

TECHNICAL UNIVERSITY OF CRETE  
SCHOOL OF MINERAL RESOURCES  
ENGINEERING

DOCTORAL THESIS

---

DEVELOPMENT OF NEW GEOSTATISTICAL  
METHODS FOR SPATIAL ANALYSIS AND  
APPLICATIONS IN RESERVES ESTIMATION AND  
QUALITY CHARACTERISTICS OF COAL DEPOSITS

---

*Author: Andrew G. Pavlides*

Advising Committee

*Advisor: Professor D.T. Hristopoulos*

*Professor Z. Agioutantis*

*Associate Professor M. Galetakis*



# Abstract

Coal is an important energy resource, especially for countries that have few other energy resources. While the use of fossil fuels is linked to several environmental challenges, these problems can be mitigated. Coal reserves are distributed in space. A better understanding of their spatial distribution would improve exploitation plans and help to better assess financial risks. This research attempts to supplement and improve existing methods of reserves estimation for coal and provide easy-to-use tools to assist in mine planning. The methods proposed herein could also be used for different ore deposits.

We introduce the spatial profitability index (SPI) that evaluates the profitability of mining each individual seam. As prices and mining costs change during the lifetime of a mine, the profitability of different mine sectors or lignite seams may change from profitable to unprofitable or vice versa. The SPI is a flexible tool that can easily and quickly investigate different economic scenarios. The SPI can be used to re-evaluate the pit limits and mine reserves with the current prices and mining costs or near-future estimates.

The SPI is applied in the multiseam lignite mine of Mavropigi, Northern Greece, to evaluate the different lignite seams based on data from 341 drill holes provided by the Public Power Corporation (PPC). The differences in the reserves estimation between the SPI-corrected data and the original data are investigated with regression kriging. The uncertainty of the prediction is investigated based on 5000 conditional simulations.

Based on the SPI, different economic scenarios are investigated that cover a large range of revenue and cost changes. The estimated reserves and revenue difference based for these scenarios are approximated by an empirical function. This function or the resulting graph can give a quick and accurate

approximation of the expected difference in reserves for specific mine sectors or the entire mine. The uncertainty of the prediction is assessed with conditional simulations.

In mining, reserves estimation is usually performed using the family of kriging methods. While these methods are the best linear estimators for Gaussian data, they require the computationally intensive inversion of a large covariance matrix. To reduce the computational load, a user-defined neighborhood radius is usually introduced. In this thesis, we investigate the recently proposed method of the *stochastic local interaction* model (SLI) as a possible alternative. This method avoids the inversion of the covariance matrix. The bandwidth parameter required by the method is self-consistently defined by the geometry of the data-set without any input from the user.

SLI models are used with three different data-sets and compared with kriging methods. The first data used for SLI estimation are the SPI-corrected data from Mavropigi mine. The second data-set is from Campbell county, in the state of Wyoming, USA to assess the distribution of the coal reserves there. The coal reserves of Campbell country constitute an important economic and energy resource for the area. The third and final data set investigated by SLI and kriging, is a non-Gaussian dataset from a gray scale photograph of Pluto. While this dataset does not involve coal deposits, it is employed in this thesis to showcase the performance of SLI methods compared to kriging methods in non-gaussian datasets.

In all three datasets SLI performs faster than the kriging methods and performs as accurately or in the case of non-Gaussian datasets more accurately than kriging methods.

## Περίληψη

Ο γαιάνθρακας αποτελεί σημαντικό ενεργειακό πόρο, ιδιαίτερα για τις χώρες που έχουν λίγες διαφορετικές πηγές ενέργειας. Παρόλο που η χρήση των ορυκτών καυσίμων συνδέεται με πολλές περιβαλλοντικές προκλήσεις, τα προβλήματα αυτά μπορούν να μετριαστούν. Η τιμή του γαιάνθρακα έχει σημαντικές διακυμάνσεις κατά την πάροδο του χρόνου. Για παράδειγμα, η τιμή του γαιάνθρακα που χρησιμοποιείται για παραγωγή ηλεκτρικής ενέργειας, υποδιπλασιάστηκε ανάμεσα στον Ιανουάριο του 2011 και τον Ιανουάριο του 2016.

Τα κοιτάσματα του γαιάνθρακα κατανέμονται στον χώρο. Καλύτερη κατανόηση της χωρικής τους κατανομής αναμένεται να βελτιώσει τα σχέδια αξιοποίησης τους και να δώσει μια καλύτερη εικόνα του οικονομικού ρίσκου. Η έρευνα αυτή επιχειρεί να συμπληρώσει και να βελτιώσει τις υπάρχουσες μεθόδους εκτίμησης αποθεμάτων γαιάνθρακα. Επιπλέον, επιχειρεί να παρέχει ευκολόχρηστα εργαλεία που βοηθούν στον σχεδιασμό εκμετάλλευσης γαιάνθρακα, ειδικά σε περιπτώσεις πολυστρωματικών κοιτασμάτων. Οι μέθοδοι που εισάγονται στην παρούσα έρευνα θα μπορούσαν να χρησιμοποιηθούν και για κοιτάσματα διαφορετικών μεταλλευμάτων.

Σε αυτή την μελέτη εισάγουμε το χωροταξικό δείκτη εκμεταλλευσιμότητας (SPI) που αξιολογεί την εκμεταλλευσιμότητα της εξόρυξης κάθε στρώματος σε πολυστρωματικά κοιτάσματα. Καθώς οι τιμές και το κόστος της εκμετάλλευσης αλλάζουν κατά την διάρκεια ζωής του ορυχείου, η εκμεταλλευσιμότητα του κάθε τομέα ή στρώματος μπορεί να μεταβληθεί από επικερδής σε μη εκμεταλλεύσιμη ή το αντίθετο. Ο δείκτης SPI είναι ένα ευέλικτο εργαλείο που μπορεί γρήγορα και εύκολα να διερευνήσει διαφορετικά οικονομικά σενάρια. Με αυτό τον τρόπο, ο SPI μπορεί να βοηθήσει στην επανεξέταση των ορίων της εκμετάλλευσης και τα

αποθέματα του ορυχείου με τις τρέχουσες τιμές ή τις προβλεπόμενες τιμές στο εγγύς μέλλον.

Ο δείκτης SPI εφαρμόζεται στο πολυστρωματικό ορυχείο της Μαυροπηγής στην Μακεδονία για την αξιολόγηση στρωμάτων λιγνίτη από 341 γεωτρήσεις της Δημόσιας Επιχείρησης Ηλεκτρισμού (ΔΕΗ). Οι διαφορές στην εκτίμηση ενεργειακών αποθεμάτων μεταξύ των διορθωμένων από τον SPI δεδομένων και των αρχικών δεδομένων διερευνώνται με kriging παλινδρόμησης (Regression kriging ή RK). Η αβεβαιότητα της εκτίμησης διερευνάται με βάση 5000 δεσμευμένες προσομοιώσεις. Με χρήση του SPI, το ενεργειακό περιεχόμενο του ορυχείου ανέρχεται σε 300-390 Pcal (349 - 454 TWh) με 90% επίπεδο εμπιστοσύνης.

Με βάση τον SPI, διερευνώνται διαφορετικές οικονομικές εκδοχές που καλύπτουν μεγάλο εύρος αλλαγών σε τιμές εσόδων και κόστους. Οι εκτιμώμενες διαφορές στα οικονομικά εκμεταλλεύσιμα αποθέματα για αυτά τα σενάρια προσεγγίζονται με μία εμπειρική συνάρτηση. Η συνάρτηση αυτή ή το γράφημα που προκύπτει μπορεί να δώσει μια γρήγορη και ακριβή προσέγγιση της αναμενόμενης διαφοράς στα ενεργειακά αποθέματα ανάλογα με τις μεταβολές στην τιμή του προϊόντος ή του κόστους εξόρυξης για συγκεκριμένους τομείς του ορυχείου, ή ολόκληρο το ορυχείο. Η αβεβαιότητα της πρόβλεψης αξιολογείται με δεσμευμένες προσομοιώσεις.

