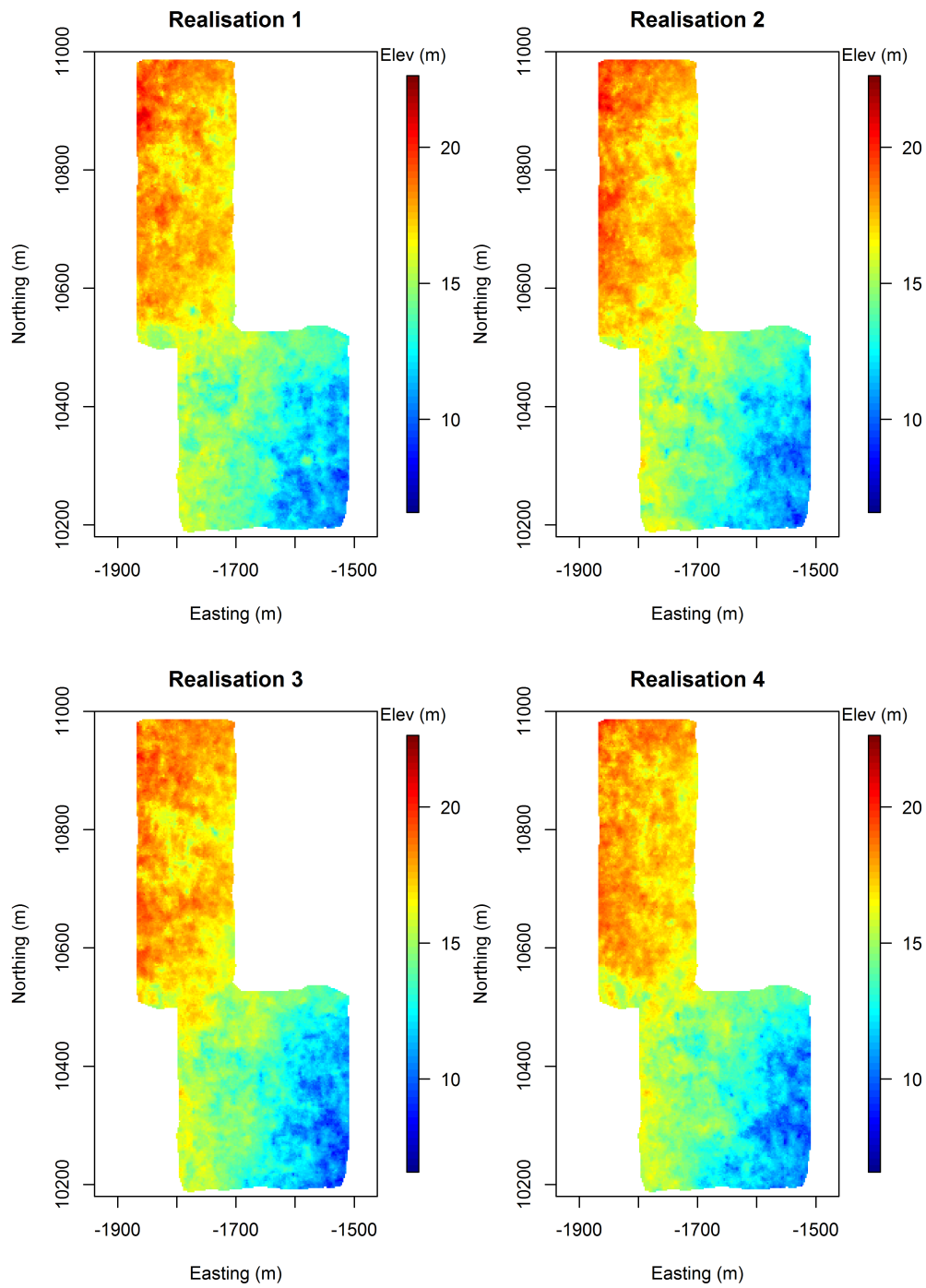


Figure F.3: First four DS realisations of the Oak bivariate simulations

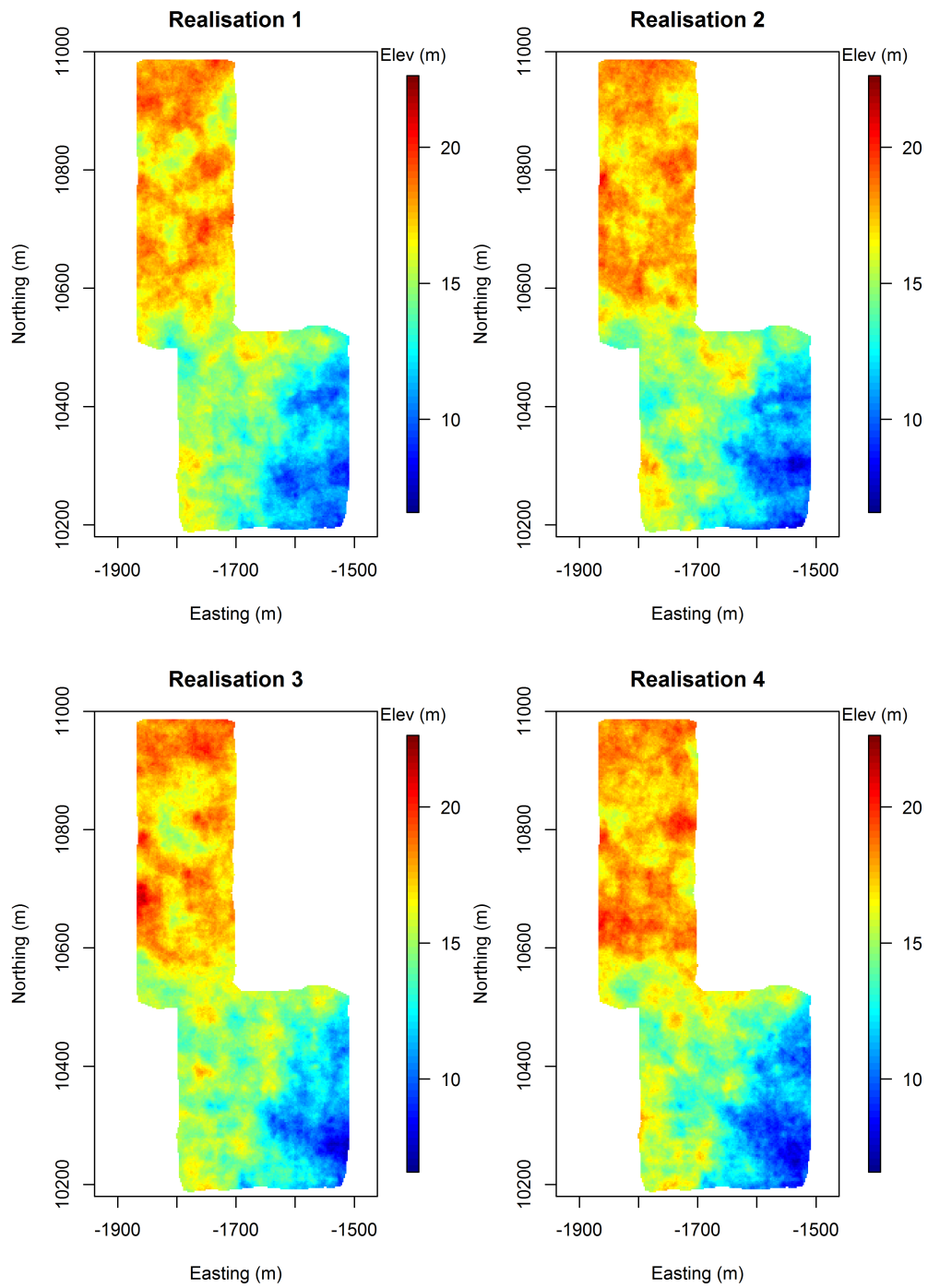


Figure F.4: First four TB realisations of the Oak bivariate simulations

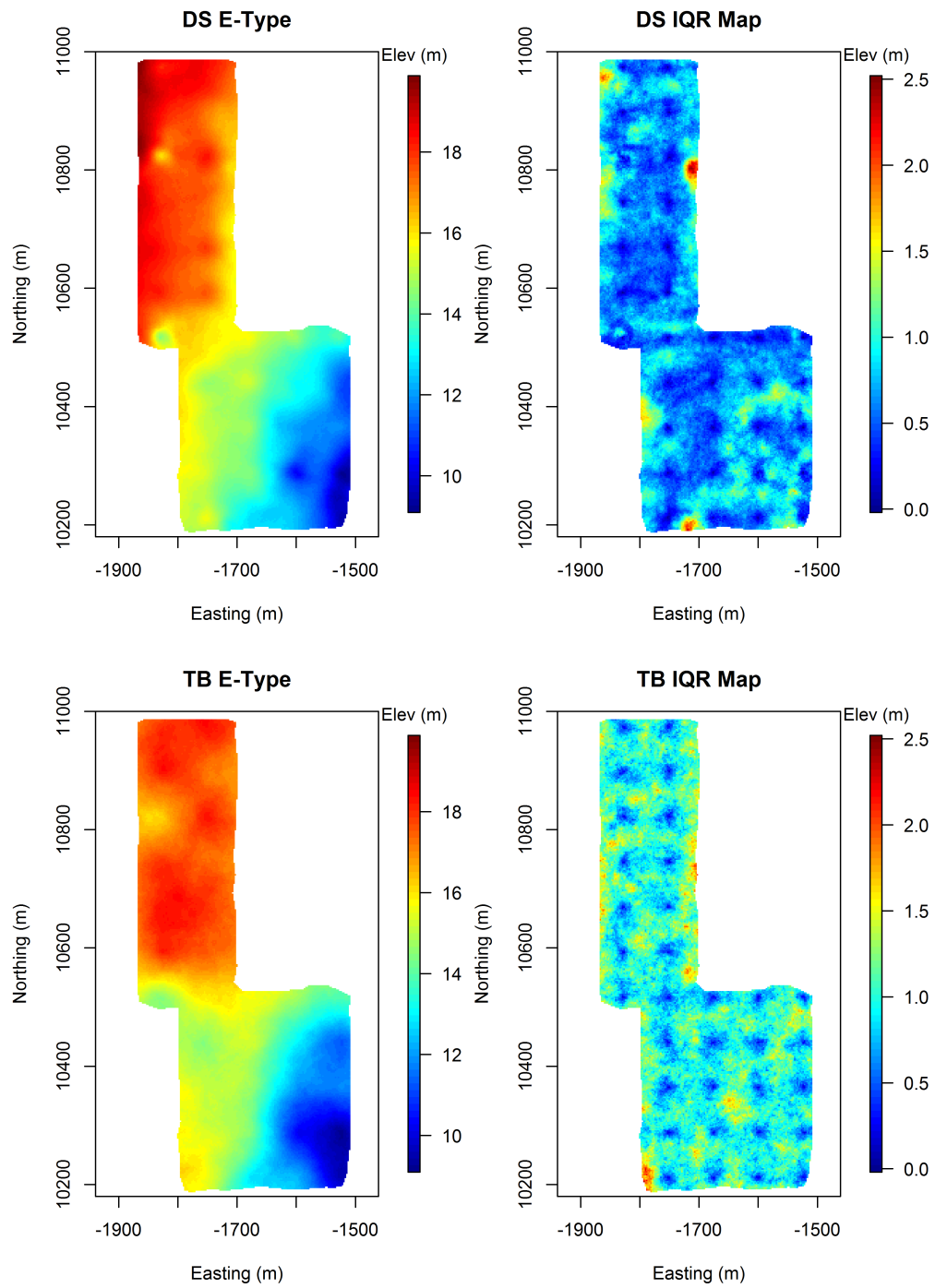


Figure F.5: Average of 40 univariate Oak realisations and the IQR maps

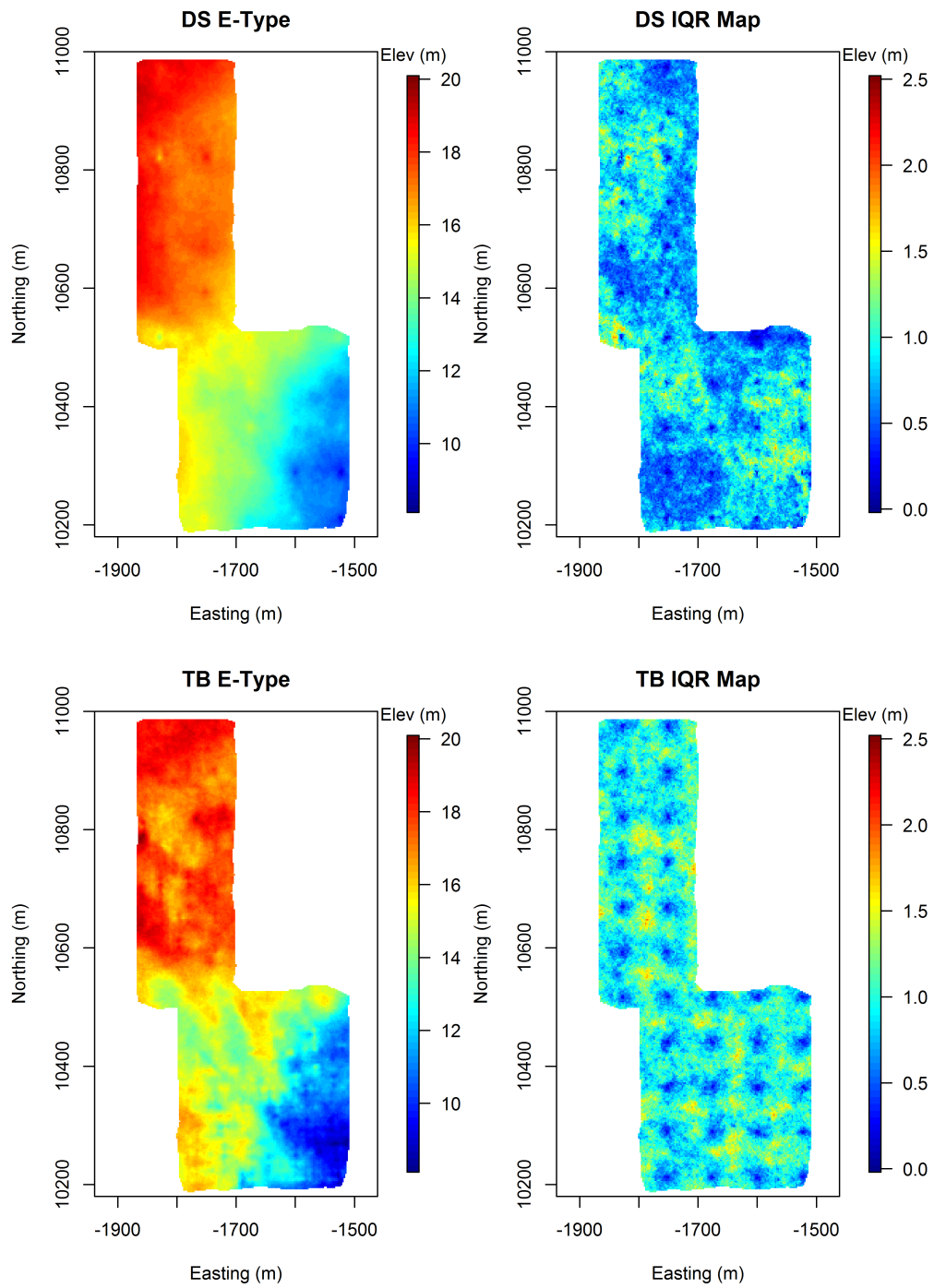
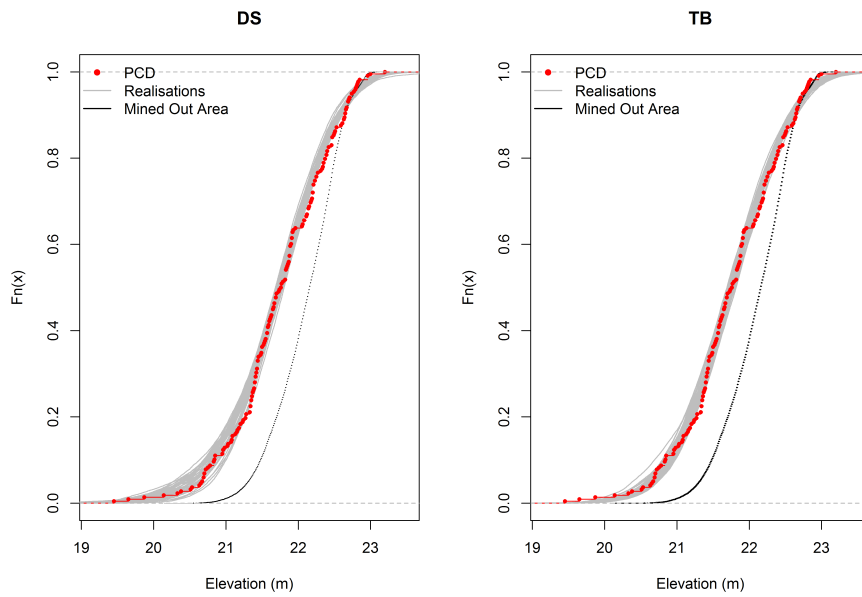