Η εμπειρική συνάρτηση που προκύπτει, μπορεί να χρησιμοποιηθεί και για να γίνει προσεγγιστική εκτίμηση του συνολικού κόστους εξόρυξης του λιγνίτη του ορυχείου ή συγκεκριμένων τομέων.

Στον τομέα της εξόρυξης, η εκτίμηση αποθεμάτων συνήθως πραγματοποιείται με χρήση των μεθόδων kriging. Ενώ αυτές οι μέθοδοι είναι οι καλύτεροι γραμμικοί εκτιμητές για γκαουσιανά δεδομένα, χρειάζονται την υπολογιστικά απαιτητική αντιστροφή μεγάλων πινάκων συνδιασποράς. Για την μείωση των υπολογιστικών απαιτήσεων, συνήθως εισάγεται μια ακτίνα εκτίμησης από τον χρήστη. Σε αυτήν την διατριβή, διερευνούμε την μέθοδο των στοχαστικών τοπικών αλληλεπιδράσεων (SLI) ως μια εναλλακτική λύση. Η καινοτομία αυτή μέθοδος αποφεύγει την αντιστροφή του πίνακα συνδιασποράς. Η παράμετρος εύρους ζώνης που

απαιτείται από τη μέθοδο είναι καθορίζεται αυτόματα από τη γεωμετρία των δεδομένων χωρίς οποιαδήποτε είσοδο από το χρήστη.

Τα μοντέλα SLI χρησιμοποιούνται με τρία διαφορετικά σύνολα δεδομένων και σε συγκρίνονται με τις μεθόδους kriging. Το πρώτο σύνολο δεδομένων που χρησιμοποιούνται για εκτίμηση με τα SLI είναι τα διορθωμένα από τον SPI δεδομένα λιγνίτη της Μαυροπηγής. Το δεύτερο σύνολο δεδομένων είναι δεδομένα γαιάνθρακα από την κομητεία Campbell της πολιτείας Wyoming των ΗΠΑ. Τα κοιτάσματα γαιάνθρακα της κομητείας του Campbell αποτελούν μια σημαντική οικονομική και ενεργειακή πηγή για την περιοχή καθώς η κομητεία περιλαμβάνει τα δύο μεγαλύτερα ανθρακωρυχεία της Γης. Το τρίτο και τελευταίο σύνολο δεδομένων που διερευνώνται από τα μοντέλα SLI και kriging είναι ένα μη γκαουσιανό σύνολο δεδομένων από μια ψηφιακή φωτογραφία του πλανήτη-νάνου Πλούτωνα. Το σύνολο δεδομένων αυτό, αν και δεν αποτελείται από δεδομένα γαιάνθρακα, διερευνάται για να εξεταστεί η αποτελεσματικότητα των SLI σε σχέση με τις μεθόδους Kriging σε μη Γκαουσιανά δεδομένα.

Και στα τρία σύνολα δεδομένων, τα μοντέλα SLI λειτουργούν γρηγορότερα από τις μεθόδους kriging. Επιπλέον δίνουν το ίδιο ακριβείς εκτίμησης με τις μεθόδους kriging για Γκαουσιανά δεδομένα ή στην περίπτωση των μη Γκαουσιανών δεδομένων, δίνουν πιο ακριβείς εκτιμήσεις.

Η εργασία αυτή ακολουθεί την ακόλουθη δομή:

Στο πρώτο κεφάλαιο, γίνεται η εισαγωγή που δίνει το κίνητρο, τους σκοπούς και την καινοτομία αυτής της μελέτης. Στο δεύτερο κεφάλαιο παρουσιάζεται η βασική θεωρία που απαιτείται για την κατανόηση αυτής της διατριβής. Στο τρίτο κεφάλαιο εισάγεται και επεξηγείται ο δείκτης εκμεταλλευσιμότητας και εφαρμόζεται στο λιγνιτωρυχείο της Μαυροπηγής στην Μακεδονία. Στο κεφάλαιο αυτό εισάγεται και η συνάρτηση της διαφοράς εκτιμώμενων αποθεμάτων (ERD). Εξετάζεται πως μπορεί να χρησιμοποιηθεί για να δώσει γρήγορες εκτιμήσεις των αποθεμάτων υπό διαφορετικές οικονομικές συνθήκες και το συνολικό κόστος εξόρυξης. Το τέταρτο κεφάλαιο παρουσιάζει τα μοντέλα στοχαστικών τοπικών αλληλεπιδράσεων και συγκρίνει την εφαρμογή τους





# Acknowledgements

A. Pavlides and D. T. Hristopulos acknowledge support from the project SPARTA 1591: “Development of Space-Time Random Fields based on Local Interaction Models and Applications in the Processing of Spatiotemporal Datasets,” which is implemented under the “ARISTEIA” Action of the operational programme Education and Lifelong Learning and is co-funded by the European Social Fund (ESF) and National Resources. We would also like to thank the Public Power Corporation of Greece for providing the data sets from Mavropigi mine.

I would like to heartily thank my advising committee, professors Dionisis Hristopulos, Zacharias Agioutantis and associate professor Michael Galetakis for their guidance and help provided during this research and especially my supervisor professor Dionisis Hristopulos for the many hours he dedicated to this work. I would also like to thank Dr Christos Roumpos, Director of mining studies division, in PPC for the assistance, advice and experience he provided us with, along with the PPC data. I would like to express my gratitude to Dr Ricardo Olea from United States Geological Survey (Reston, Virginia, USA) for his input and help about the Campbell county reserves.

Finally, I would like to thank my friends and colleagues in the Geostatistics Laboratory of the Mineral Resources Engineering School of Technical University of Crete. Dr Ivi Tsantili, Dr Milan Žukovič and Dr Manolis Varouhakis for their assistance with the SLI models and SSRF variogram model and Dr Vasiliki Mouslopoulou for her assistance with identifying the presence and significance of tectonic faults in the Mavropigi Mine. Mr Manolis Petrakis, Vasilis Androulakis and Sarantis Kyritsis for providing crucial technical support. Last but not least, I would like to thank Mrs Viki Agou for her invaluable assistance in writing this thesis.



# Contents

<b>1</b>	<b>Introduction</b>	<b>23</b>
1.1	Motivation . . . . .	23
1.2	Objectives . . . . .	25
1.3	Innovation . . . . .	26
1.4	Structure . . . . .	26
<b>2</b>	<b>Preliminary Information</b>	<b>29</b>
2.1	Mathematical concepts . . . . .	29
2.1.1	Random variables and random fields . . . . .	29
2.1.2	Covariance . . . . .	30
2.1.3	Statistical homogeneity . . . . .	31
2.1.4	Statistical isotropy . . . . .	32
2.2	Spatial analysis . . . . .	32
2.2.1	Trend model . . . . .	33
2.2.2	Variogram models . . . . .	33
2.2.3	Spartan spatial random field model . . . . .	35
2.2.4	Kriging . . . . .	36
2.2.5	Inverse distance weights . . . . .	40
2.3	Maximum likelihood estimation . . . . .	41
2.3.1	Correlated random variables . . . . .	42
2.4	Simulation . . . . .	43
2.5	Spatial model validation . . . . .	44