Figure F.6: Average of 40 bivariate Oak realisations and the IQR maps

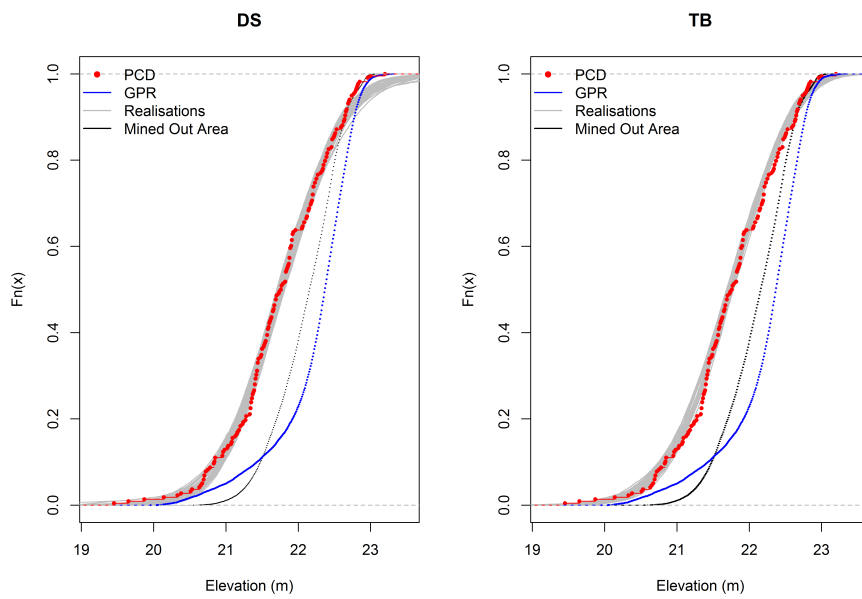
# Appendix G

## Statistical and Structural Analyses Results



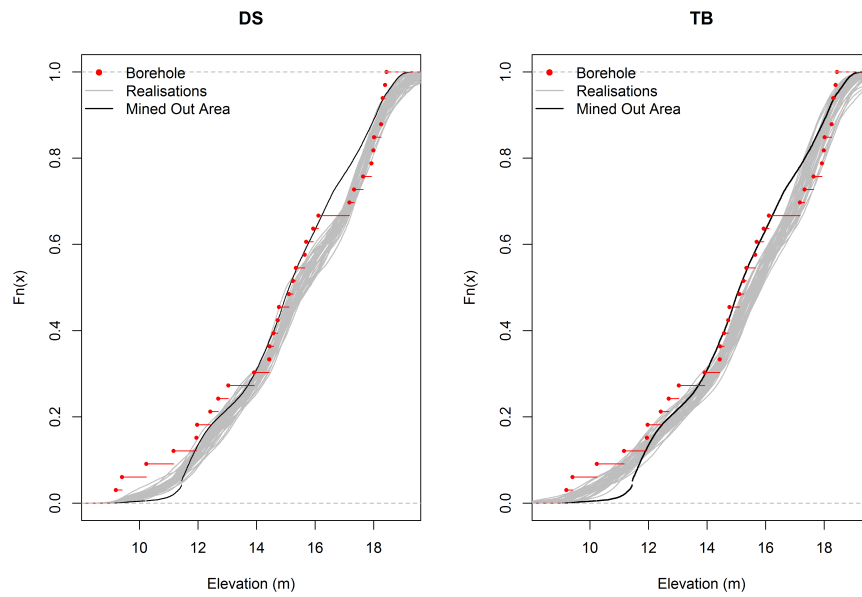


(a) Univariate Kumbur

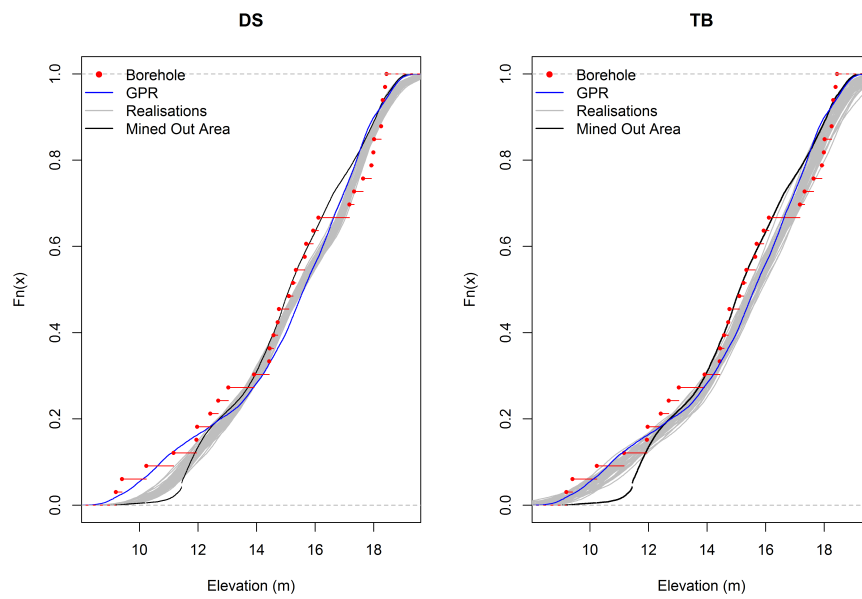


(b) Bivariate Kumbur

Figure G.1: Reproduction of the experimental quantiles for Kumbur simulations

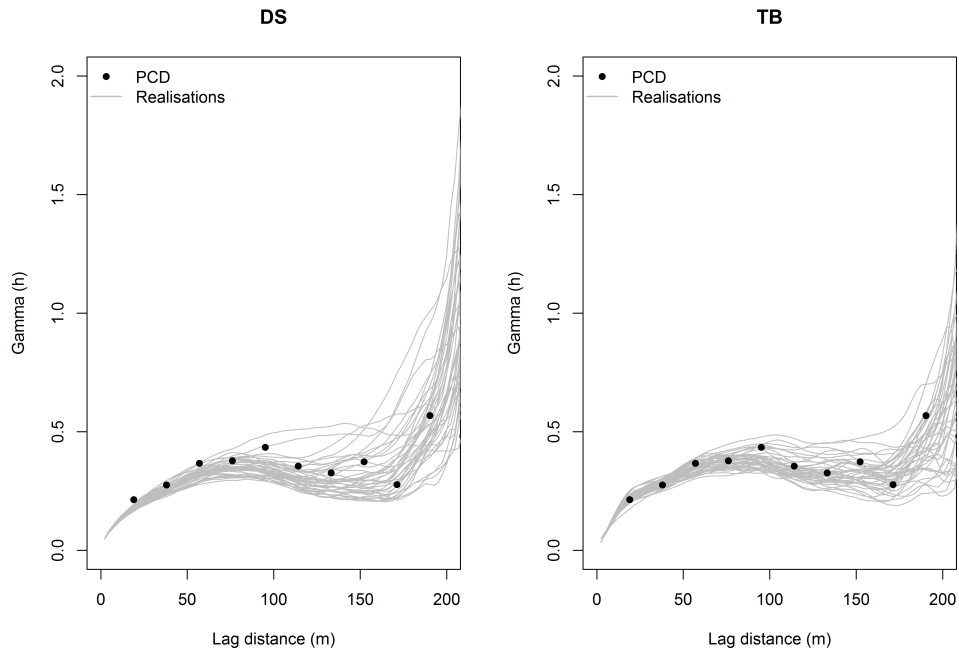


(a) Univariate Oak

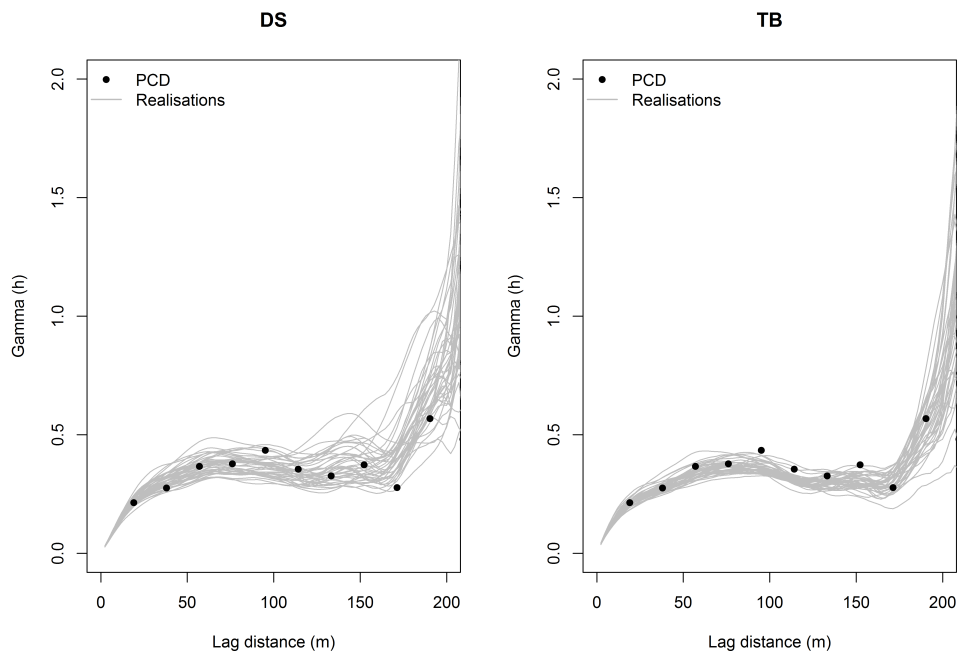


(b) Bivariate Oak

Figure G.2: Reproduction of the experimental quantiles for Oak simulations

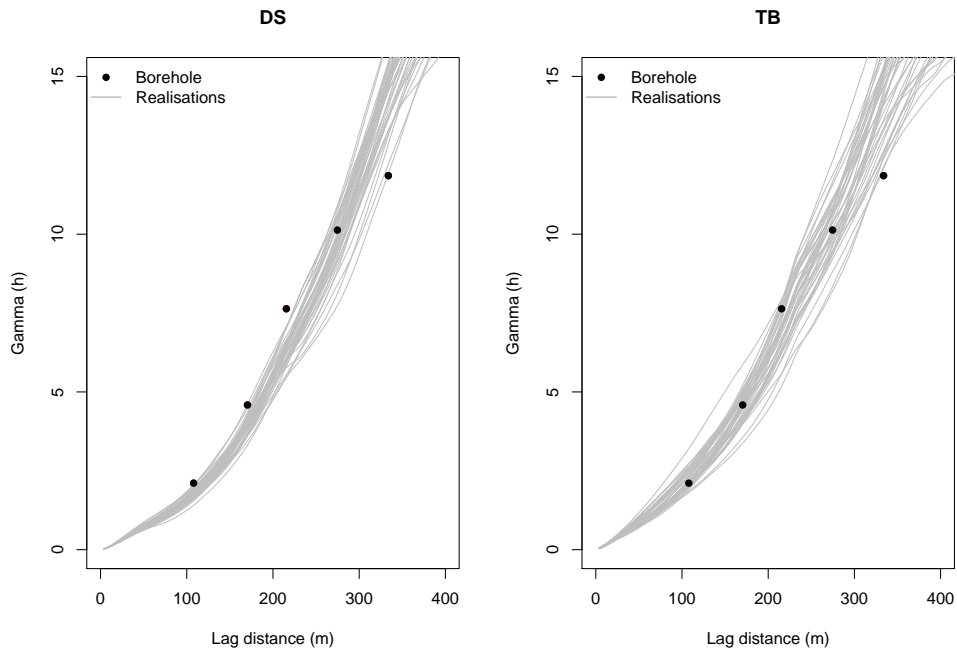


(a) Univariate Kumbur

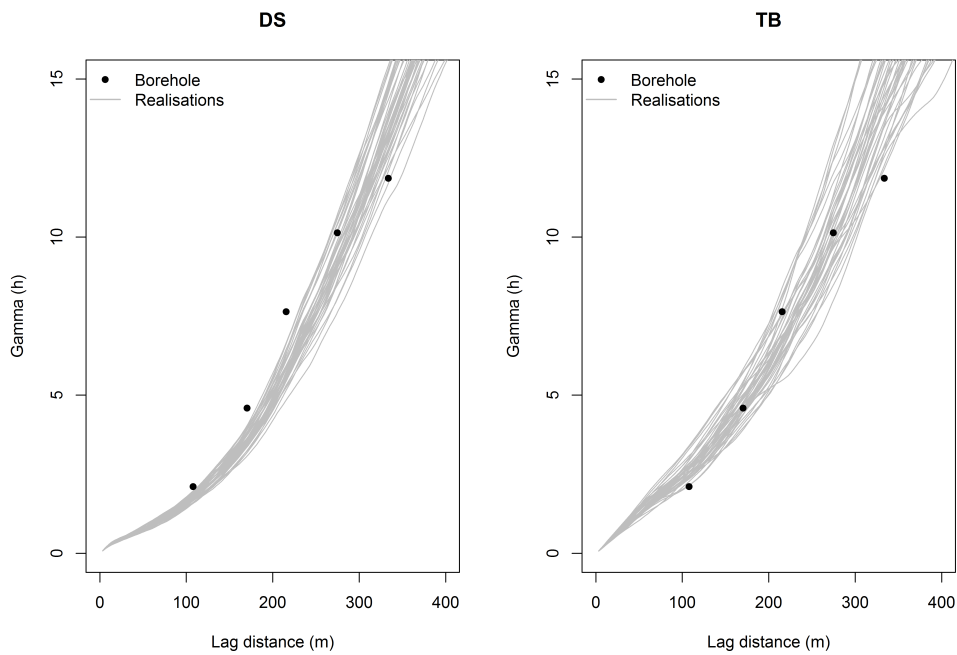


(b) Bivariate Kumbur

Figure G.3: Reproduction of the variograms for the Kumbur mine simulations

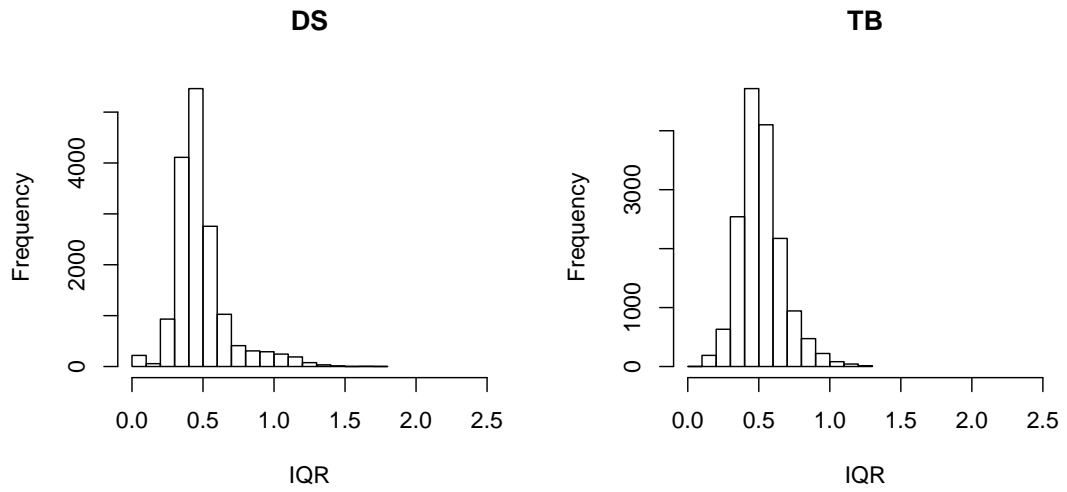


(a) Univariate Oak

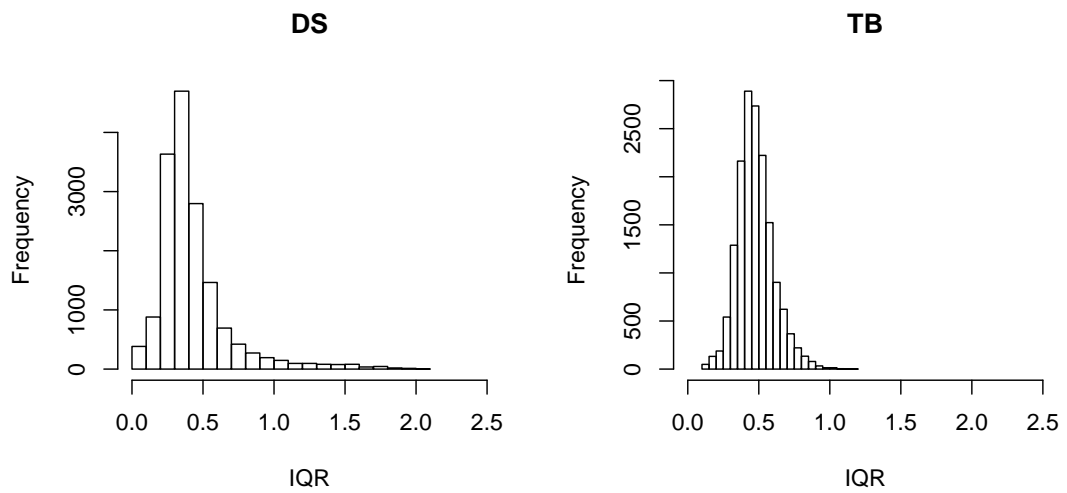


(b) Bivariate Oak

Figure G.4: Reproduction of the variograms for the Oak mine simulations

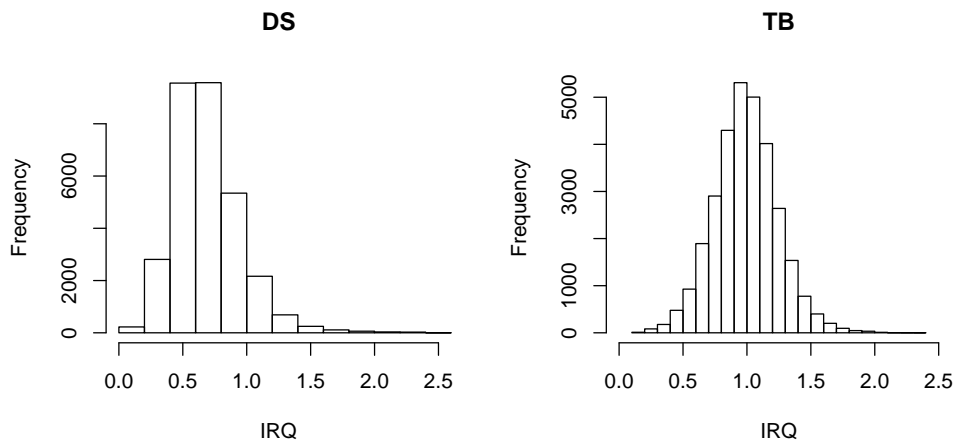


(a) Univariate Kumbur

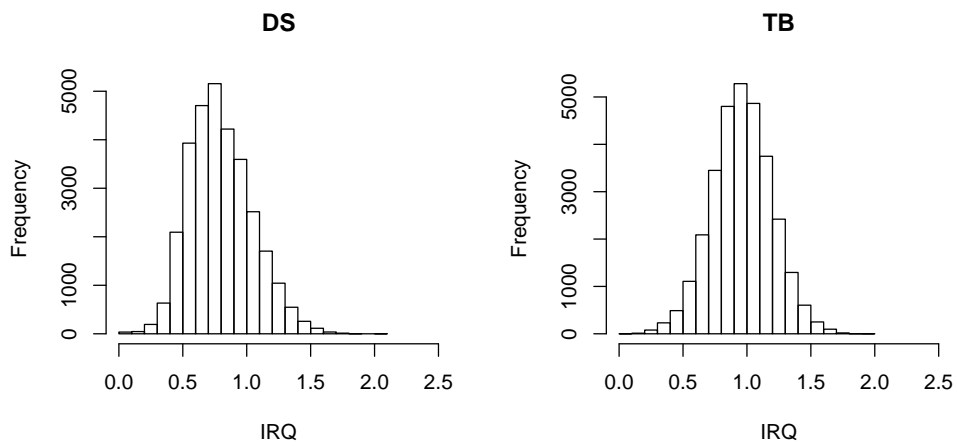


(b) Bivariate Kumbur

Figure G.5: Distribution of the interquartile ranges of the simulated elevation values at each grid node



(a) Univariate Oak



(b) Bivariate Oak

Figure G.6: Distribution of the interquartile ranges of the simulated elevation values at each grid node

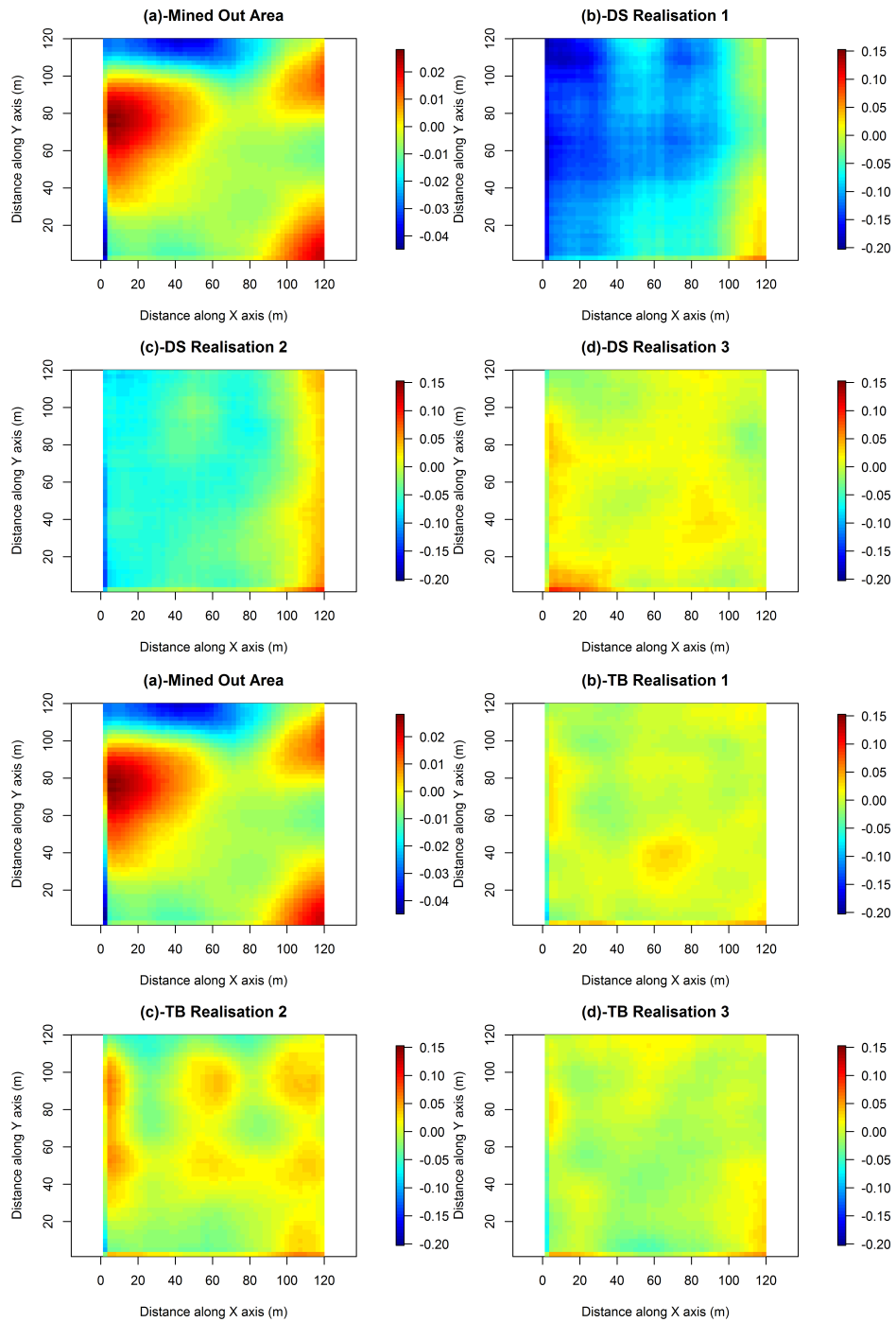


Figure G.7: Cumulant maps of the Kumbur mined-out floor (reference image) and the first 3 DS and TB univariate Kumbur realisations

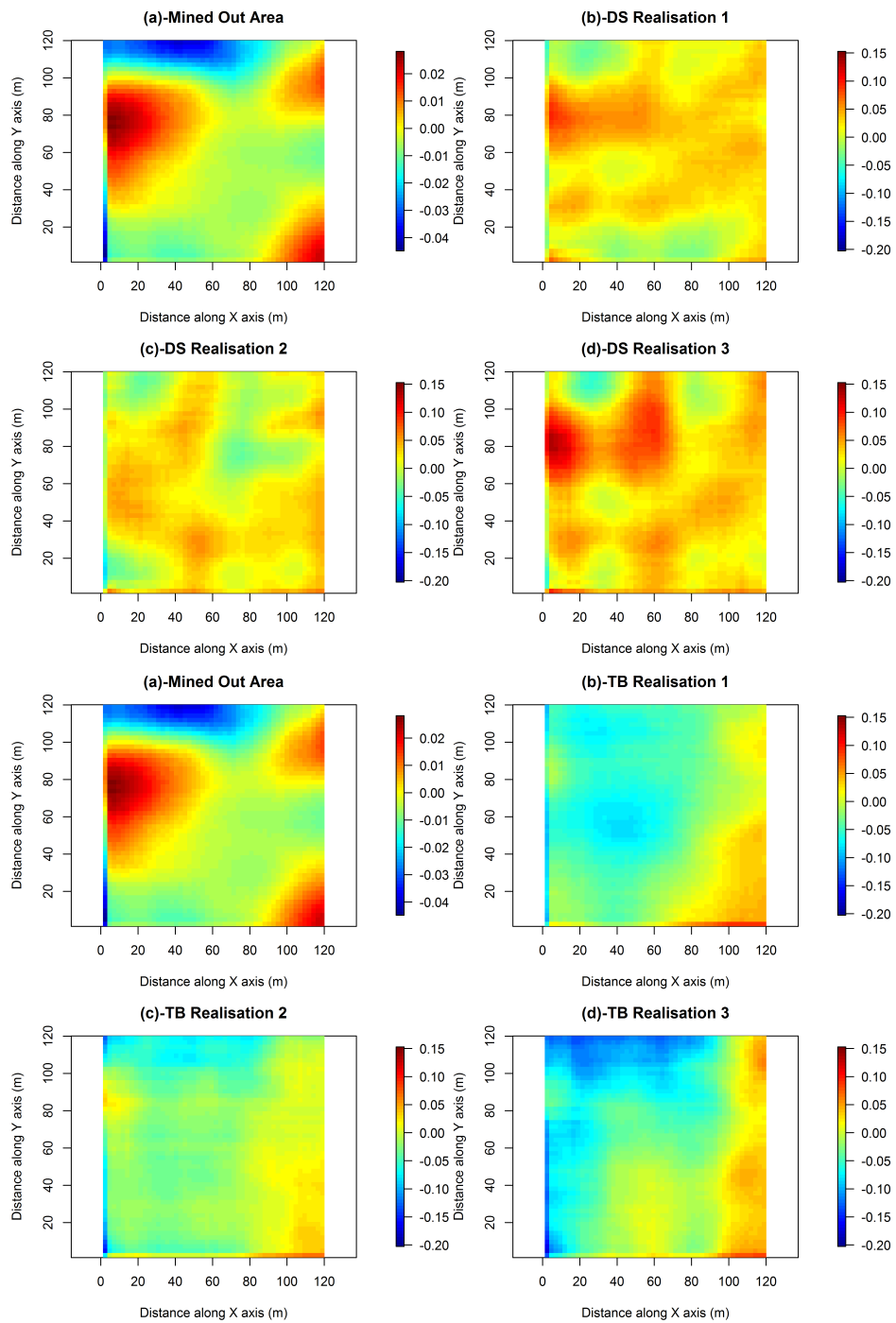


Figure G.8: Cumulant maps of the Kumbur mined-out floor (reference image) and the first 3 DS and TB bivariate Kumbur realisations



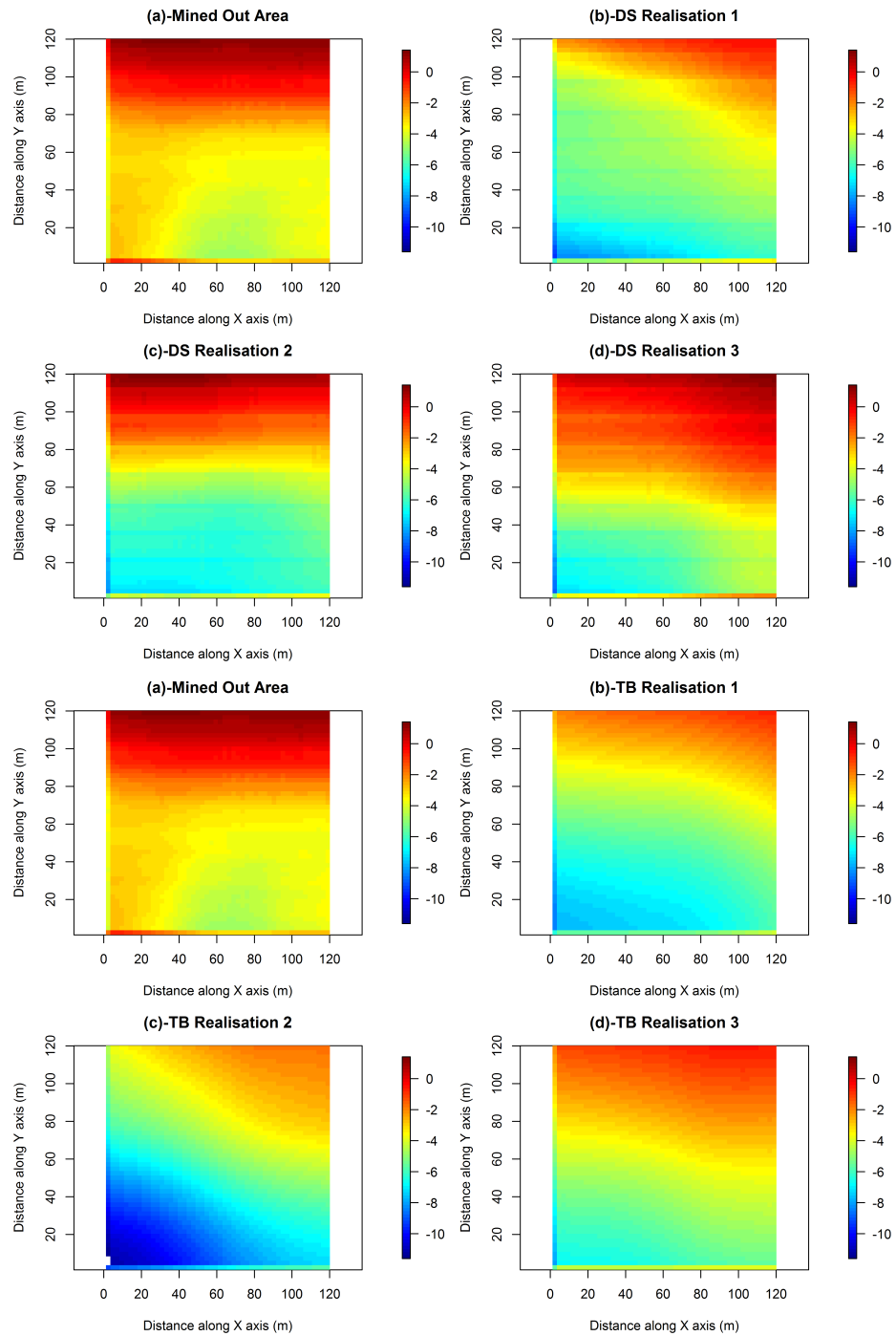


Figure G.9: Cumulant maps of the Oak mined-out floor (reference image) and the first 3 DS and TB univariate Oak realisations