2.5.1	Validation measures . . . . .	44
2.6	Coal mining terms . . . . .	47
2.6.1	Existing economic indices . . . . .	48
2.6.2	pit limit optimization . . . . .	49
<b>3</b>	<b>Spatial Profitability Index</b>	<b>51</b>
3.1	Introduction . . . . .	51
3.1.1	Mavropigi lignite deposit . . . . .	52
3.2	Spatial profitability index . . . . .	53
3.2.1	Definition . . . . .	54
3.2.2	SPI and mid-term mine planning . . . . .	57
3.2.3	Supplementing thresholds for SPI . . . . .	59
3.3	Data and exploratory analysis . . . . .	60
3.3.1	Spatial data analysis . . . . .	64
3.4	Regression kriging reserves estimation . . . . .	68
3.4.1	Cross-validation analysis of the spatial model . . . . .	68
3.4.2	Estimation of the total energy content . . . . .	69
3.4.3	Assessment of energy content uncertainty . . . . .	72
3.5	Empirical equation for reserves correction . . . . .	72
3.5.1	ERD and gross mine profit . . . . .	76
3.6	SPI-based pit bottom elevation . . . . .	80
3.7	Summary . . . . .	81
<b>4</b>	<b>Stochastic Local Interaction Model</b>	<b>85</b>
4.1	Introduction . . . . .	85
4.2	Review of SLI theory . . . . .	85
4.2.1	Notation . . . . .	86
4.2.2	The SLI model . . . . .	88
4.2.3	Precision matrix formulation . . . . .	90
4.2.4	Model estimation . . . . .	90
4.2.5	Prediction with SLI . . . . .	91

<i>CONTENTS</i>	13
4.3 Case study of Mavropigi mine . . . . .	92
4.3.1 Optimal kernel selection . . . . .	93
4.3.2 Method comparisons . . . . .	94
4.4 Coal reserves estimation in Campbell county . . . . .	97
4.4.1 Coal reserves estimation with SLI . . . . .	98
4.4.2 Coal prediction with kriging . . . . .	100
4.4.3 Method comparisons . . . . .	102
4.5 Grayscale image of Pluto . . . . .	104
4.5.1 Grayscale prediction with SLI . . . . .	106
4.5.2 Grayscale prediction with kriging . . . . .	106
4.5.3 Method comparisons . . . . .	108
4.6 Summary . . . . .	109
<b>5 Simulations</b>	<b>111</b>
5.1 Introduction . . . . .	111
5.2 Tested transformations . . . . .	112
5.2.1 Regression kriging on the transformed data sets . . . . .	113
5.3 Conditional simulations . . . . .	114
5.4 Simulated energy content reserves . . . . .	116
5.5 Analysis of reserves simulation results . . . . .	117
5.6 Summary . . . . .	119
<b>6 Conclusions</b>	<b>121</b>
6.1 Conclusions . . . . .	121
6.2 Progress towards the objectives . . . . .	122
6.3 Suggestions for future studies . . . . .	123
<b>Appendices</b>	<b>135</b>
<b>A Appendix: SPI code</b>	<b>137</b>



# List of Figures

3.1	Schematic of drill-hole in multiseam lignite mine. The elevation is measured in terms of meters above sea level. . . . .	58
3.2	Positions of drill holes and mine boundaries (blue online) in Mavropigi mine and the lines dividing the sectors (red lines) .	63
3.3	Experimental directional variograms (tolerance $\pi/4$ rad) at the directions $0, \pi/4, \pi/2, 3\pi/4$ for a) the PPC data and b) the SPI corrected data. . . . .	65
3.4	Experimental (markers) and optimal SSRF (lines) variogram models for the two LEC area density datasets. The horizontal axis represents normalized distances. The function $\gamma_{in}(r)$ (crosses for the empirical variogram and dotted line for the SSRF model) corresponds to the SPI-corrected LEC area density. whereas $\gamma_{PPC}(r)$ (filled circles for the empirical variogram and dash-dot lines for the SSRF model) corresponds to the LEC area density based on the mine's pit bottom elevation as determined by PPC. . . . .	67
3.5	Regression kriging interpolation maps using orthogonal grid cells of dimensions $19.5 \text{ m} \times 22.5 \text{ m}$ . (a) Map of LEC area density in Mavropigi mine based on geological lignite data. (b) Map of the kriging error standard deviation. (c) Difference between LEC area density estimates based on the original versus the SPI-corrected data. . . . .	71

- 3.6 Histograms of energy content (Pcal) based on 5 000 Monte Carlo conditional simulations for (a) the original dataset and (b) the SPI-corrected data. LEC values between the 5% and 95% percentile based on the 5000 simulations are shown. . . . 73
- 3.7 Estimated difference of energy reserves versus  $\delta_c$  (markers) plotted against the theoretical model of equation (3.16) (solid red line) for the entire mine (100% is 353 Pcal) and for sector 3 (100% is 82 Pcal). The envelope defining the 90% confidence interval based on 2 000 simulations is also shown (green online). 75
- 3.8 Estimated difference of energy reserves versus  $C_c$  (markers) plotted against the theoretical model of equation (3.17) (solid line) for the whole mine and Sector 3, assuming lignite price  $T = 15 \text{ €/Gcal}$ . . . . . 78
- 3.9 Estimated difference  $\delta R$  in the reserves of the Mavropigi mine with an increase of the maximum mining cost threshold from  $C_c = C$  to  $C'_c = C + \delta C$ . . . . . 79
- 3.10 Suggested open pit bottom elevation (height above sea level) for various  $\delta_c$  values. The bottom elevation shown does not account for technical constraints. . . . . 82
- 4.1 Histogram of the SLI bandwidth for the exponential kernel, obtained for the Mavropigi mine data set using  $k=1$  for the  $k$ -nearest neighbor. . . . . 95
- 4.2 Interpolation maps of the LEC area density in Mavropigi mine using SPI with a critical threshold  $\delta_c = 1.1$ . (a) Kriging map (b) SLI map. . . . . 96
- 4.3 Power River Basin location in southeast Montana and north-east Wyoming, in the USA. The Campbell county is surrounded by the red line. . . . . 97
- 4.4 Drill hole locations in Campbell county, Wyoming. Locations above the dotted line belong to area 1 and locations below the dashed line belong to area 2. There is a 7% overlap between the two areas. . . . . 98



4.5	Normal probability plot of the coal data. The data (blue dots) are sorted along the X-axis according to value. The Y-axis represents the quantiles. The red line represents the theoretical normal distribution. . . . .	99
4.6	Histogram of the bandwidth for Campbell county (a) Area 1 (b) Area 2 . . . . .	100
4.7	SLI map for coal seam thickness (m) in Campbell county, Wyoming in the USA. . . . .	101
4.8	SSRF variogram model (equation (2.10)) for coal seam thickness (m) in Campbell county (line) and experimental variogram values (circles). The model parameters are in Table 4.7.	102
4.9	Regression kriging interpolation map for coal seam thickness (m) in Campbell county, Wyoming in the USA. . . . .	103
4.10	Image of the dwarf planet Pluto, 88 by 88 pixels. . . . .	104
4.11	Histogram of the Pluto intensity values. . . . .	105
4.12	Image of the Training set, containing 20% of the pixels from the original Pluto image. . . . .	105
4.13	Histogram of the Laplacian $L_s = \nabla^2 x_s$ of the intensity values $x_s$ of the entire Pluto image of figure 4.10. Values of $L_s < 0$ are set to 0 and values of $L_s > 255$ are set equal to 255. . . .	106
4.14	Histogram of bandwidths for the Pluto image. . . . .	107
4.15	Pluto image reconstructed by the SLI model with the parameters of Table 4.2 and the tricube kernel. . . . .	107
4.16	SSRF variogram model (equation (2.10), line) and experimental variogram (circles) for pixel grayscale intensity for the training set of Pluto. The model parameters are in Table 4.10.	108
4.17	Pluto image recreated by ordinary kriging . . . . .	108
5.1	Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the square root transformation of LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS. . . . .	115

5.2	Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the logarithmic transformation of LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS. . . . .	115
5.3	Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the non-trasnformed LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS. . . . .	116
5.4	Normal probability plots for the LEC reserves estimations. 5000 simulations were conducted using CDKC simulation for the three different data sets. . . . .	117
5.5	Histograms of energy content (cal) based on 5 000 Monte Carlo conditional simulations. . . . .	118