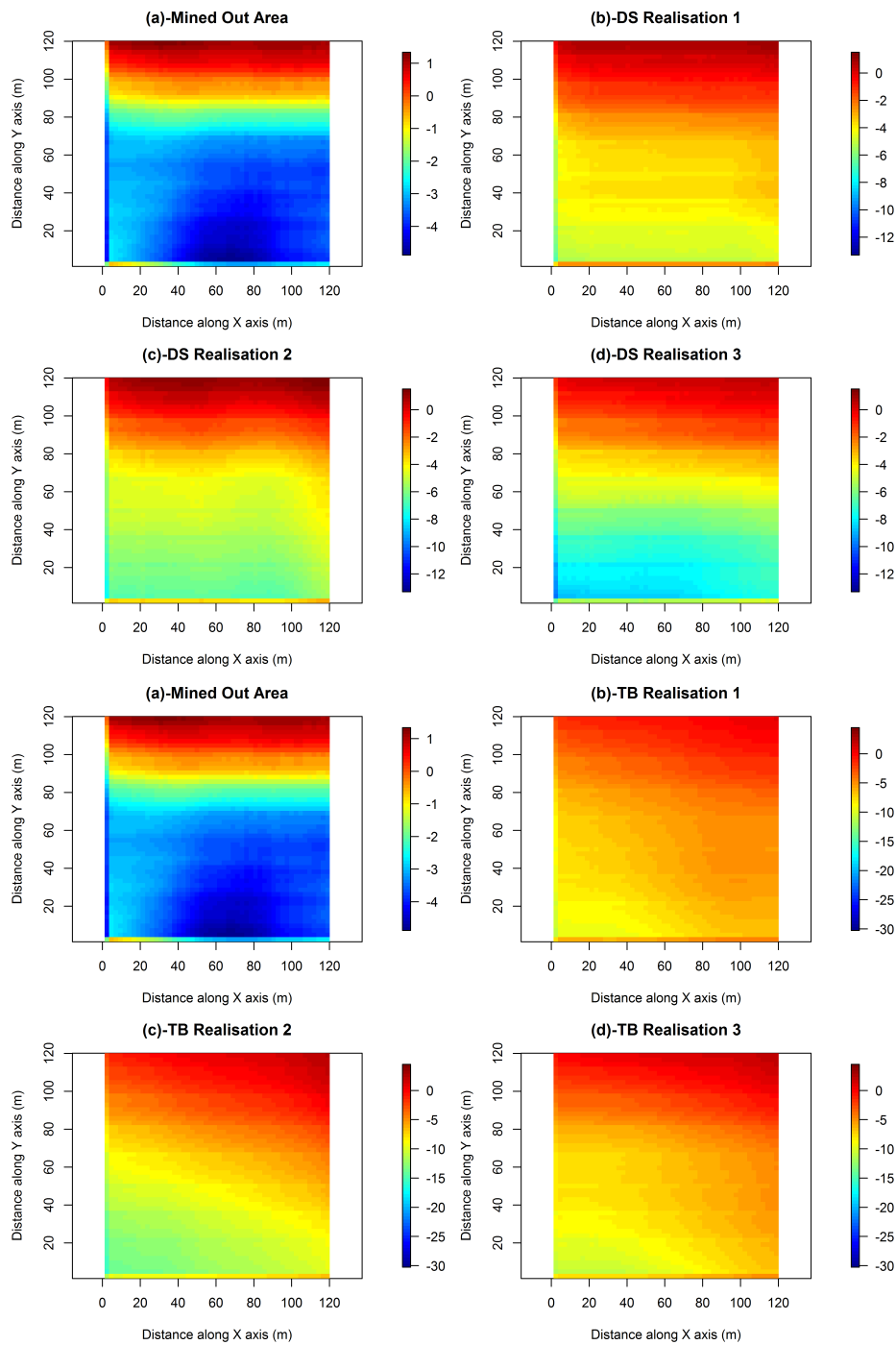
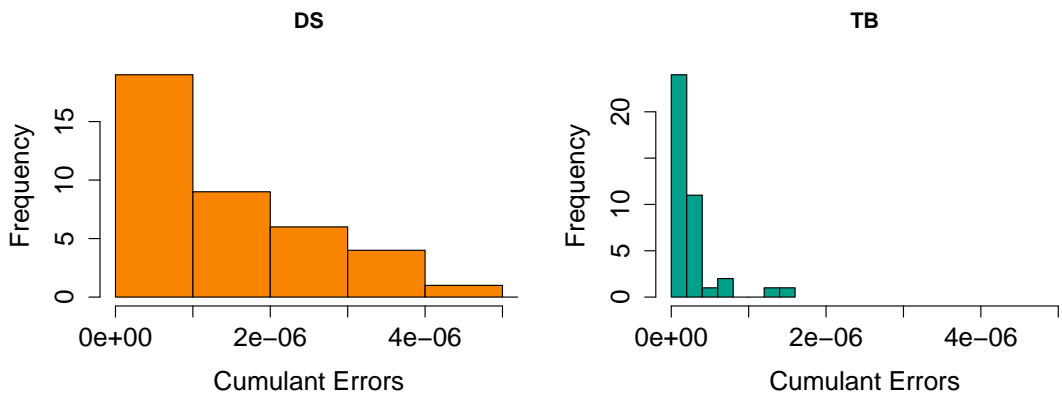
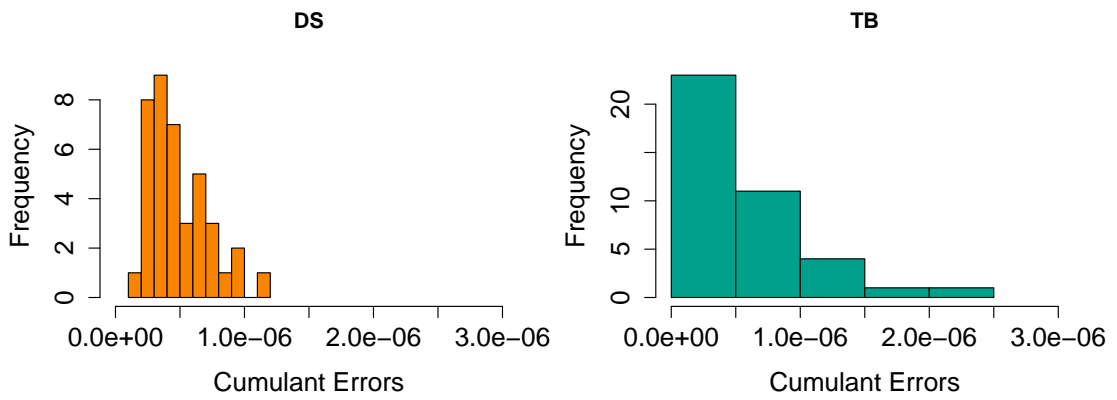


Figure G.10: Cumulant maps of the Oak mined-out floor (reference image) and the first 3 DS and TB bivariate Oak realisations

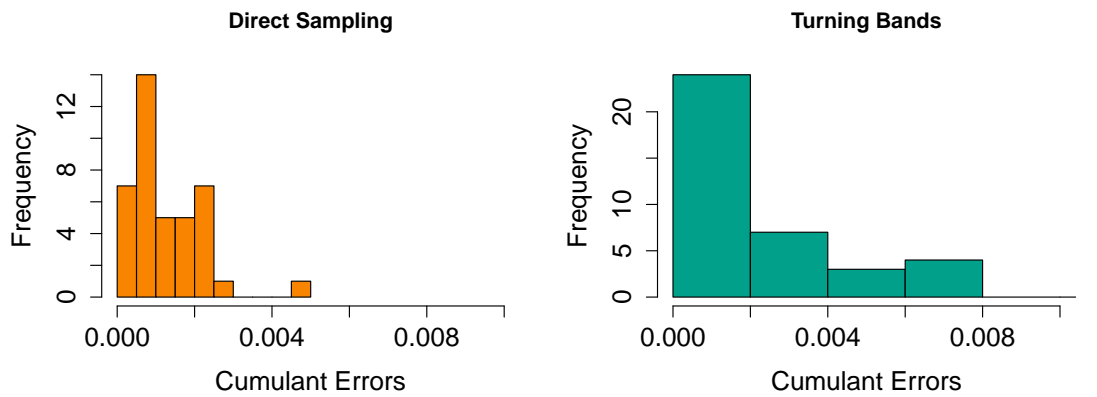


(a) Univariate Kumbur

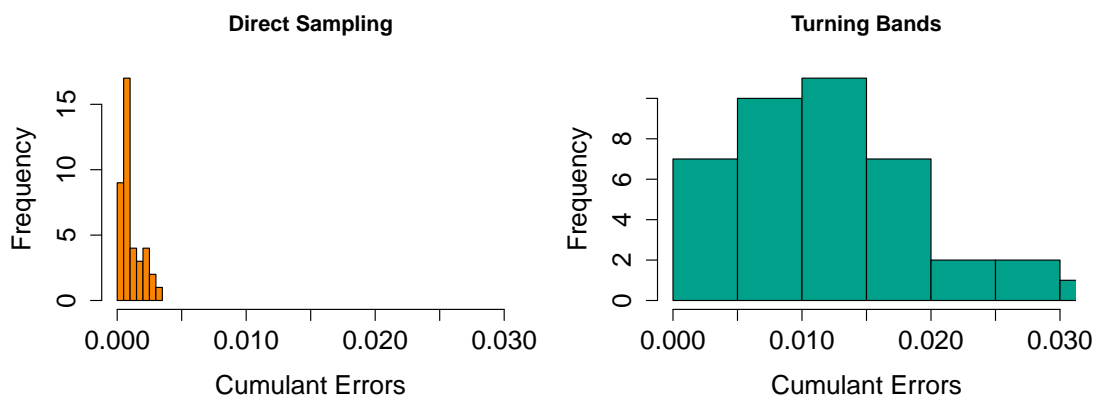


(b) Bivariate Kumbur

Figure G.11: Distributions of the cumulant errors for the univariate Kumbur simulations



(a) Univariate Oak



(b) Bivariate Oak

Figure G.12: Distributions of the cumulant errors for the univariate Oak simulations

# Appendix H

## Declaration of Authorship

I, Yasin Dagasan, contributed:

- %75 of the design and analysis of the research study and authoring the paper entitled “Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study”. O.E. assisted with the design, analysis and supervision of the study and reviewed the paper. E.T. supervised the study and reviewed the paper.
- %80 of the design and analysis of the study and authoring the paper entitled *Pilot-point optimisation of mining boundaries for lateritic bauxite deposits: finding the trade-off between dilution and ore loss*. P.R. and J.S. assisted with the design and analysis of the study and reviewed the paper. O.E. assisted with the problem set-up, supervised the study and reviewed the paper. E.T. supervised the study and reviewed the paper.
- %85 of the design and analysis of the study and authoring the paper entitled “Automatic parameter tuning of multiple-point statistical simulations of lateritic bauxite deposits. P.R. and J.S. assisted with the design and analysis of the study and reviewed the paper. O.E. and E.T. supervised the study and reviewed the paper.
- %85 of the design and analysis of the study and authoring the paper entitled “Multiple-point statistical simulation of the ore boundaries for the lateritic bauxite deposits. O.E. supervised the study, assisted with the design and analysis of the study and reviewed the paper. P.R. and J.S. assisted with the design and analysis of the study and reviewed the paper. E.T. supervised the study and reviewed the paper.

Yasin Dagasan



Date: 25.05.2018

I, endorse that this level of contribution by the candidate indicated above is appropriate.

Dr. Oktay Erten

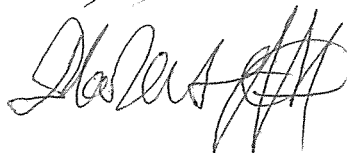
Co-supervisor and co-author



Date: 23.05.2018

Prof. Erkan Topal

Co-supervisor and co-author



Date: 23/05/2018

Prof. Philippe Renard

Co-author



Date: 24.05.2018

Dr. Julien Straubhaar

Co-author



Date: 24.05.2018

# Appendix I

## Copyright Licenses for typeset Chapters 5 and 6

Minerals is a completely Open Access journal with the included article: Dagasan, Y., Renard, P., Straubhaar, J., Erten, O., Topal, E. (2018). Automatic Parameter Tuning of Multiple-Point Statistical Simulations for Lateritic Bauxite Deposits. *Minerals*, 8 (5), 220. DOI: 10.3390/min8050220, freely reproduced completely with no copyright infringements as a part of this thesis by the first author. See <https://creativecommons.org/licenses/by-nc-nd/4.0/> for more information if necessary.

**SPRINGER NATURE LICENSE  
TERMS AND CONDITIONS**

May 13, 2018

This Agreement between Curtin Univeristy -- Yasin Dagan ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4341791140323
License date	May 04, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Natural Resources Research
Licensed Content Title	Pilot Point Optimization of Mining Boundaries for Lateritic Metal Deposits: Finding the Trade-off Between Dilution and Ore Loss
Licensed Content Author	Yasin Dagan, Philippe Renard, Julien Straubhaar et al
Licensed Content Date	Jan 1, 2018
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	>50,000
Author of this Springer Nature content	yes
Title	PhD Student
Instructor name	Roger Thompson
Institution name	Curtin Univeristy
Expected presentation date	Jun 2018
Order reference number	9380
Requestor Location	Curtin Univeristy Locked Bag 30  Kalgoorlie, Western Australia 6433 Australia Attn: Yasin Dagan
Billing Type	Invoice
Billing Address	Curtin Univeristy Locked Bag 30  Kalgoorlie, Australia 6433 Attn: Yasin Dagan
Total	0.00 AUD

**Terms and Conditions**

**Springer Nature Terms and Conditions for RightsLink Permissions**  
**Springer Customer Service Centre GmbH (the Licensor)** hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:



1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).  
  
If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.
2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.
3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.
4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.
5. Where '**reuse in a dissertation/thesis**' has been selected the following terms apply: Print rights for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline ([www.sherpa.ac.uk/romeo/](http://www.sherpa.ac.uk/romeo/)).
6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines <http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/>), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.
7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to [Journalpermissions@springernature.com](mailto:Journalpermissions@springernature.com)/[bookpermissions@springernature.com](mailto:bookpermissions@springernature.com) for these rights.
8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.
9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.
10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

## **Appendix — Acknowledgements:**

### **For Journal Content:**

Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)]

### **For Advance Online Publication papers:**

Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)]

### **For Adaptations/Translations:**

Adapted/Translated by permission from [the Licensor]: [Journal Publisher (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION**  
(Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

**Note: For any republication from the British Journal of Cancer, the following credit line style applies:**

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:

Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [**JOURNAL ACRONYM**])

**For Book content:**

Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author(s)**] [**COPYRIGHT**] (year of publication)

**Other Conditions:**

Version 1.0

**Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

# Appendix J

## Paper 1

Copyright: The Journal of Southern African Institute of Mining and Metallurgy is a completely Open Access journal with the included article: Dagan, Y., Erten, O., Topal, E. (2018). Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study. *The Journal of South African Institute of Mining and Metallurgy*, 118 (2), 173-184. DOI: 10.17159/2411-9717/2018/v118n12a11, freely reproduced completely with no copyright infringements as a part of this thesis by the first author. See <https://creativecommons.org/licenses/by-nc-nd/4.0/> for more information if necessary.



# Accounting for a spatial trend in fine-scale ground-penetrating radar data: a comparative case study

by Y. Dagasana\*, O. Erten\*, and E. Topal†

## Synopsis

In geostatistics, one of the challenges is to account for the spatial trend that is evident in a data-set. Two well-known kriging algorithms, namely universal kriging (UK) and intrinsic random function of order  $k$  (IRF- $k$ ), are mainly used to deal with the trend apparent in the data-set. These two algorithms differ in the way they account for the trend and they both have different advantages and drawbacks. In this study, the performances of UK, IRF- $k$ , and ordinary kriging (OK) methods are compared on densely sampled ground-penetrating radar (GPR) data acquired to assist in delineation of the ore and waste contact within a laterite-type bauxite deposit. The original GPR data was first pre-processed to generate prediction and validation data-sets in order to compare the estimation performance of each kriging algorithm. The structural analysis required for each algorithm was carried out and the resulting variograms and generalized covariance models were verified through cross-validation. The variable representing the elevation of the ore unit base was then estimated at the unknown locations using the prediction data-set. The estimated values were compared against the validation data using mean absolute error (MAE) and mean squared error (MSE) criteria. The results show although IRF- $k$  slightly outperformed OK and UK, all the algorithms produced satisfactory and similar results. MSE values obtained from the comparison with the validation data were 0.1267, 0.1322, and 0.1349 for IRF- $k$ , OK, and UK algorithms respectively. The similarity in the results generated by these algorithms is explained by the existence of a large data-set and the chosen neighbourhood parameters for the kriging technique.

## Keywords

ground-penetrating radar, geostatistics, nonstationarity, universal kriging, ordinary kriging, intrinsic random function of order  $k$ .

## Introduction

In the geological modelling of the laterite-type deposits, the exploration boreholes may be sparsely spaced for two major reasons: (1) the attributed grades do not tend to vary significantly across the deposit; (2) the overall exploration costs must be minimized. However, since the geological contact between the ore and underlying waste unit fluctuates in a rather complex manner, one should not rely solely on the interpolation of the sparsely-spaced borehole data, as the estimates would not reflect the actual variability in the surface. Therefore, it is imperative that the sparsely-spaced borehole data be supplemented with appropriate geophysical information for more accurate resource estimations (Erten *et al.*, 2015, 2013). Ground-penetrating radar (GPR) has widely been used to acquire a

complementary dense data-set to better delineate the interface between two geological units (bauxite/ferricrete) (Francke and Nobes 2000; Francke and Parkinson 2000; Watts 1997). Due to the tropical weathering and leaching mechanisms that generate laterite-type deposits, the geological interface measured through GPR appears to be well correlated with the easting (X) and northing (Y) coordinates of the surface, which indicates the presence of a spatial trend (McLennan, Ortiz, and Deutch, 2006; Leuangthong, Lyall, and Deutsch, 2002). Geostatistical techniques are the main tools which are used to estimate the elevation of the interface at non-sampled locations, and there are different approaches used to account for the trend in a data-set. In this paper, we estimate the elevation of the geological interface through ordinary kriging (OK), universal kriging (UK), intrinsic random function of order  $k$  (RF- $k$ ) methods and compare the estimation performances.

Geostatistical techniques are based on the theory of regionalized variables (RVs) and are used to estimate an attribute of interest at non-sampled locations (Goovaerts, 1997; Journel and Huijbregts, 1978). The idea behind the theory is that a RV  $Z(\mathbf{u}_\alpha), \{\alpha+1, \dots, n\}$  is considered to be the realization of an order-two random function (RF)  $Z(\mathbf{u})$  and is assumed to have been generated according to a probability density function (Matheron, 1971; Olea, 1974). Due to this characteristic of the RV, there is a spatial correlation between the samples, which allows the prediction of  $Z^*(\mathbf{u})$  at each non-sampled location  $\mathbf{u}$ . The prediction of  $Z^*(\mathbf{u})$  requires the covariance function  $C(\mathbf{h})$  of  $Z(\mathbf{u})$  to be known, and this statistical inference can practically be made from an available realization if the realization

\* Mining and Metallurgical Engineering Department, WA School of Mines, Curtin University, Kalgoorlie, WA, Australia.

† School of Mining and Geosciences, Nazarbayev University, Astana, Kazakhstan.

© The Southern African Institute of Mining and Metallurgy, 2018. ISSN 2225-6253. Paper received Apr. 2016; revised paper received Aug. 2017.

## Accounting for a special trend in fine-scale ground-penetrating radar data

exhibits stationary characteristics. If there is a trend in the realization/data-set and the realization exhibits nonstationary characteristics, the change in the value of RV is no longer dependent on the lag distance  $\mathbf{h}$ , but is also dependent on the location of the RV. In the case of a nonstationarity, there are basically two geostatistical methods to account for the trend in the data-set: universal kriging and intrinsic random function of order  $k$  (IRF- $k$ ). These two methods differ in the way they detect the trend and the type of structural function used to describe the spatial relationship.

In order to estimate the value of an attribute at a non-sampled location, the kriging algorithm requires the computation of a system of equations with a known variogram  $\gamma(\mathbf{h})$ . The problem with UK arises when the underlying variogram or covariance function is not known. In UK, the RF  $Z(\mathbf{u})$  is comprised of a trend  $m(\mathbf{u})$  and a residual  $R(\mathbf{u})$  component. The variogram of residuals can be used to calculate the underlying variogram, but this requires the determination of the trend component  $m^*(\mathbf{u})$ . Calculation of  $m^*(\mathbf{u})$ , on the other hand, requires the variogram to be known. One solution to this circular problem is refining both the variogram and the trend estimates iteratively, as mentioned in Neuman and Jacobson (1984). However, it has some associated drawbacks, such as underestimation of the underlying variogram and extreme difficulties in determining the degree of trend or the underlying variogram from the residuals (Armstrong, 1984; Cressie, 1993). Another solution to these problems exists if there is a particular direction or subzone where there is a sub-stationary zone in a nonstationary data-set. Journel and Rossi (1989) proposed that experimental variograms inferred in these directions or subzones can be used to estimate the trend optimally. Chilès (1976) gives an example of such practice assuming an isotropic variogram model computed from the sample variogram in stationary directions or subzones.

IRF- $k$  was introduced by Matheron (1973) due to the practical difficulties in the application of the UK approach. It basically decomposes the trend and covariance structure through increments of a sufficient order to filter out the trend and to achieve stationarity (Chilès and Delfiner, 2012). The kriging system of the IRF- $k$  method is identical to the UK method except that the variogram employed in UK is replaced by the generalized covariance (GC) in the IRF- $k$  method. Contrary to UK, where the trend is required to be estimated beforehand as a linear combination of known, linearly independent functions to obtain the stationary residuals, nonstationarity in IRF- $k$  is accounted for through the calculation of the GC. The advantage of using GC is that it has wider class of admissible functions compared to the variogram and the automatic detection of the parameters of GC makes the application easier (Delhomme, 1978). On the other hand, the use of GC creates some hurdles in practical modelling since the method requires identification of the order  $k$  and it is difficult to interpret the GC (Cressie, 1993). Both UK and IRF- $k$  methods have advantages and disadvantages and as a result, the choice between the two methods is based heavily on the practical difficulties of fitting functions (Buttafuoco and Castrignano, 2005). Christensen (1990) states that the IRF- $k$  and UK are identical provided that the GC is identified correctly.

UK, IRF- $k$ , and OK have been compared in several publications. Journel and Rossi (1989) compared OK and UK in a case study in which they regionalized seam thickness and coal quality variables. The results indicated that the OK and UK methods gave similar results. It was also concluded that any kriging algorithm with moving data windows is equal to considering a nonstationary random function model with a mean re-estimated at each new location. Similarly, Zimmerman *et al.* (1999) compared the performances of four interpolation algorithms, two of which were OK and UK. They stated that although UK was expected to outperform OK in situations where trends exist, OK performed slightly better than UK. Odeh, McBratney, and Chittleborough (1994) modelled soil variables by using different interpolation methods, and stated that OK was the most inferior of all the methods implemented, including UK. Odeh, McBratney, and Slater (1997) compared the performance of several prediction models, including OK and IRF- $k$ , in prediction of soil parameters. The IRF- $k$  method performed slightly better than OK. It is therefore apparent that the performance ranking of UK, IRF- $k$ , and OK algorithms varies between different investigations.

The objective of this study is to predict the variability in elevation of the base of the ore unit and compare the performance of different kriging estimators using densely sampled GPR data. The contribution of this paper is the implementation of, and comparison of, the performance of, different kriging algorithms in the case of a bauxite laterite deposit. This was mainly done in order to ascertain which kriging algorithm is more suitable for bauxite base elevation data with a spatial trend. Since GPR data represents the elevation of an interface surface, nonstationarity was intuitively expected. This was also confirmed by the omnidirectional variograms computed in the initial data analysis. The omnidirectional variograms revealed the existence of nonstationarity in the prediction data-set, requiring handling of the trend by UK and IRF- $k$  methods. Therefore, these methods were employed to account for the spatial trend in the data-set along with the OK algorithm, which considers the spatial trend to be constant. In order to evaluate the performances of each kriging algorithm, GPR data representing the elevation was resampled to form validation and prediction data-sets. The validation data-set was used to assess the performance of the UK, IRF- $k$ , and OK estimators.

### General review of methods

#### Geostatistical theory

The RF can be represented by the following model:

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}) \quad [1]$$

where  $m(\mathbf{u})$  represents the trend component and  $R(\mathbf{u})$  represents the random part having a covariance function in two-dimensional space  $\mathbf{u}=(x,y)$ . The covariance  $C(\mathbf{h})$  is defined as the following:

$$C(\mathbf{h}) = E[R(\mathbf{u})R(\mathbf{u} + \mathbf{h})] \quad [2]$$

where  $\mathbf{h}$  is the lag distance. Considering the second-order

## Accounting for a special trend in fine-scale ground-penetrating radar data

stationarity assumption,  $R(\mathbf{u})$  is assumed to have a zero mean value and the trend  $m(\mathbf{u})$  is assumed to be constant (Oliver and Webster, 2015). Hence,  $C(\mathbf{h})$  is equal to:

$$C(\mathbf{h}) = E [Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h}) - m(\mathbf{u})^2] \quad [3]$$

In situations where the mean is not constant, the covariance cannot exist. In these cases, the assumption of stationarity is weakened to the one called *intrinsic stationarity* (Matheron, 1963), where the expected differences  $E[Z(\mathbf{u}+\mathbf{h})-Z(\mathbf{u})]$  are equal to zero and the covariance of the random part, which is used to measure the spatial relations of the residuals, is replaced by the variance of the differences:

$$Var[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = E [\{Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})\}^2] = 2\gamma(\mathbf{h}) \quad [4]$$

where  $\gamma(\mathbf{h})$  represents the variogram at lag  $\mathbf{h}$ . The following relationship between the variogram  $\gamma(\mathbf{h})$  and covariance  $C(\mathbf{h})$  function exists for a RF which is second-order stationary:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \quad [5]$$

where  $C(\mathbf{0})$  represents the variance  $\sigma^2$  of the RF.

### Ordinary kriging (OK)

In OK, the RV is estimated at an unsampled location  $\mathbf{u}$  as a linear combination of available neighbouring data  $\{Z(\mathbf{u}_\alpha), \alpha=1, \dots, n\}$  (Goovaerts, 2000).

$$Z_{OK}^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}^{OK} Z(\mathbf{u}_\alpha) \quad [6]$$

As with all kriging algorithms, the objective of OK is to minimize the estimation variance as well as ensure the unbiasedness of the estimator. OK weights  $\lambda_{\beta}(\mathbf{u})$  are determined to satisfy this unbiasedness and minimized variance goals. The following OK system is used to obtain the kriging weights:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta}^{OK} \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_{OK}(\mathbf{u}) = \gamma(\mathbf{u}_\alpha - \mathbf{u}) & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}^{OK} = 1 \end{cases} \quad [7]$$

where  $\mu(\mathbf{u})$  represents the Lagrange parameter used to minimize the variance. The only information needed for the OK system is the variogram value corresponding to every lag distance  $\mathbf{h}$ . The kriging variance of OK is calculated by:

$$\sigma_{OK}^2(\mathbf{u}) = C_R(\mathbf{0}) - \sum_{\alpha=1}^n \lambda_{\alpha}^{OK}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{OK}(\mathbf{u}) \quad [8]$$

### Universal kriging (UK)

In UK, the trend component is modelled as a smoothly varying deterministic function of  $\mathbf{u}$  and is expressed as:

$$m(\mathbf{u}) = \sum_{l=0}^L a_l f_l(\mathbf{u}) \quad \text{for } l = 0, \dots, L \quad [9]$$

where  $m(\mathbf{u})$  represents the local mean,  $a_l$ , the unknown coefficients of the trend function, and  $f_l(\mathbf{u})$  represent the known functions of the coordinates  $\mathbf{u}$  and are called *trial* or

*base functions* (Kitanidis, 1997; Rossi and Deutsch, 2014). The residual component  $R(\mathbf{u})$  is modelled as a second-order stationary RF with a zero-mean,  $E[Z(\mathbf{u})]=0$ . Combining the trend and the residual, the RV is represented by the following equation:

$$Z(\mathbf{u}) = \sum_{l=0}^L a_l f_l(\mathbf{u}_\alpha) + R(\mathbf{u}) \quad [10]$$

The trend can be modelled as a low-order polynomial function of the spatial coordinates  $\mathbf{u}$ . Previous studies have indicated that increasing the number of order of trend functions does not lead to better modelling of the trend (Journel and Rossi, 1989).

To satisfy the unbiasedness condition  $E[Z^*(\mathbf{u})-Z(\mathbf{u})]=0$ , Equation [9] can be rewritten as:

$$\sum_{\alpha=1}^n \lambda_{\alpha}^{UK} f_l(\mathbf{u}_\alpha) = f_l(\mathbf{u}) \quad \text{for } l = 0, \dots, L \quad [11]$$

where the set of  $L$  constraints is termed the *universality* or *unbiasedness conditions*. The kriging system satisfying these requirements is then defined by:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta}^{UK} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \sum_{l=0}^L \mu_l^{UK} f_l(\mathbf{u}_\alpha) = C(\mathbf{u}_\alpha - \mathbf{u}) & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}^{UK} = 1 \\ \sum_{\beta=1}^n \lambda_{\beta}^{UK} f_l(\mathbf{u}_\beta) = f_l(\mathbf{u}) & l = 0, \dots, L \end{cases} \quad [12]$$

where  $\lambda_{\beta}$  are the UK weights and  $\mu_l$  are the Lagrange parameters used to determine the coefficients  $\lambda_{\beta}$  minimizing the error variance.

The UK estimate variance is defined as:

$$\sigma_{UK}^2(\mathbf{u}) = C_R(\mathbf{0}) - \sum_{\alpha=1}^n \lambda_{\alpha}^{UK}(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}) - \sum_{k=0}^K \mu_k^{UK}(\mathbf{u}) f_k(\mathbf{u}) \quad [13]$$

The inference of the residual variogram  $\gamma_R(\mathbf{h})$  is, however, not straightforward, as the only available data is in fact  $Z$  values, not  $R$  values. The experimental variogram of  $Z$  values is defined by:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E \{ [Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2 \} \\ &= E \{ [R(\mathbf{u}) + m(\mathbf{u}) - R(\mathbf{u}+\mathbf{h}) - m(\mathbf{u}+\mathbf{h})]^2 \} \\ &= 2\gamma_R(\mathbf{h}) + [m(\mathbf{u}) - m(\mathbf{u}+\mathbf{h})]^2 \end{aligned} \quad [14]$$

For UK, the underlying (trend-free) variogram is assumed to be known (Armstrong, 1984). To predict this variogram, the form of the trend should be known. However, to estimate the form of the trend, the variogram must be known. Various approaches have been developed to resolve this circular problem. The most common approach is to select a direction or a subzone in which the trend  $m(\mathbf{u})$  can be negligible and calculating the experimental covariance along these selected directions or within this subzone (Atkinson and Lloyd, 2007; Chihi *et al.*, 2000; Journel and Rossi, 1989; Myers, 1989).

### Intrinsic random function of order $k$ (IRF- $k$ )

Intrinsic random function of order  $k$  was developed by Matheron (1973) and later refined by Delfiner (1976). It is an alternative tool used in nonstationary data to remove the trend by filtering out the low-order polynomials and is commonly used when it is difficult to infer the underlying

## Accounting for a special trend in fine-scale ground-penetrating radar data

variogram of the variable of interest (Wackernagel, 2002). In the IRF- $k$  method, the deterministic functions  $f_l(\mathbf{u})$  used to represent the trend model  $m(\mathbf{u})$  in Equation [9] are restricted to those which are only translation-invariant and pairwise orthogonal (Wackernagel, 2002).

Considering a set of weights  $\lambda_{\alpha}$  applied to particular points  $\mathbf{u}_{\alpha}$ , a discrete measure  $\mathbf{u}_{\alpha}$  is defined as:

$$\lambda = \sum_{\alpha=1}^n \lambda_{\alpha} \delta_{\mathbf{u}_{\alpha}} \quad [15]$$

where  $\delta_{\mathbf{u}_{\alpha}}$  represents the Dirac measure at point  $\mathbf{u}_{\alpha}$ . Any linear combinations of the weights with RV at locations  $\mathbf{u}_{\alpha}$  are defined as:

$$Z(\lambda) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) \quad [16]$$

The expression shown in Equation [16] is called the allowable linear combination of order  $k$  (ALC- $k$ ) in Equation [17]

$$\sum_{\alpha=1}^n \lambda_{\alpha} f_l(\mathbf{u}_{\alpha}) = 0 \text{ for } l = 0, \dots, L \quad [17]$$

holds true for all monomials of order  $\leq k$ . Considering a nonstationary  $Z(\mathbf{u})$ , if the expression given in Equation [16] is second-order stationary regardless of any translation  $\mathbf{h}$  and whatever the ALC- $k$   $\lambda$ , the RV  $Z(\mathbf{u})$  is called as IRF- $k$ .