# List of Tables

3.1	Hypothetical energy content per lignite seam . . . . .	57
3.2	Hypothetical economic parameters of the three lignite seams shown. . . . .	57
3.3	Economic parameters of the Mavropigi mine sectors. The selling price of lignite is $T = 15 \text{ €/Gcal}$ and the assumed critical threshold is $\delta_c = 1.1$ . . . . .	62
3.4	Statistics of lignite energy content (LEC) area density and SPI-corrected LEC area density ( $\text{Gcal/m}^2$ ) for Mavropigi mine. SPI-corrected data are based on a threshold $\delta_c = 1.1$ . . . . .	63
3.5	Optimal parameter values for the linear trend model (3.11). $R$ is the Pearson correlation coefficient between the data and the trend. . . . .	64
3.6	Anisotropic ratio for both detrended data sets estimated with MLE method. $\rho$ is the anisotropic ratio estimated with MLE and $\phi$ is the direction of the anisotropy in radians. . . . .	66
3.7	Optimal parameters of the SSRF variogram models for LEC area density ( $\gamma_{PPC}(r)$ ) and SPI-corrected LEC area density ( $\gamma_{SPI}(r)$ ) data. The estimation of parameters was based on weighted least squares. The length parameter $\xi$ is non-dimensional because it is expressed in normalized coordinates. . . . .	68

3.8	Validation measures for regression kriging. Model 1 refers to the original data with SSRF variogram and Model 2 to SPI-corrected data with SSRF variogram model. $\rho$ is Pearson's correlation coefficient, $r_S$ is Spearman's correlation coefficient, $\tau$ is Kendall's correlation coefficient, ME is the mean error (Gcal/m <sup>2</sup> ), MAE is the mean absolute error (Gcal/m <sup>2</sup> ), MaxAE is the maximum absolute error (Gcal/m <sup>2</sup> ) and RMSE is the root mean square error (Gcal/m <sup>2</sup> ). . . . .	69
3.9	Lignite energy content for each sector of Mavropigi mine for the original data and the SPI corrected data. . . . .	70
3.10	Simulation statistics. Dataset 1 corresponds to data without use of the SPI while dataset 2 are the data corrected by the SPI. The following symbols are used: $\bar{X}$ : average energy content; $\sigma_X$ : standard deviation of energy content; $x_{5\%}$ : 5% percentile; $x_{95\%}$ : 95% percentile . . . . .	72
3.11	Parameters of the ERD sigmoid function model (3.16) used to estimate changes in energy reserves. "Main" corresponds to the ERD parameters of the model based on the SPI-corrected dataset for the different critical thresholds. "Lower" corresponds to the ERD parameters for the function representing the lower limit of the 90% confidence interval, whereas "Upper" corresponds to the ERD parameters for the upper limit. The parameters for Sector 3 are denoted as "Main(S3)", "Lower(S3)" and "Upper(S3)". The confidence intervals are based on 2000 Monte Carlo simulations at each investigated critical threshold. . . . .	76
3.12	Coefficients of equation (3.17) for the entire Mavropigi mine and for Sector 3. . . . .	77
4.1	Examples of Kernel functions [2]. The normalized distance is $u = r/h$ , where $r$ is the Euclidean distance and $h$ is the kernel bandwidth. $I_A(u)$ is the indicator function of set $A$ . $I_A(u) = 1$ , if $u \in A$ and $I_A(u) = 0$ , if $u \notin A$ . . . . .	88
4.2	Parameters of SLI for the tested kernels. $\bar{h}$ is the mean bandwidth and $h_m$ is the median of the bandwidth . . . . .	93

4.3	Validation measures of SLI for different kernel functions, using LVO cross-validation. Time needed and reserves estimated are included. $\rho$ is Pearson's correlation coefficient, $r_S$ is Spearman's correlation coefficient, $\tau$ is Kendall's correlation coefficient, ME is the mean error, MAE is the mean absolute error, MaxAE is the maximum absolute error and RMSE is the root mean square error. Trian represents the triangular kernel function, Expo the exponential kernel function, Quad the quadratic kernel function and Tric the tricubic kernel function.	94
4.4	Validation measures for the estimations with SLI compared to Regression Kriging. $\rho$ is Pearson's correlation coefficient, $r_S$ is Spearman's correlation coefficient, $\tau$ is Kendall's correlation coefficient, ME is the mean error, MAE is the mean absolute error, MaxAE is the maximum absolute error and RMSE is the root mean square error. . . . .	96
4.5	SLI parameters of equation 4.9 for the two neighborhoods of the Campbell county . . . . .	100
4.6	Optimal parameter values $\alpha_0, \dots, \alpha_5$ for the quadratic trend model of equation (4.22). $R$ is the Pearson correlation coefficient between the data and the trend. . . . .	101
4.7	SSRF variogram model parameters . . . . .	102
4.8	Leave-one-out cross-validation measures for the estimations with SLI compared to Regression Kriging for the coal reserves of Campbell county. $\rho$ is Pearson's correlation coefficient, $r_S$ is Spearman's correlation coefficient MAE is the mean absolute error and RMSE is the root mean square error. Time is the time required for the prediction. . . . .	104
4.9	Parameters of the SLI model used for the Pluto image reconstruction. . . . .	106
4.10	SSRF variogram model parameters fit on Pluto image training set. . . . .	107

4.11	Validation measures for the predictions of the 6 184 missing points with the SLI model compared to ordinary kriging for the reconstruction of the Pluto image based on the 1 560 data points. $\rho$ is Pearson's correlation coefficient, MAE is the mean absolute error and RMSE is the root mean square error. Time refers to the computational time required for the reconstruction.	109
5.1	Optimal parameter values for the linear trend model (5.1) for the non-transformed and transformed data. $X$ : Non-transformed data (lignite energy content); $\sqrt{X}$ Square root transform of the data; $\ln(X)$ : Logarithmic transform of the data. $R$ is the Pearson correlation coefficient between the data and the trend. . . . .	112
5.2	Exponential variogram model parameters for the non-transformed and transformed data. $X$ : Non-transformed data (lignite energy content); $\sqrt{X}$ Square root transform of the data; $\ln(X)$ : Logarithmic transform of the data. . . . .	113
5.3	Energy content for Mavropigi mine estimated with Regression Kriging and cross-validation correlation coefficients for the three data sets of section 5.2. $X$ : Non-transformed data (lignite energy content); $\sqrt{X}$ Square root transform of the data; $\ln(X)$ : Logarithmic transform of the data. . . . .	114
5.4	Simulation statistics. $\bar{X}$ : average energy content; $\sigma_X$ : Standard deviation of energy content; $x_{2.5\%}$ : 2.5% percentile; $x_{97.5\%}$ : 97.5% percentile; bias = $\bar{X} - \hat{X}_{OK}$ . . . . .	118
A.1	drill-hole format. ID and ID-core are the ID numbers assigned to the drill-hole and the drill-hole core data. Floor and ceiling correspond to the height above sea level for the floor and ceiling of the core in question. W%, AWF% and CO <sub>2</sub> % are the water content, ash water free content and CO <sub>2</sub> content for the core. LCV is the Lower Calorific Value (in kcal/kg) of the core. X, Y, Z are the cardinal coordinates of the drill-hole head. . . . .	138

# Chapter 1

## Introduction

### 1.1 Motivation

In spite of environmental concerns related to fossil fuels, coal remains an important energy resource which contributes to the energy independence of countries that do not possess other energy resources. Currently, approximately 30% of the electricity generation in the European Union is coal-based. The exploitation of coal contributes significantly to job creation; for example, in 2012 the coal industry provided  $\approx 240,000$  jobs within the European Union [4]. The value of thermal coal varies significantly with time [5].