The variogram in IRF- $k$  is replaced by a new function called the GC function (Chiles and Delfiner, 2012). GC is denoted by  $K(\mathbf{h})$  and is used to describe the correlation structure of the random part  $R(\mathbf{u})$ . The GC function of an IRF- $k$  is a symmetric function,  $K(\mathbf{h})=K(-\mathbf{h})$ , and satisfies the condition:

$$C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} K(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) \quad [18]$$

An example to GC function is known the as *polynomial generalized covariance function*. The equation of the polynomial generalized covariance function is given as:

$$K(\mathbf{h}) = \sum_{p=0}^k (-1)^{p+1} b_p |\mathbf{h}|^{2p+1} \quad [19]$$

The conditions on  $b_p$  are satisfied if  $b_p \leq 0 \forall p$ .

The intrinsic kriging system minimizing the variance is expressed in terms of the GC and shown in the following:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} K(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) - \sum_{l=0}^k \mu_l f_l(\mathbf{u}_{\alpha}) = K(\mathbf{u}_{\alpha} - \mathbf{u}) & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} f_l(\mathbf{u}_{\beta}) = f_l(\mathbf{u}) & l = 0, \dots, L \end{cases} \quad [20]$$

The estimation variance is defined by:

$$\sigma^2(\mathbf{u}) = K(\mathbf{0}) - \sum_{\alpha=1}^n \lambda_{\alpha} K(\mathbf{u}_{\alpha} - \mathbf{u}) + \sum_{k=0}^K \mu_k f_k(\mathbf{u}) \quad [21]$$

### Case study

#### Geological setting

The region where the mine site is located is composed of Proterozoic and Palaeozoic basement in the eastern part. This basement comprises acid intrusives, extrusives, and metamorphics. Overlying the basement to the west, Mesozoic and Cenozoic sediments dominate. The sediments were

intensively weathered, which played a crucial role in the formation of laterites rich in alumina and bauxites. The bauxite sits above an almost entirely kaolinized pallid zone (Morgan, 1992).

The bauxite deposit in the mine area is thought to have formed from *in-situ* chemical weathering of kaolinite, quartz, and iron oxide minerals (Loughnan and Bayliss, 1961). The occurrence of the alumina-rich was horizon controlled by climate, vegetation cover, chemical conditions, bedrock composition and texture, groundwater circulation, relief, time, and tectonic conditions (Gow and Lozej, 1993). There is a regolith zone in the mine area comprising, from top to bottom, post-weathering sediments (red soil), bauxitic cement, pisolitic bauxite, nodular ferricrete, the kaolinite zone, and the saprolitic zone (Bardossy and Aleva 1990).

#### The GPR survey

An electromagnetic (EM) wave that travels through shallow ground shows different responses to subsurface structures with varying electromagnetic properties such as dielectric permittivity, conductivity, and electromagnetic permeability. GPR utilizes the dielectric permittivity contrast that exists between the geological structures (ASTM D 6432-99 2005). The EM wave emitted into the ground from the transmitting antenna at the surface is reflected back when there is a difference in the electrical properties of the subsurface structures. This reflected wave is then received by a receiving antenna and recorded as a function of time (Davis and Annan, 1989). Being a function of depth, antenna spacing, and average radar-wave velocity, the time taken for the wave to travel to the interface and back up to the surface is called the *two-way travel time*. Knowing the radar wave velocities, this two-way travel time is then converted into depth.

#### Selection of GPR method to map the lateral variability at the mine site

The main aim of the survey was to map the lateral variability at the bauxite/ferricrete interface. As the area to be surveyed was large (360 × 800 m), the method chosen needed to be easy to implement as well as provide fast data acquisition so that the surveying results could be checked immediately. Other considerations in selecting a suitable geophysical method were spatial resolution capability, cost-effectiveness, and data processing requirements (Erten, 2012).

In order to be sure about the applicability of the method selected, the petrophysical properties established in previous work on samples collected from another location within the mine site were considered. The laboratory results revealed that the conductivities of the bauxite and ferricrete are rather low and there is a dielectric permittivity contrast between the bauxite and the ferricrete. This indicated that the emitted waves would reflect from the bauxite interface, favouring the use of the GPR method.

#### The data

The mine area chosen in this case study is approximately 360 × 800 m in size. The data-set comprises the GPR pick-points acquired from the surface of the mine area using a radar device. The specifications of the radar device are given by Francke and Utsi (2009). The total areal coverage of the GPR survey at the mine area was 142 300 m<sup>2</sup> with GPR profiles

## Accounting for a special trend in fine-scale ground-penetrating radar data

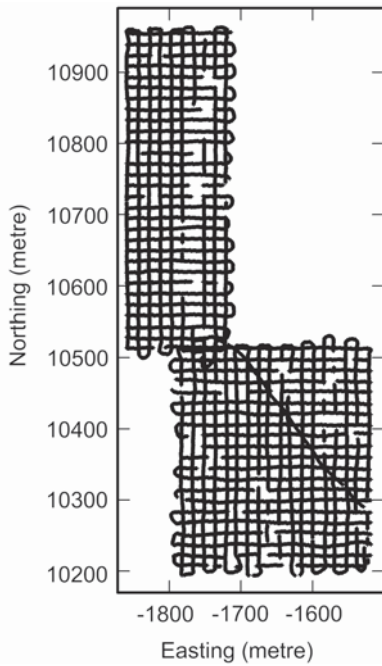


Figure 1—GPR pick-points collected from the mine area

17 940 m in length, which provided 64 670 GPR pick-points distributed along the GPR profiles at 0.25 m interval (Figure 1). The GPR profiles were arranged in a square grid with a line spacing of about 15 × 15 m (Erten, 2012).

Due to the nature of GPR data acquisition, it is expected that there might be multiple points that have the same coordinates. Therefore, the data-set was first pre-processed to mask the potential duplicates in order to avoid any kriging matrix instability in the estimation process. This was carried out by masking all the sample points that were within 0.25 m of the other data-points. The number of raw GPR pick-points, after processing, reduced to 30 630. In order to compare the performance of each kriging method, the data was split into two parts: prediction and validation data-sets. The prediction data-set was generated by re-sampling the GPR pick-points randomly on a regular 15 × 15 m grid. This process yielded 735 pick-points as a prediction/training data-set to be used in the estimation. The remaining 29 895 pick-points were kept for the validation. The flow chart of the data processing and the methodology are presented in Figure 2.

### Exploratory data analysis

The target variable to be regionalized in this study is the

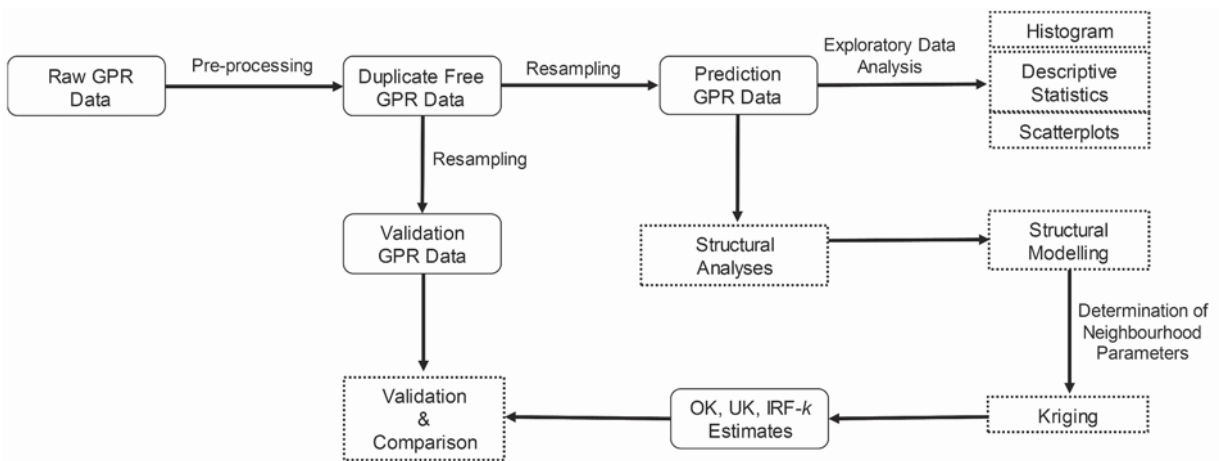


Figure 2—Pre-processing of the GPR data and the methodology

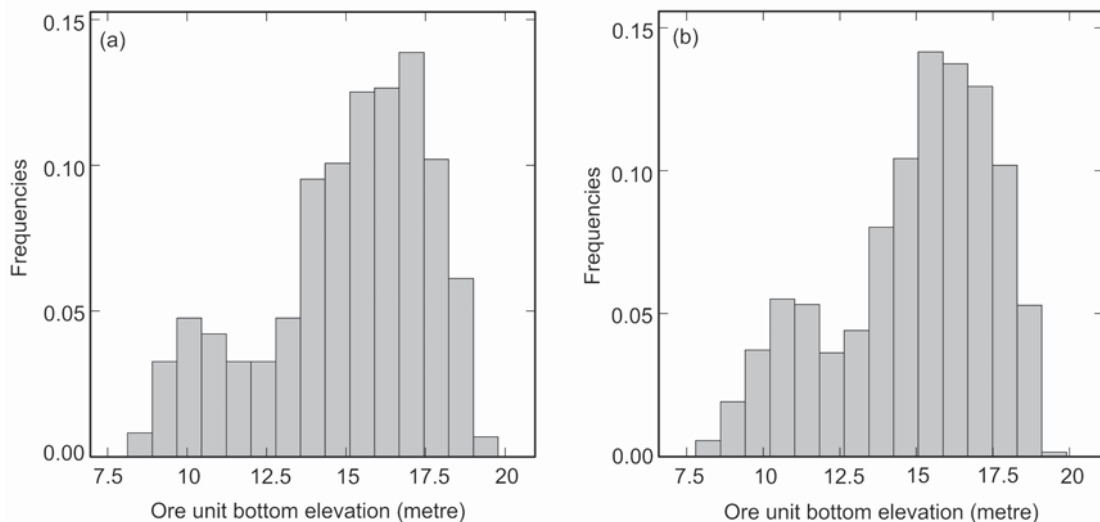


Figure 3—(a) Histograms of prediction GPR data-set and (b) the original raw GPR data



## Accounting for a special trend in fine-scale ground-penetrating radar data

Table 1

### Descriptive statistics of the GPR data for elevation of the ore unit base

Variable	Unit	Count	Mean	SD	Min.	Max.	CV
Ore unit base elevation prediction data	m	735	15.01	2.60	8.14	19.77	0.17
Ore unit base elevation original data	m	30,630	14.92	2.57	7.81	19.90	0.17

elevation of the bauxite/ferricrete interface. This variable was obtained through the GPR survey, the depth to the interface being calculated from the two-way travel time. This depth was then subtracted from the ground surface elevation to obtain the elevation of the bauxite/ferricrete interface. The unit of the variable is metres (m).

The histograms of both prediction and validation data corresponding to the elevations of the base of the ore unit are shown in Figure 3.

As it can be seen in Figure 3, the histograms of the prediction and validation data are closely similar. This resemblance is also revealed in the descriptive statistics of these two data-sets given in Table 1.

The calculated coefficients of skewness of the prediction/training and validation data-sets are  $-0.70$  and  $-0.64$ , respectively.

Scattergrams plotted against X and Y coordinates in Figure 4 reveals that the elevation variable decreases towards the east and increases towards the north, which suggests a possible trend dipping in southeasterly direction.

### Structural analysis

The raw omnidirectional variograms of the ore unit base elevation were computed from the prediction data with 50 lags, having a lag distance of 15 m and a lag tolerance of 7.5 m. This lag distance was the average distance between the samples in the prediction data-set. The experimental omnidirectional variogram is shown in Figure 5.

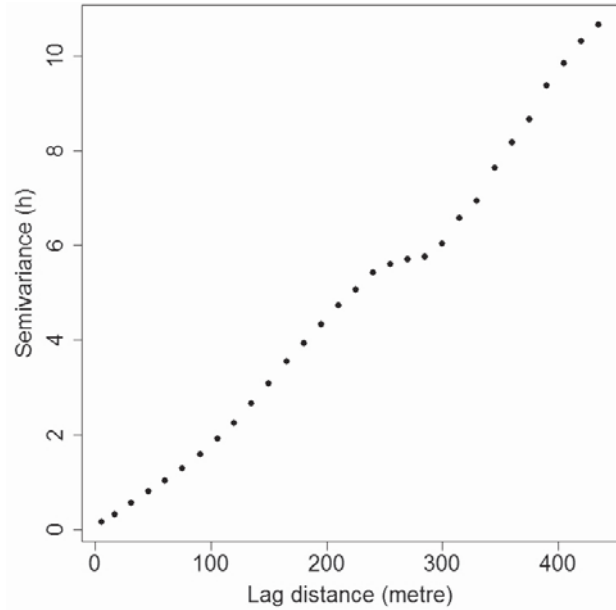


Figure 5—Omnidirectional variogram of prediction data

The apparent increase in the variogram parallel to the increasing lag distance also confirms the existence of a trend or nonstationarity of the elevation variable. This conclusion is supported by the scattergrams shown in Figure 4.

A variogram map was computed to detect the maximum and minimum spatial continuity directions as well as any possible anisotropies. The variogram map and experimental variograms in these directions are shown in Figure 6.

The maximum spatial continuity in the variogram map is seen in the  $N20^\circ$  direction as the variogram values do not change significantly along this direction. On the other hand, perpendicular to this direction, which is  $N110^\circ$ , the elevation varies more rapidly, indicating the minimum spatial continuity direction. In the  $N70^\circ$  direction, there are relatively high variogram values. As seen in Figure 6b, the variogram values in the  $N70^\circ$  direction are higher only for the lag values

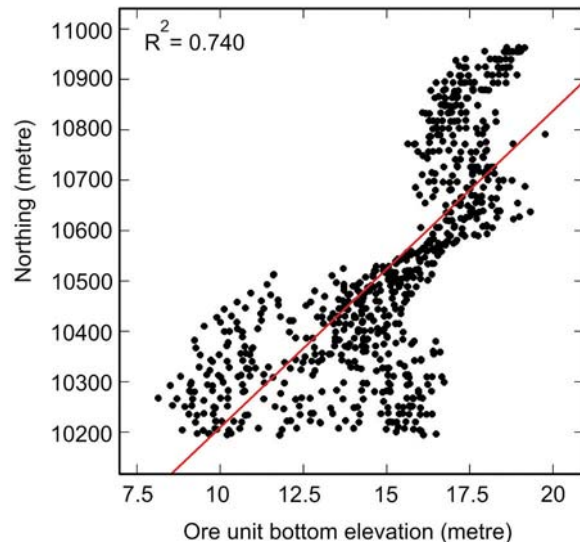
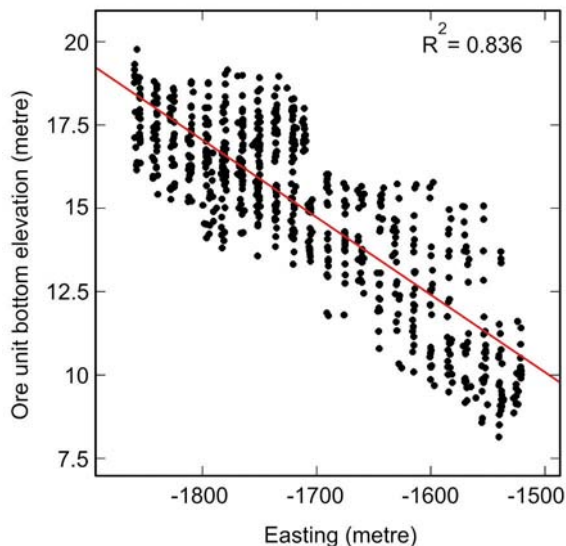


Figure 4—Elevation changes along X and Y coordinates

## Accounting for a special trend in fine-scale ground-penetrating radar data

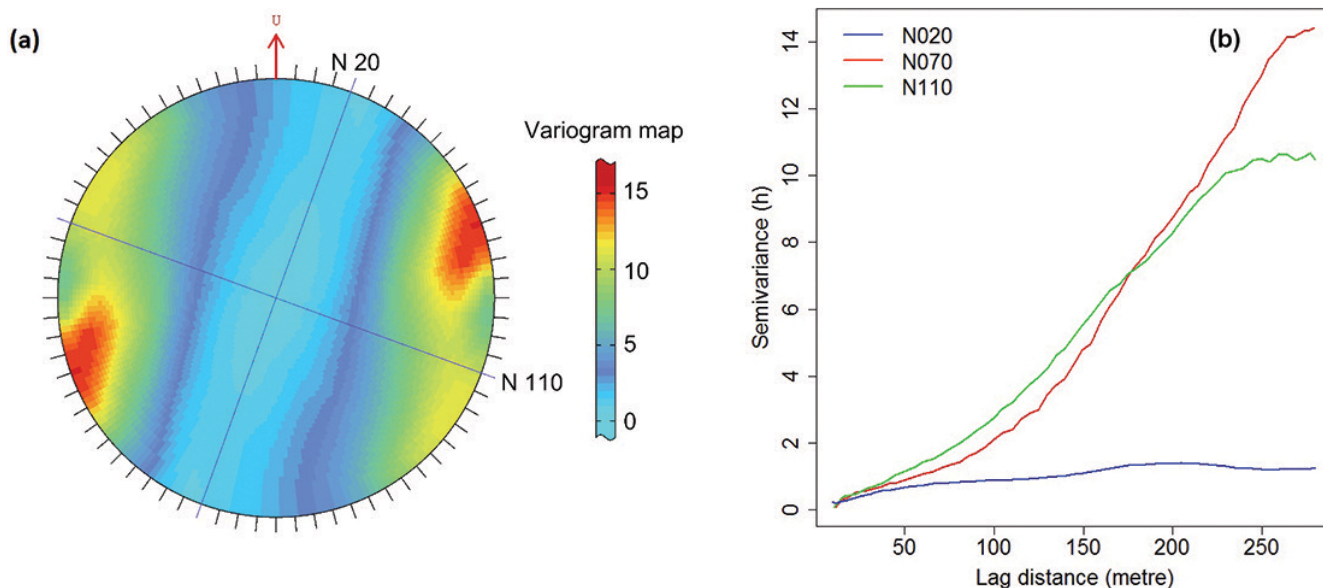


Figure 6—(a) Variogram map of the elevation variable (prediction-training data-set) and (b) experimental variograms in N20°, N70°, and N110° directions

greater than 180. Since the GPR data is densely sampled, only samples that are around 80–100 m apart from the estimation grid node are considered for the estimation and are of greatest importance. This fact confirms the selection of the N110° direction as the minimum spatial continuity direction, since the experimental variogram values in the N110° direction are higher in the 0–100 lag distance range.

### Structural modelling for UK

UK was the first approach used in this study to account for the nonstationarity seen in the data. UK requires a prior determination of the  $L$  trend functions  $F_L(\mathbf{u})$  and the covariance  $C_R(\mathbf{h})$  of the residual component  $R(\mathbf{u})$  inferred from the residual variogram  $\gamma_R(\mathbf{h})$  (Goovaerts, 1997). The  $L$  trend functions were identified by testing low-order polynomial functions,  $k \leq 2$ , by locally fitting the polynomial functions through the ordinary least-squares method. The trend component identification of the elevation variable is summarized in Table II.

In order to identify the order  $k$  of the trend component, the degree of trend, a number of varying orders of polynomials were selected and fitted to the data-points. In this process, errors obtained from the fitting of different orders of  $k$  polynomials are calculated at each point and ranked by order of absolute magnitude. The ranks obtained from each point are then averaged and the order  $k$  having the lowest average rank selected as true  $k$ . In this case study,

polynomials of order 0, 1, and 2 were tried. As is seen from Table II, the smallest mean rank value was 1.850, and it was yielded by the first-order polynomial,  $k=1$ . Therefore, the first-order trend function, which is comprised of  $1, x, y$  monomials, was selected as the best-fit trend. It would, therefore, be expected that the OK, UK, and IRF- $k$  estimates would be similar due to the linear model fitted to the trend.

The first step in determining the underlying variogram for UK was to estimate the coefficients of the drift function with a least-squares based estimator. Once the coefficients of the drift function are determined, the residuals are computed by subtracting the drift from the data. An experimental variogram of the residuals is then calculated and a model is fitted to the residuals experimental variogram. The bias associated with the experimental variogram is computed and an iteration is applied to compute the corrected experimental

Table II

Trend component identification summary			
Trend	Mean error	Mean squared error	Mean rank
1 x y	0.0025	0.6008	1.850
No trend	-0.0006	0.8851	2.149
1 x y x <sup>2</sup> xy y <sup>2</sup>	0.0082	1.221	2.001

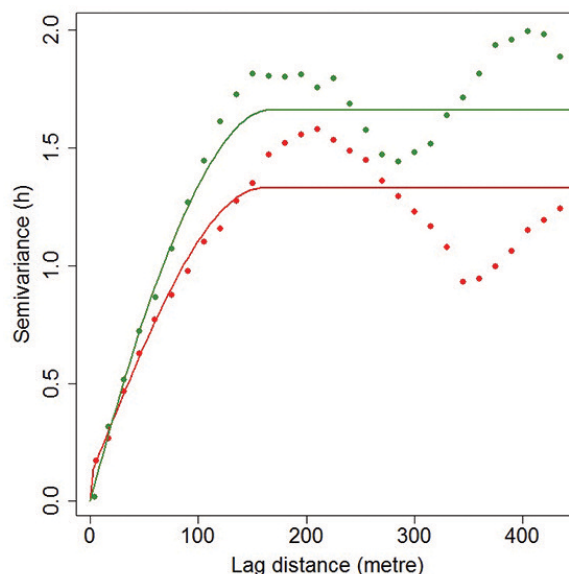


Figure 7—Experimental and theoretical variogram of inferred residuals

## Accounting for a special trend in fine-scale ground-penetrating radar data

variogram. This iteration is carried out  $n$  times,  $n$  being a user-defined maximum iteration number. Variogram fitting for UK was carried out by using ISATIS software. The mathematical expression of the model fitted to the underlying variogram model is as follows:

$$\gamma_R(\mathbf{h}) = 0.34Sp_{h_{b_1}}\left(\frac{20^\circ}{48}\right) + 1.06Sp_{h_{b_2}}\left(\frac{20^\circ}{256}\right)$$

The mathematical model fitted to the underlying variogram is given in Figure 7.

### Structural modelling for IRF- $k$

The second approach used to account for the nonstationarity was the IRF- $k$  method. Nonstationary modelling with the IRF- $k$  method involves three steps: (1) determination of order  $k$ , (2) inference of the optimal GC model, and (3) kriging the elevation variable based on the inferred GC. The GC model is readjusted (if necessary) based on the comparison of the obtained error with the theoretical standard deviations (Delfiner, 1976). Trend analysis has already been carried out for the UK case and the results were summarized in Table II. It was found that the best fit trend for the variable was the order 1 trend ( $k=1$ ).

In order to infer the optimal GC, several arbitrary models of GCs were proposed and tested. The optimality of a GC model was determined on the basis of the ratio called 'Jackknife' which is basically the number indicating the ratio of the theoretical variance to the experimental one (Farkhutdinov *et al.*, 2016). The calculated Jackknife numbers were ranked in ascending order and the GC model corresponding to the Jackknife number closest to unity was selected as the optimal GC. The tested GC models and their scores can be seen in Table III.

It can be seen in Table III that the order 1 GC function yielded the Jackknife score closest to unity. Hence, order 1 type of the GC function was selected with a sill value of 3.148 and a range value of 170 m. Any possible anisotropy that may exist in the data-set is accounted for by the polynomial function which filters out the trend (Delfiner, 1976).

### Structural modelling for OK

In the OK case, the residual variogram  $\gamma_R(\mathbf{h})$  was inferred by calculating the experimental variogram of  $Z(\mathbf{u})$  along the direction in which the trend  $m(\mathbf{u})$  was deemed negligible. The deemed trend-free direction was detected as N20° from the variogram map shown in Figure 6, and the experimental variogram and the model fitted are shown in Figure 8.

However, since the GPR data is densely sampled, only the samples that are 80–100 m away from any estimation grid

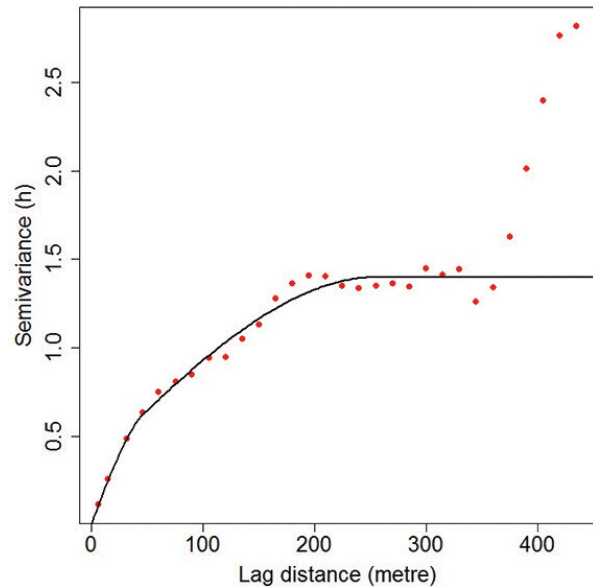


Figure 8—Experimental variogram and the fitted variogram model of the elevation variable in-trend-free direction

are considered and are of greater importance. Therefore, the variogram model reaching the sill value at range value of 256 is considered as robust.

Two spherical models without a nugget variance provided the best fit to the experimental variogram computed by using 15 m as the lag distance and 0.5  $\mathbf{h}$  as the lag tolerance. The mathematical representation of the fitted model is shown in the following:

$$\gamma_R(\mathbf{h}) = 0.34Sp_{h_{b_1}}\left(\frac{20^\circ}{48}\right) + 1.06Sp_{h_{b_2}}\left(\frac{20^\circ}{256}\right)$$

### Cross-validation

The cross-validation technique was used to assess the accuracy of the variogram models fitted to the experimental variograms. It utilizes diagnostic statistics and the accuracy of the prediction is evaluated through various tools (Webster and Oliver, 2001). The criteria used to estimate the accuracy in this study are the mean absolute error (MAE) and mean squared deviation ratio (MSDR). The MAE should ideally be zero, which satisfies the unbiasedness condition. The MSDR is basically the ratio of the computed squared errors to the kriging variances, and the closer the MSDR is to unity, the better the model for kriging (Oliver 2010). These two criteria are calculated as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |Z^*(u_{\alpha_i}) - Z(u_{\alpha_i})|$$

$$MSDR = \frac{1}{N} \sum_{i=1}^N \frac{(Z^*(u_{\alpha_i}) - Z(u_{\alpha_i}))^2}{\sigma^2(\mathbf{u})}$$

where  $N$  is the number of data values (which is 735 for this study),  $Z(u_{\alpha_i})$  is the true value,  $Z^*(u_{\alpha_i})$  is the predicted value, and  $\sigma^2(\mathbf{u})$  is the kriging variance. The results of the cross-validation technique that was implemented to assess the accuracy of the variograms used for UK, IRF- $k$ , and OK are given in Table IV.

Table III

### Tested GC function models and their Jackknife scores

Order 1 GC	Spline	Spherical	Cubic	Jackknife score
3.148	0.000	0.000	0.000	0.902
0.000	1.469	0.000	0.000	0.733
0.000	0.000	2.062	0.000	0.892
0.000	0.000	0.000	8.087	0.624

## Accounting for a special trend in fine-scale ground-penetrating radar data

Table IV

### Results of cross-validation for UK, IRF-*k*, and OK

Variograms of algorithms	MAE	MSDR
OK	0.00047	1.02941
IRF- <i>k</i>	0.00153	0.97807
UK	0.00145	1.21735

The results of the diagnostic statistics show that, although the variogram modelled for OK has the minimum MAE and the closest MSDR value to unity, all of the computed MAE values are small, indicating that there is no significant difference among the error statistics. Based on these results, all of the modelled variograms are considered to be appropriate for spatial prediction.

### Prediction of ore unit base elevation

The first step of the prediction of the elevation of the ore unit base is to define an estimation grid that is capable of covering the whole area containing the data-points as well as minimizing the extrapolation. In this case study, a two-dimensional grid having the following properties was created: the origin is  $X_0 = -1872.00$  m,  $\gamma_0 = 10181.00$  m, the dimensions of the mesh are  $dx = 5.00$  m,  $dy = 5.00$  m, the number of meshes is 75 along the  $x$  direction and 163 along the  $y$  direction, resulting in 12 225 meshes in total. This was then followed by constructing a polygon delineating the boundaries of the data, which defines the resource estimation area. The number of meshes within the polygon boundary is 7 089.

### Neighbourhood parameters

A moving type of neighbourhood was used in all the aforementioned algorithms to estimate the ore unit base elevations. In order to make the data around the estimation grid evenly distributed, eight angular sectors around the estimation grid were defined with a minimum of four samples in each sector. Delfiner (1976) states that the number of

angular sectors should at least be twice the number of unbiasedness conditions, and considering the linear trend, there should at least be six angular sectors around the data. Hence, the number of angular sectors selected satisfies the given rule of thumb. Based on the parameters selected, a minimum of 32 samples around the estimation grid were used in each search neighbourhood. The radius of the search window circle was determined to be 170 m. However, since GPR pick-points were densely sampled, 32 samples around any estimation grid node do not fall further than 80 m radius. In another words, points that are 80 m away from any estimation grid node are not taken into account for estimations through kriging techniques due to this neighbourhood selection. This selected neighbourhood was used for all the kriging methods.

### Predictions with UK, IRF-*k*, and OK techniques

UK was used to estimate the ore unit base elevation variable by making use of the inferred underlying variogram of the residuals and the trend function, which describe the spatial relationships between the sample data. The spatial maps of the ore unit base elevation obtained from the UK estimation are given in Figure 9.

Contrary to UK, IRF-*k* does not require the trend and the underlying variogram of the residuals to be determined beforehand, since it has its own automatic structure identification algorithm allowing it to pick up the best set of parameters within a preselected set of models (Chiles and Delfiner, 2012). Hence, parameters describing the spatial relationship for IRF-*k* algorithm in this study were automatically detected. These parameters are the trend and the chosen optimal GC model. The spatial maps of the estimates from IRF-*k* method are given in Figure 10.

In addition to these nonstationary methods, OK was implemented by neglecting the nonstationarity present in the data. In this method, the variogram model obtained along the trend-free direction was used as a structural input describing the spatial relationship. The maps of the results produced by OK are given in Figure 11.