The exploitation of coal is linked to several environmental challenges. The burning of coal in power plants releases large amounts of  $\text{CO}_2$  in the atmosphere. Byproducts of the coal combustion are fly ash, bottom ash and sludge. The coal may also contain sulfides that release sulfur dioxide ( $\text{SO}_2$ ) when burned or contribute to toxic hydrogen sulfide ( $\text{H}_2\text{S}$ ) release [59]. Another issue of concern is the large quantities of waste that are disposed in areas around the mine [71]. Yet, the environmental problems can be mitigated, at least to some extent [12, 43]. For example, fly ash and harmful gases can be captured and neutralized by means of various filters if the ash and sulfide content are accurately estimated. In addition, suitable planning strategies allow for waste to be disposed in exploited parts of the mine.

Both fossil-based and renewable energy resources share a common feature: they are distributed in space (renewable resources are also distributed in

time). It is necessary to understand quantitatively the variations in the spatial distribution in order to optimize exploitation plans, correctly assess investment risks, and timely compensate for spatial variations in the quality of the energy product.

Kriging methods are the standard geostatistical tool used to estimate the spatial variability of mineral reserves. These methods are commonly used to estimate the quality and quantity characteristics of coal deposits. These methods require the time consuming and computationally intensive inversion of the covariance matrix. Kriging methods benefit from search neighborhoods the researcher or engineer must determine independently. Furthermore, Kriging methods work best if the data set is close to the normal distribution.

Geostatistical analysis is commonly used for reserves estimation and the initial exploitation planning of coal or different resource mines. However, the exploitation lasts for several years. The price of the product, the mining costs, the environmental reclamation cost, the cost to expropriate the land and other costs usually fluctuate significantly during the duration of the exploitation. For example, the price of thermal coal was nearly halved between January 2011 and January 2016 [5]. Changes in costs or product price could render unprofitable the exploitation of certain parts of the mine that were considered profitable based on previous economic estimates [6] thus changing the reserves.

Changes in the price of the product and cost of the exploitation can lead to different mining plans and possibly change the pit limits of the mine. The period that price estimates are considered reliable for most commodities has a time horizon of three to five years, while the initial mining plans may have been designed decades in the past with outdated price predictions. There is a lack of geostatistical tools that can assist in revising the original mine planning by incorporating current predictions of near-future prices that differ from possibly outdated economic predictions made at the time of the initial planning.



## 1.2 Objectives

One of the goals of this research is to show that geostatistical tools [14, 60] can be used to efficiently analyze the spatial variability and estimate uncertainties, and that this information can be used to investigate economic scenarios, environmental concerns, and issues related to mine exploitation.

Kriging methods are the best linear estimators for Gaussian data [39, 23]. However, they require the computationally intensive inversion of covariance matrices and are commonly applied with a neighborhood, the radius of which must be chosen by the user. In this dissertation, a recently proposed alternative method is tested, the *stochastic local interaction* model (SLI). This method avoids the inversion of the covariance matrix. Instead, it is based on a joint probability density function defined by an energy functional which involves local interactions between the data. The SLI employs a bandwidth parameter which is defined by the geometry of the data-set without any input from the user.

An objective of this research is to provide flexible and easy-to-use tools that can assist in defining the optimal pit limits during the initial mine planning and to also provide adjustments of those limits under evolving economic conditions. Such tools should generate reliable estimates in the near future (3-5 years). Motivated by this need, we introduce the spatial profitability index (SPI) that evaluates the profitability of each individual seam and assists in defining whether it is profitable to extract that seam, hence assisting in the definition of the pit limits. In contrast with pit-limit delimitation algorithms, the SPI algorithm does not evaluate blocks but individual seams. The spatial estimations made using the SPI-evaluated drill-hole data can subsequently be used by open-pit delimitation algorithms such as the Lerchs and Grossman algorithm [34] to better estimate pit limits. Changing economic conditions are investigated by using different economic scenarios and using the SPI to re-evaluate the pit limits and reserves. The SPI is flexible enough to be used for re-evaluation of pit limits in different sectors of the mine under changing economic parameters.

As prices and costs change during the exploitation, so do the profitable reserves of the product and therefore, the expected revenue. One of the objectives of this research is to assist in the quick and efficient re-evaluation

of the reserves and expected revenue of each sector or the entire mine. The approach taken is the construction of an empirical function based on the estimated reserves difference (ERD) resulting from application of the SPI. The resulting graph quickly gives an accurate approximation of the reserves and revenue difference in different economic scenarios for specific sectors or the entire mine. In their current form, the SPI and ERD are limited to reserves evaluation for already planned sectors. Their use could possibly lead to change in the planned pit bottom of such sectors.

### 1.3 Innovation

The SPI is a new tool for evaluating the profitability of mining seams. Unlike other indexes as the discounted cash flow which are either financial in nature but do not account for spatial variability or incorporate the spatial variability of physical variables but not their financial significance, SPI is both spatial and financial in nature. The method to calculate the SPI is also straightforward and can be easily applied to different multiseam deposits.

While estimations of reserves in different financial situations have been conducted in the past [42], the change has not been modeled by equations, nor it has been tied to changes in gross mine profit. The ERD is constructed in a way that makes possible to get an accurate and fast estimation of *the total mining cost* for the seams or sectors investigated.

The SLI interpolation method is used in mineral resources data for the first time. This method gives significantly faster predictions of comparable accuracy to the kriging methods for data close to the normal distribution. For data sets in this study that deviate from the normal distribution, the SLI gives more accurate predictions than kriging with reduced computational times.

### 1.4 Structure

The remainder of this thesis is structured as follows:

The second chapter comprises preliminary information. It defines random fields, which are used in geostatistics to represent spatial variability

and some of their important properties, such as the covariance function and the concepts of stationarity and isotropy. Further-more, the chapter briefly describes the steps usually followed in spatial analysis studies. We give the definition of the trend, variogram models, the Spartan covariance model and finally a detailed synopsis of kriging methods. Next is a presentation of the conditional simulation method used in this research, which is based on decomposition of the covariance matrix combined with kriging conditioning. The validation measures used to assess the estimation results and compare between the various methods tested are presented next. This chapter concludes with an explanation of the mining terms used in this research, as well as the commonly used methods and financial indices this research attempts to supplement.

The third chapter introduces the SPI and the case study of Mavropigi lignite mine in Northern Greece. After a brief presentation of the Mavropigi lignite mine, the SPI and the threshold it compares with are defined. Next, the algorithm used to calculate SPI is given followed by an example that illustrates the use of the algorithm. Different thresholds than the accepted ratio of revenue by cost are investigated next. The exploratory analysis of the drill-hole data of Mavropigi mine, broken in five sectors, follows. Two sets of data are used in the calculations: the SPI corrected data and the original data that have not been evaluated with SPI. Next, the spatial analysis based on regression kriging for both datasets is presented along with uncertainty assessment based on conditional simulations. In this chapter, we also present the construction of the estimated reserves difference (ERD) empirical function and discuss how it can be used to generate accurate and fast estimations of reserves changes and the total mining cost for the mine. Finally an example is presented which illustrates the application of SPI could change the pit bottom elevation during the long term or the mid term mine planning.

The fourth chapter presents the SLI model, explores its application to three case studies and shows comparisons with kriging methods. The theoretical background that describes the SLI and the precision matrix is given first, followed by a description of spatial prediction using the SLI model. The SLI model for the Mavropigi mine and the results are compared with the kriging results of the previous chapter. The second case study inves-

tigated involves coal drill-hole data from Campbell county in the state of Wyoming in the USA. After exploratory spatial analysis is conducted, regression kriging and SLI are utilized to generate estimates and the results are compared. The third and final data set examined, is a non-Gaussian dataset from a gray scale digital image of Pluto. SLI results are compared to those obtained with ordinary kriging. While this dataset does not involve coal deposits, it is employed in this thesis to showcase the performance of SLI methods in non-gaussian datasets.

The fifth chapter focuses on the simulations of the Mavropigi data set that are tested for uncertainty assessment of the reserves. Two transformations of the data set (logarithm and square root) are investigated. Regression kriging is applied to the two transformed data sets and the results are compared to the kriging results obtained from the original data set. Since all datasets exhibit similar performance with each other, simulations based on all three datasets are conducted and their results are compared. The chapter closes by an analysis of our decision to use the simulation results of the unmodified data set.

The sixth and final chapter is comprises a synopsis of this research. It details the progress achieved towards reaching the objectives set in the first chapter and concludes with suggestions for further studies.

A list of publications that were produced as a result of this research is presented next before the section that contains the bibliographical references. An appendix with the matlab code used to calculate the SPI follows.

# Chapter 2

## Preliminary Information

### 2.1 Mathematical concepts

Earth science data are distributed through space or time [15, 14]. Geostatistics comprises a collection of mathematical methods that can be used to model and characterize spatial attributes, based on the theory of random fields.

#### 2.1.1 Random variables and random fields

A *random variable*  $X$  is a point-like variable that can take values, denoted by  $x$ , from an ensemble of possible (probable) values. A random variable is called discrete if takes values  $x_i$ , where  $i=1, \dots, N$ , in a countable set, i.e., the frequency of appearance of each value is determined by a probability distribution function. A random variables  $X$  is called continuous if it takes values  $x$  in a continuum set. Then, the probability that  $X$  takes values in an infinitesimal interval around  $x$  is determined by a probability density function (pdf)[49, 16].

The expectation  $E[X]$  of a random variable  $X$  is the mean value of the random variable for all of its states. If the probability distribution of  $X$  has a pdf  $f(x)$ , then the expected value is

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx. \quad (2.1)$$

A stochastic process is a collection of random variables representing the evolution of some system of random values over time. In a stochastic process there are several (often infinite) directions in which the process may evolve. This is the probabilistic counterpart to a deterministic system.

A random field  $X_{\mathbf{s}}$  is a collection of random variables distributed over the spatial domain of interest. The vector  $\mathbf{s}$  corresponds to the position of a point in the study area. At its most basic, discrete case, a random field is a list of random numbers whose indices are mapped into a  $n$ -dimensional space. More generally, the values might be defined over a continuous domain. Values in a random field are often spatially correlated in one way or another.

Mathematical properties characterizing the random variables are extended for the case of variables distributed in space. A random field consists of a set of random variables that describe the spatial (or space–time) changes of an attribute. Therefore, a random field may be viewed as a multidimensional random variable. Due to the interdependence of physical quantities at different locations in space, random fields have unique mathematical properties that distinguish them from a set of independent random variables. In this study, a random field is denoted as  $X_{\mathbf{s}}$  or  $X(\mathbf{s})$ . Values corresponding to a unique realization of random field  $X_{\mathbf{s}}$  are denoted with  $x(\mathbf{s})$  [65, 32].

Fluctuation of the random field  $X(\mathbf{s})$  is a random field  $X'(\mathbf{s})$  the values of which are the fluctuations of the values of the  $X(\mathbf{s})$  around the expected value in those locations. As such, fluctuation  $X'(\mathbf{s})$  is  $X'(\mathbf{s}) = X(\mathbf{s}) - E[X(\mathbf{s})]$ .

Random variables and random fields are described by an ensemble of states. Each state (realization) is also a sample of the field with a corresponding probability which is determined by the multidimensional joint probability density function of the field [65, 32, 1].

### 2.1.2 Covariance

The *Covariance function*, or simply covariance  $c_x(\mathbf{s}_1, \mathbf{s}_2)$  of a random field  $X$  is a measure of how much the fluctuation of the field at point  $\mathbf{s}_1$  influences the fluctuation of the field at the point  $\mathbf{s}_2$ . It is defined according to equation (2.2).

$$\begin{aligned}
c_X(\mathbf{s}_1, \mathbf{s}_2) &= E[(X(\mathbf{s}_1) - E[X(\mathbf{s}_1)])(X(\mathbf{s}_2) - E[X(\mathbf{s}_2)])] \\
&= E[X(\mathbf{s}_1)X(\mathbf{s}_2)] - E[X(\mathbf{s}_1)]E[X(\mathbf{s}_2)]
\end{aligned} \tag{2.2}$$

in the above  $E[\cdot]$  denotes the expectation over the ensemble of the random field states. It follows from equation (2.2) that  $c_X(0) = \sigma_X^2$ , where  $\sigma_X^2$  is the variance of the random field  $X$ .

For two random vectors  $\mathbf{X}_1, \mathbf{X}_2$ , the *covariance matrix* is

$$\begin{aligned}
\mathbf{C}(\mathbf{X}_1, \mathbf{X}_2) &= E[(\mathbf{X}_1 - E[\mathbf{X}_1])(\mathbf{X}_2 - E[\mathbf{X}_2])^T] \\
&= E[\mathbf{X}_1\mathbf{X}_2^T] - E[\mathbf{X}_1]E[\mathbf{X}_2]^T,
\end{aligned} \tag{2.3}$$

where  $\mathbf{X}^T$  is the transpose of vector  $\mathbf{X}$ . The element  $C_{i,j}$  of the covariance matrix is equal to the covariance  $c(X_i, X_j)$  as defined in equation (2.2) [14].

### 2.1.3 Statistical homogeneity

A random field  $X(\mathbf{s})$  is *statistically homogeneous in the weak sense* if the mean value (expectation) is constant, meaning that  $m_X(\mathbf{s}) = m_X$  and the covariance function depends only on the distance vector  $\mathbf{r} = \mathbf{s}_1 - \mathbf{s}_2$  between two points i.e.,  $c_X(\mathbf{s}_1, \mathbf{s}_2) = c_X(\mathbf{r})$ . The second condition guarantees that the variance of a statistically homogeneous field is constant. Hence, the dependence of the value of random field  $X(\mathbf{s})$  on other locations  $\mathbf{s}_i$ , is determined only by the distance between  $\mathbf{s}_i$  and  $\mathbf{s}$  and not by the actual coordinates of these locations.

These above conditions define statistical homogeneity in the weak sense. A random field is statistically homogeneous in the strong sense if the multidimensional pdf for  $N$  points, where  $N$  is any positive integer, remains unchanged by transformations that change the location of the points without changing the distances between them.

Practically, statistical homogeneity implies that there are no spatial trends, so that the spatial variability of the field can be attributed to fluctuations around a constant level equal to the mean value. In practice it is difficult

to test homogeneity in the strong sense since the multidimensional joint pdf of the random field is usually unknown while the covariance is easier to estimate [32, 1].

### 2.1.4 Statistical isotropy

*Isotropy* implies uniformity in all directions. A field is statistically isotropic if it is statistically homogeneous and its covariance function depends only on the distance  $r$ , but not on the direction of the distance vector  $\mathbf{r}$ . Therefore, if the covariance function is isotropic, the field is by definition statistically homogeneous, but not vice versa.

*Anisotropy*, on the contrary, implies that the spatial variability depends on the direction. The covariance of an anisotropic random field depends on both the distance  $r$  and the direction of vector  $\mathbf{r}$ . A random field is considered anisotropic if the directional covariances have different sill or correlation length values [48, 23].

## 2.2 Spatial analysis

Spatial analysis provides tools for resolving variations of various properties at different scales and for identifying potentially useful correlations between variables. Spatial analysis finds applications in energy resources exploitation [20, 23, 58], earth sciences [15, 10], meteorology [41, 38] and agriculture [19] among other fields similar to the way statistical analysis finds applications in numerous fields from material sciences to economics [24]. The geostatistical viewpoint offers a good balance between flexibility and ease of use [14, 52, 45].