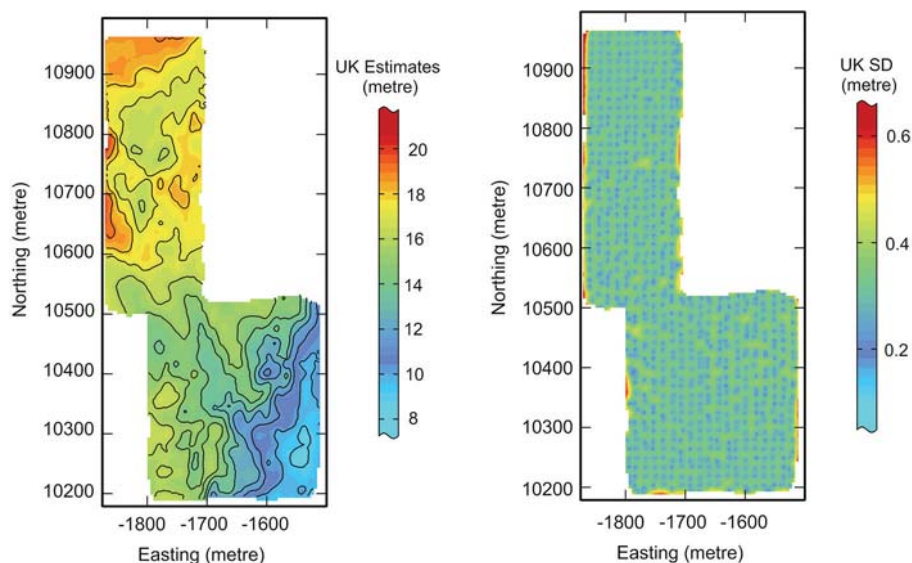


Figure 9—Maps of ore unit base elevation estimates and the associated standard deviation (SD) map calculated by the UK method

## Accounting for a special trend in fine-scale ground-penetrating radar data

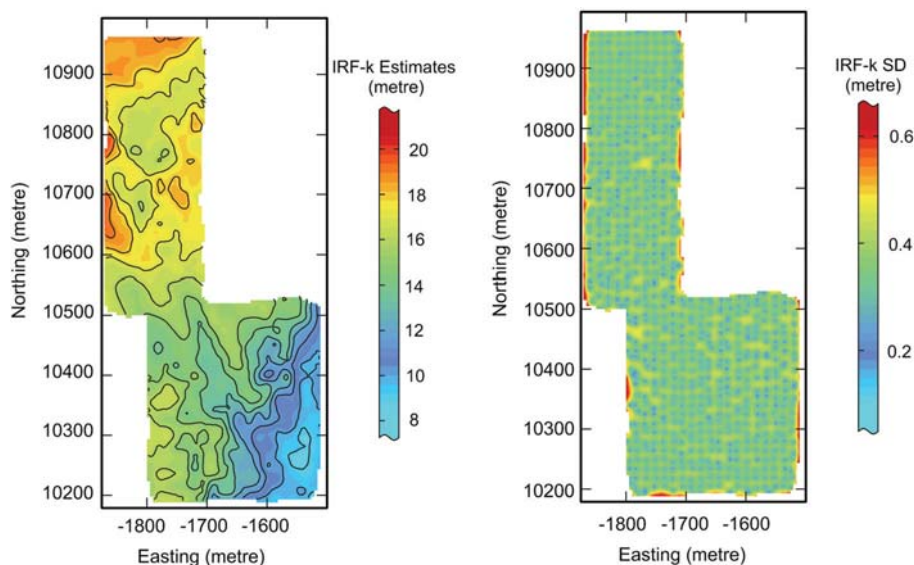


Figure 10—Maps of ore unit base elevation estimates and the associated standard deviation (SD) map calculated by the IRF-*k* method

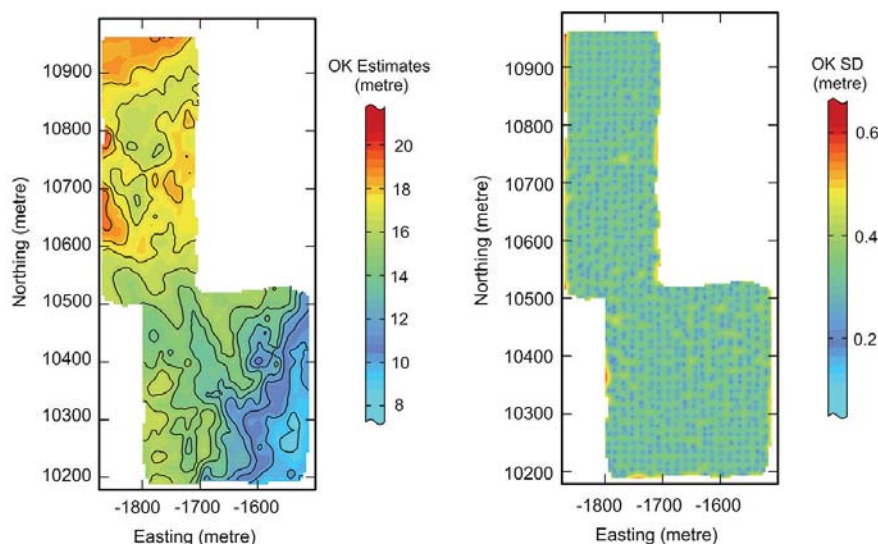


Figure 11—Maps of ore unit base elevation estimates and the associated standard deviation (SD) map calculated by the OK method

### Results and discussion

The performance of the kriging estimators was tested by comparing the validation data with the kriged data. This was done by comparing the mean squared error (MSE) and MAE values. In order to compute these errors, kriged data, which was collocated with the validation data, was copied from each of the UK, IRF-*k*, and OK maps and used to calculate the error associated with each estimation algorithm. The results of MSE and MAE are shown in Table V.

Variable	GPR elevation		
	IRF- <i>k</i>	OK	UK
MAE	0.2445	0.2506	0.2535
MSE	0.1267	0.1322	0.1349

The differences between the errors calculated from the different kriging algorithms are similar and the errors are not significant, considering the mean 15.01 m of the prediction data-set.

Scattergrams generated by plotting the kriged values obtained from the three kriged algorithms against the validation data yielded almost identical results. The coefficients of correlations were 0.990, 0.990, and 0.991 for UK, OK, and IRF-*k*, respectively.

Although IRF-*k* slightly outperformed the other predictor algorithms, the estimation errors were not significant enough to conclude that the more sophisticated IRF-*k* algorithm outperformed OK in this particular case study. The similarity in the results yielded by these techniques is thought to be due to factors such as the densely sampled GPR elevation variable and the selected neighbourhood parameters.

The elevation variables were densely sampled and a maximum of 32 samples were used in the estimation of the

## Accounting for a special trend in fine-scale ground-penetrating radar data

variable at an unknown grid node due to the neighbourhood definition. Since 32 samples were almost always within a 80–100 m radius, the trend was probably not apparent within such a radius. Therefore, the effect of the trend was not experienced to any significant degree, since all the samples used in the estimation were from the immediate vicinity of the estimation grid.

### Conclusions

The focus of this case study was to implement and compare the performances of different geostatistical estimators in the case of a trend apparent in a densely sampled GPR data-set. The UK and IRF- $k$  methods were implemented to account for the trend seen in the data-set and OK was implemented by considering the spatial trend to be constant.

The performances were assessed by comparing the kriged values with the preselected validation data for each kriging algorithm. The results of the comparisons with the validation data have shown that IRF- $k$  outperformed the other algorithms considering MAE and MSE criteria. However, the differences between the results were not sufficiently significant for one kriging algorithm to stand out among the others. For example, the MSE values obtained from the comparison with the validation data were 0.1267, 0.1322, and 0.1349 for the IRF- $k$ , OK, and UK algorithms, respectively. Similarly, the plotted scattergrams demonstrated a similar outcome; the coefficient of correlations obtained from plotting the kriged values against the validation data were 0.990, 0.990, and 0.991 for UK, OK, and IRF- $k$ , respectively. This similarity was mainly due to the large data-set and neighbourhood parameters chosen.

### References

- ARMSTRONG, M. 1984. Problems with universal kriging. *Journal of the International Association for Mathematical Geology*, vol. 16. pp. 101–108. doi:10.1007/Bf01036241
- ASTM D 6432-99. 2005. Standard guide for using the surface ground penetrating radar method for subsurface investigation. ASTM International, West Conshohocken, PA.
- ATKINSON, P.M. and LLOYD, C.D. 2007. Non-stationary variogram models for geostatistical sampling optimisation: An empirical investigation using elevation data. *Computers and Geosciences*, vol. 33. pp. 1285–1300.
- BARDOSSY, G. and ALEVA, G.J.J. 1990. Lateritic Bauxites. Elsevier, Amsterdam.
- Buttafuoco, G. and Castrignano, A. 2005. Study of the spatio-temporal variation of soil moisture under forest using intrinsic random functions of order  $k$ . *Geoderma*, vol. 128. pp. 208–220. doi:10.1016/j.geoderma.2005.04.004
- CHIH, H., GALLI, A., RAVENNE, C., TESSON, M., and DE MARSILY, G. 2000. Estimating the depth of stratigraphic units from marine seismic profiles using nonstationary geostatistics. *Natural Resources Research*, vol. 9. pp. 77–95.
- CHILÈS, J-P. 1976. How to adapt kriging to non-classical problems: three case studies. *Advanced Geostatistics in the Mining Industry: Proceedings of the NATO Advanced Study Institute*, Istituto di Geologia Applicata of the University of Rome, Italy, 13–25 October 1975. Springer. pp. 69–89.
- CHILÈS, J.P. and DELFINER, P. 2012. *Geostatistics: Modeling Spatial Uncertainty*. Wiley.
- CHRISTENSEN, R. 1990. The equivalence of predictions from universal kriging and intrinsic random-function kriging. *Mathematical Geology*, vol. 22. pp. 655–664. doi:10.1007/Bf00890514
- CRESSIE, N. 1993. *Statistics for Spatial Data*. Wiley-Interscience, New York.
- DAVIS, J.L. and ANNAN, A.P. 1989. Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy. *Geological Prospecting*, vol. 37. pp. 531–551.
- DELFINER, P. 1976. Linear estimation of non stationary spatial phenomena. *Advanced Geostatistics in the Mining Industry*. Springer. pp. 49–68.
- DELHOMME, J.P. 1978. Kriging in the hydrosiences. *Advances in Water Resources*, vol. 1. pp. 251–266. doi:10.1016/0309-1708(78)90039-8
- ERTEN, O. 2012. Profiling and mining control to mitigate dilution effect from SiO<sub>2</sub> at the base of a bauxite deposit. PhD thesis, University of Queensland.
- ERTEN, O., KIZIL, M.S., TOPAL, E., and McANDREW, L. 2013. Spatial prediction of lateral variability of a laterite-type bauxite horizon using ancillary ground-penetrating radar data. *Natural Resources Research*, vol. 22. pp. 207–227.
- ERTEN, O., McANDREW, L., KIZIL, M.S., and TOPAL, E. 2015. Incorporating fine-scale ground-penetrating radar data into the mapping of lateral variability of a laterite-type bauxite horizon. *Mining Technology*, vol. 124. pp. 1–15.
- FALLON, G.N., FULLAGAR, P.K., and SHEARD, S.N. 1997. Application of geophysics in metalliferous mines. *Australian Journal of Earth Sciences*, vol. 44. pp. 391–409.
- FARKHUTDINOV, A., GOBLET, P., DE FOUQUET, C., and CHERKASOV, S. 2016. A case study of the modeling of a hydrothermal reservoir: Khankala deposit of geothermal waters. *Geothermics*, vol. 59, Part A. pp. 56–66. ttp://dx.doi.org/10.1016/j.geothermics.2015.10.005
- FRANCKE, J. and UTSI, V. 2009. Advances in long-range GPR systems and their applications to mineral exploration, geotechnical and static correction problems. *First Break*, vol. 27. pp. 85–93.
- FRANCKE, J.C. and NOBES, D.C. 2000. A preliminary evaluation of GPR for nickel laterite exploration. *Proceedings of the Eighth International Conference on Ground Penetrating Radar*, Gold Coast, Queensland. Noon, D.A., Stickley, G.F., and Longstaff, D. (eds). SPIE, Bellingham, WA. pp. 7–12.
- FRANCKE, J.C. and PARKINSON, G. 2000. The new role of geophysics in nickel laterite exploitation and development. *Proceedings of the Mining Millennium/PDAC Convention*, Toronto, Canada. Prospectores and Developers Association of Canada.
- GOOVAERTS, P. 1997. *Geostatistics for Natural Resource Evaluation*. Oxford University Press, New York.
- GOOVAERTS, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, vol. 228. pp. 113–129. doi: 10.1016/S0022-1694(00)00144-X
- GOW, N.N. and LOZEJ, G.P. 1993. Bauxite. *Geoscience Canada*, vol. 20. pp. 9–16.
- ISATIS. 2016. Software manual, 9.0 edn. Avon Cedex, France.
- JOURNEL, A.G. and HUIJBREGTS, C.J. 1978. *Mining Geostatistics*. Academic Press, London.
- JOURNEL, A.G. and ROSSI, M.E. 1989. When do we need a trend model in kriging? *Mathematical Geology*, vol. 21. pp. 715–739.
- KITANIDIS, P.K. 1997. *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press.
- LEUANGTHONG, O., LYALL, G., and DEUTSCH, C. 2002. Multivariate geostatistical simulation of a nickel laterite deposit. *APCOM 2002: Proceedings of the 30th International Symposium on the Application of Computers and Operations Research in the Mineral Industry*, University of Alaska. Bandopadhyay, S. (ed.). Society for Mining, Metallurgy, and Exploration, Littleton, CO. pp. 261–273.
- LOUGHNAN, F.C. and BAYLISS, P. 1961. The mineralogy of the bauxite deposits near Weipa, Queensland. *American Mineralogist*, vol. 46. pp. 209–217.
- MATHERON, G. 1963. Principles of geostatistics. *Economic Geology*, vol. 58. pp. 1246–1266.
- MATHERON, G. 1971. The theory of regionalised variables and its applications. *Ecole des Mines, Cahiers du Centre de Morphologie Mathématique*, Paris. Vol. 5. pp. 211.

## Accounting for a special trend in fine-scale ground-penetrating radar data

- MATHERON, G. 1973. The intrinsic random functions and their applications *Advances in Applied Probability*, pp. 439–468.
- McLENNAN, J.A., ORTIZ, J.M., and DEUTCH, C.V. 2006. Geostatistical simulation of optimum mining elevations for nickel laterite deposits. *CIM Bulletin*, vol. 1, no. 6. pp. 1–9.
- MORGAN, C.M. 1992. Geology of the spheres at Weipa. *Travaux de Comité Internationale pour l'Etude des Bauxites, de Alumine et de l'Aluminium*, vol. 22. p. 61–74.
- MYERS, D.E. 1989. To be or not to be...stationary? That is the question. *Mathematical Geology*, vol. 21. pp. 343–362.
- NEUMAN, S.P. and JACOBSON, E.A. 1984. Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Journal of the International Association for Mathematical Geology*, vol. 16. pp. 499–521. doi: 10.1007/Bf01886329
- ODEH, I., McBRATNEY, A., and SLATER, B. 1997. Predicting soil properties from ancillary information: non-spatial models compared with geostatistical and combined methods. *Proceedings of Geostatistics Wollongong 96*. Vol. 1. Baafi, E.Y. and Scofield, N.A. (eds). Kluwer. pp. 1008–1019.
- ODEH, I.O.A., McBRATNEY, A.B., and CHITTLEBOROUGH, D.J. 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, vol. 63. pp. 197–214.
- OLEA, R.A. 1974. Optimal contour mapping using universal kriging. *Journal of Geophysical Research*, vol. 79. pp. 695–702. doi: 10.1029/JB079i005p00695
- OLIVER, M.A. 2010. *Geostatistical Applications for Precision Agriculture*. Springer.
- OLIVER, M.A. and WEBSTER, R. 2015. *Basic Steps in Geostatistics : the Variogram and Kriging*. Springer.
- ROSSI, M.E. and DEUTSCH, C.V. 2014. *Mineral Resource Estimation*. Springer.
- SINGH, N.K. 2007. Ground penetrating radar (GPR) 'mineral base profiling and orebody optimisation'. *Proceedings of the 6th International Heavy Minerals Conference 'Back to Basics'*, Hluhluwe, South Africa. Southern African Institute of Mining and Metallurgy, Johannesburg. pp. 179–182.
- WACKERNAGEL, H. 2002. *Multivariate Geostatistics*. Springer, Berlin.
- WATTS, A. 1997. Exploring for nickel in the 90s, or 'Til depth us do part! Geophysics and Geochemistry at the Millenium. *Proceedings of Exploration 97: Fourth Decennial International Conference on Mineral Exploration*, Toronto, Canada. Gubins, A.G. (ed.) Prospector and Developers Association of Canada. pp. 1003–1013.
- WEBSTER, R. and OLIVER, M.A. 2001 *Geostatistics for environmental scientists. Statistics in Practice*. Wiley, Chichester, UK.
- ZIMMERMAN, D., PAVLIK, C., RUGGLES, A., and ARMSTRONG, M.P. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology*, vol. 31. pp. 375–390. ♦

**4<sup>th</sup> YOUNG PROFESSIONALS CONFERENCE**

**CREATING A SUSTAINABLE AFRICA MINERALS INDUSTRY THROUGH APPLIED INNOVATION**

**18–19 September 2018, Focus Rooms  
The Core, Sunninghill, Sandton**

### BACKGROUND

The minerals industry is a backbone for most of the economies in the African continent, thus, it is vital that the exploitation of the mineral resources is conducted profitably. However, without any paradigm shift in the tactics we employ to mine these resources, the mineral wealth we possess will not be of any benefit to the current and future generations. There need to be a shift in the way we exploit the resources in order to ensure longevity of current operations and enable mining of deep-level complex orebodies in a safe, healthy and profitable manner. This can be achieved through integration of 1st, 2nd, 3rd and 4th revolutions to create a sustainable minerals industry

### FOR FURTHER INFORMATION CONTACT:

Conference Co-ordinator: Camielah Jardine • E-mail: [camielah@saimm.co.za](mailto:camielah@saimm.co.za)  
Tel: +27 11 834 1273/7 • [www.saimm.co.za](http://www.saimm.co.za)



*Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.*