In contrast with the simpler distance-based methods (e.g., inverse square distance weighted [34], geostatistical methods also provide quantitative measures of uncertainty. In addition, distance-based methods do not account for spatial correlations, the type of probability distribution followed by the data, geological discontinuities, and the extent of the deposit [47]. Geostatistical methods have successfully been used to investigate energy resources such as solar energy [70, 69], wind speed [13], and geothermal potential [9] [57, pp. 6-



14]. Geostatistics have found extensive applications in coal mining [47, 43] as well as other mineral resources [23, 36]

Geostatistics was based on the work of Danie Krige [39] who developed a linear optimal stochastic interpolator (named Kriging after him) for mineral reserves estimation. Kriging methods have been widely used to estimate coal reserves in multiseam mines [56, 46, 63]. Geostatistical analysis of drill-hole lignite data allows estimating the quantity of inorganic material in lignite seams, the local sulfide content [64], the location of methane gas sources in underground mines, and the gas retained in such sources [37]. Using such spatial information, it is possible to plan coal blending strategies, selective exploitation, as well as enrichment and beneficiation of reserves to lessen the environmental impact [8, 43].

### 2.2.1 Trend model

Random fields can be represented as  $X(\mathbf{s}) = m_X(\mathbf{s}) + X'(\mathbf{s})$ . The function  $m_X(\mathbf{s})$  is the *trend function* that represents the deterministic spatial correlations between the data, which are usually of long range.  $X'(\mathbf{s})$  is a random field that corresponds to the fluctuations of  $X(\mathbf{s})$  around the trend. The expectation of the fluctuation is  $E[X'(\mathbf{s})] = 0$ .

In most applications that involve mineral resources data, there is no theoretical evidence to suggest a particular type of analytical trend model. Because the concept of trend  $m_X(\mathbf{s})$  is usually associated with a smoothly varying component of the variability of  $X_s$  in space, it is typically modeled with low-order polynomials [14, 48, 23].

### 2.2.2 Variogram models

Geostatistical analysis is based on the *variogram function*  $\gamma(\mathbf{s}, \mathbf{s} + \mathbf{r})$ , where  $\mathbf{s}$  is the position vector and  $\mathbf{r}$  the lag (distance) vector. The variogram describes the spatial correlations of the spatial random field  $X(\mathbf{s})$ . It is defined by means of the following equation in which  $E[\cdot]$  denotes the expectation over the ensemble of the random field states [45]

$$\gamma(\mathbf{r}) = \frac{1}{2} E [\{X(\mathbf{s}) - X(\mathbf{s} + \mathbf{r})\}^2] . \quad (2.4)$$

In equation (2.4) it is assumed that the random field is either statistically stationary or that it has stationary increments, so that the variogram depends only on  $\mathbf{r}$  and not on  $\mathbf{s}$ . For the quality parameters of coal this assumption is generally true [43].

If the random field is stationary, the variogram function is connected to the covariance  $C(\mathbf{r})$  as follows

$$\gamma(\mathbf{r}) = C(0) - C(\mathbf{r}). \quad (2.5)$$

It follows from equation (2.4) that  $\gamma(0) = 0$ . As explained in section 2.1.2,  $C(0) = \sigma^2$ , where  $\sigma^2$  is the variance of the random field. As  $|\mathbf{r}| \rightarrow \infty$ ,  $C(\mathbf{r}) \rightarrow 0$ , hence the variogram of a stationary random field has a sill equal to the variance  $\sigma^2$  of the random field. In practice, the experimental variogram, which is estimated from the data, may show a *discontinuity* at the origin equal to  $C_0$ . This represents unresolvable fluctuations or measurement errors and is known as the nugget effect, where  $C_0$  is the nugget variance [23]. The *correlation length* or characteristic length is a normalization parameter of the distance  $\mathbf{r}$  thus defining the interval within which the field value at one point significantly affects the value at another point [45].

To emulate the variogram at any distance, theoretical variogram models are fit on the experimental variogram [7, 45]. Some commonly used theoretical variogram models include the exponential, gaussian and spherical models. Their respective equations are listed below. The symbol  $\sigma_x^2$  denotes the variance,  $\|\mathbf{r}\|$  is the Euclidean norm of the lag vector  $\mathbf{r}$ , and  $\xi$  is the correlation length.

Exponential

$$\gamma_x(r) = \sigma_x^2 [1 - \exp(-\|\mathbf{r}\|/\xi)]. \quad (2.6)$$

Gaussian

$$\gamma_x(r) = \sigma_x^2 [1 - \exp(-\|\mathbf{r}\|^2/\xi^2)]. \quad (2.7)$$

Spherical

$$\gamma_x(r) = \begin{cases} \sigma_x^2 \left[ 1.5 \left( \frac{\|\mathbf{r}\|}{\xi} \right) - 0.5 \left( \frac{\|\mathbf{r}\|}{\xi} \right)^3 \right] & \text{if } \|\mathbf{r}\| \leq \xi \\ \sigma_x^2 & \text{if } \|\mathbf{r}\| \geq \xi. \end{cases} \quad (2.8)$$

Matern

$$\gamma_x(r) = \sigma_x^2 \left[ 1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \sqrt{2\nu} \frac{r}{\xi} \right)^\nu K_\nu \left( \sqrt{2\nu} \frac{r}{\xi} \right) \right]. \quad (2.9)$$

The parameter  $\nu > 0$  is the smoothness parameter, which controls the continuity of the random field.  $\Gamma(\cdot)$  is the gamma function, and  $K_\nu(\cdot)$  is the modified Bessel function of the second kind of order  $\nu$  [67, 61].

### 2.2.3 Spartan spatial random field model

Spartan Spatial Random Fields (SSRFs) are a recently proposed family of geostatistical models with environmental applications [67, 21, 72] as well as applications with coal deposit evaluation [53, 51]. SSRFs are generalized Gibbs random fields with a kernel function that acts as a filter for the fluctuations. The term Spartan indicates parametrically compact models that involve a small number of parameters. The SSRF are defined by means of physically motivated spatial interactions between the field values [67, 31].

In three spatial dimensions the SSRF model is given by [67, 30, 33]

$$C(\mathbf{r}) = \begin{cases} \frac{\eta_0}{2\pi\Delta} e^{-\frac{|\mathbf{r}|}{\xi} \beta_2} \left( \frac{\sin(\frac{|\mathbf{r}|}{\xi} \beta_1)}{\frac{|\mathbf{r}|}{\xi}} \right) & |\eta_1| < 2, \\ \frac{\eta_0}{8\pi} e^{-\frac{|\mathbf{r}|}{\xi}} & \eta_1 = 2, \\ \frac{\eta_0}{4\pi\Delta} \left( e^{-\frac{|\mathbf{r}|}{\xi} \omega_1} - e^{-\frac{|\mathbf{r}|}{\xi} \omega_2} \right) \frac{1}{(\omega_1 - \omega_2) \frac{|\mathbf{r}|}{\xi}} & \eta_1 > 2. \end{cases} \quad (2.10)$$

In equation (2.10),  $\eta_0$  is the scale factor that determines the magnitude of

the fluctuations,  $\eta_1$  is the rigidity coefficient, and  $\xi$  is the characteristic length that determines the range of spatial correlations. The remaining coefficients are given by  $\beta_{1,2} = |2 \mp \eta_1|^{1/2}$ ,  $\omega_{1,2} = \left(\frac{\eta_1 \mp \Delta}{2}\right)^{1/2}$ , and  $\Delta = |\eta_1^2 - 4|^{1/2}$ . The correlation length of the Spartan model is determined by both  $\eta_1$  and  $\xi$ . For  $\eta_1 = 2$  the Spartan covariance of equation (2.10), the exponential model of equation (2.6) is recovered.

The SSRF model was shown to perform competitively in interpolation studies that involved radioactivity dose rate [21], coal reserves [53, 51] and groundwater level [66]. It is a flexible covariance model with two parameters  $(\eta_1, \xi)$ . Motivated by the success of this model on environmental and coal data, the Spartan Spatial Random Field (SSRF) covariance model is one of the models tested in the geostatistical analysis of this research.

#### 2.2.4 Kriging

Kriging involves a family of stochastic spatial interpolation methods that can be used to estimate the value of a random field  $X(\mathbf{u})$  at an unmeasured point  $\mathbf{u}$  by means of a linear combination of the measurements at  $n$  nearby points  $\mathbf{s}_1, \dots, \mathbf{s}_n$ . These points ideally involve all the sampling locations. If this choice leads to too-heavy computational load, the neighboring points are restricted to those found within a *kriging neighborhood* which is defined by the user. The estimation process is usually repeated at every node of a grid suitably defined for the particular application. This allows the creation of maps representing the isopleth contours of the random field. For example, if the yearly precipitation is measured in a few stations distributed in an area, kriging methods can be used to create a map of precipitation in a grid that encompasses the entire area [1]. These maps can be accompanied by an estimate of reliability, which determines the uncertainty of the estimation at each point.

The predicted value of the field at the estimation point is expressed according to the following linear combination:

$$\hat{X}(\mathbf{u}) = \sum_{i=1}^n \lambda_i X(\mathbf{s}_i). \quad (2.11)$$

In equation (2.11),  $\hat{X}(\mathbf{u})$  is the estimation at the unmeasured point  $\mathbf{u}$ , and  $\lambda_i$  are linear weights that correspond to each of the  $n$  points in the kriging neighborhood. The prediction error is  $\epsilon(\mathbf{u}) = X(\mathbf{u}) - \hat{X}(\mathbf{u})$ . In kriging methods, the linear weights  $\lambda_i$  are calculated by minimizing the error variance of equation (2.12)

$$\sigma_E^2(\mathbf{u}) = \text{Var} \left[ X(\mathbf{u}) - \hat{X}(\mathbf{u}) \right]. \quad (2.12)$$

The variance may be minimized under constraints as in the case of ordinary kriging.

This leads to a linear system of equations which is expressed in terms of the covariance or the variogram. In contrast with deterministic methods, such as inverse distance weights, Kriging provides a measure of prediction uncertainty based on the kriging variance  $\sigma_K^2(\mathbf{u})$  [39, 17, 43].

### Simple Kriging

Simple kriging (SK) is applied if the mean  $m_x$  is known and constant throughout the random field, i.e.  $E[X(\mathbf{s})] = m_x$ . If  $X'(\mathbf{s}) = X(\mathbf{s}) - m_x$  the kriging estimator is defined by the following equation:

$$\hat{X}(\mathbf{u}) = m_x + \sum_{i=1}^{n(\mathbf{u})} \lambda_i X'(\mathbf{s}_i). \quad (2.13)$$

Since  $E[X'(\mathbf{s})] = 0$ , by definition the error variance of equation (2.12) becomes

$$\sigma_{E,SK}^2(\mathbf{u}) = E \left[ \left\{ \hat{X}'(\mathbf{u}) - X'(\mathbf{u}) \right\}^2 \right]. \quad (2.14)$$

After minimization over the weights, the system of  $n(\mathbf{u})$  linear equations used to calculate the linear kriging weights  $\lambda_i$  is expressed by the system of equations shown in (2.15) [17, 39]

$$\sum_{j=1}^{n(\mathbf{u})} \lambda_j c_x(\mathbf{s}_i - \mathbf{s}_j) = c_x(\mathbf{s}_i - \mathbf{u}), \quad i = 1, \dots, n(\mathbf{u}). \quad (2.15)$$

The above system of linear equations may also be expressed as the matrix

equation

$$\mathbf{C}_{i,j}\lambda_j = \mathbf{C}_{i,u}. \quad (2.16)$$

The matrix  $\mathbf{C}_{i,j}$ , represents the covariance matrix between the data points. The vector  $\mathbf{C}_{i,u}$  represents the values of the covariance function between the sample points  $\mathbf{s}_i$  and the estimation point  $\mathbf{u}$ . Since  $c_x(0) = \sigma_x^2$ , the linear system is equivalent to

$$\begin{bmatrix} \sigma_x^2 & \dots & c_x(\mathbf{s}_1 - \mathbf{s}_n) \\ c_x(\mathbf{s}_2 - \mathbf{s}_1) & \dots & c_x(\mathbf{s}_2 - \mathbf{s}_n) \\ \vdots & \vdots & \vdots \\ c_x(\mathbf{s}_n - \mathbf{s}_1) & \dots & \sigma_x^2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} c_x(\mathbf{s}_1 - \mathbf{u}) \\ c_x(\mathbf{s}_2 - \mathbf{u}) \\ \vdots \\ c_x(\mathbf{s}_n - \mathbf{u}) \end{bmatrix} \quad (2.17)$$

The reliability of the prediction is determined by the square root of the variance of the estimation error  $\sigma_{E,SK}^2(\mathbf{u})$  which is determined from equation (2.18). The quantity  $\sigma_{E,SK}(\mathbf{u})$  is the standard deviation of the Gaussian distribution that describes the random variable  $\hat{X}(\mathbf{u})$  [23, 14].

$$\sigma_{E,SK}^2(\mathbf{u}) = \sigma_x^2 - \sum_{i=1}^{n(\mathbf{u})} \sum_{j=1}^{n(\mathbf{u})} C_{u,i} C_{i,j}^{-1} C_{j,u}. \quad (2.18)$$

### Ordinary kriging

Ordinary kriging (OK) is applied if the mean  $m_x(\mathbf{u})$  is constant but unknown inside the local neighborhood of the estimation point. The mean  $m_x(\mathbf{s})$  may vary from neighborhood to neighborhood if the ordinary kriging is not applied over the entire domain.

The unknown local mean is removed from the linear estimator by forcing the sum of the kriging weights to be equal to one. This constraint enforces the zero bias condition. The ordinary kriging estimator  $\hat{X}(\mathbf{u})$  is thus written as

$$\hat{X}(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i X(\mathbf{s}_i), \quad (2.19)$$

$$\text{with } \sum_{i=1}^{n(\mathbf{u})} \lambda_i = 1. \quad (2.20)$$

In the case of ordinary kriging, the minimum mean square error should be calculated using the constraint imposed by the zero bias condition of equation (2.20). The minimization of the error variance under the zero bias condition makes use of the Lagrange multiplier method for constrained optimization [18].

These conditions lead to the linear system of equations (2.21) and (2.22) for the linear weights, where the constant  $\mu$  is the Lagrange parameter.

$$\sum_{j=1}^{n(\mathbf{u})} \lambda_j c_X(\mathbf{s}_i - \mathbf{s}_j) + \mu = c_X(\mathbf{s}_i - \mathbf{u}), \quad i = 1, \dots, n(\mathbf{u}), \quad (2.21)$$

$$\sum_{i=1}^{n(\mathbf{u})} \lambda_i = 1. \quad (2.22)$$

The linear system of equations that gives the OK weights  $\lambda_i$  is written in the form of matrices as shown in equation (2.23)

$$\begin{bmatrix} \sigma_X^2 & \dots & c_X(\mathbf{s}_1 - \mathbf{s}_n) & 1 \\ c_X(\mathbf{s}_2 - \mathbf{s}_1) & \dots & c_X(\mathbf{s}_2 - \mathbf{s}_n) & 1 \\ \vdots & \vdots & \vdots & \vdots \\ c_X(\mathbf{s}_n - \mathbf{s}_1) & \dots & \sigma_X^2 & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} c_X(\mathbf{s}_1 - \mathbf{u}) \\ c_X(\mathbf{s}_2 - \mathbf{u}) \\ \vdots \\ c_X(\mathbf{s}_n - \mathbf{u}) \\ 1 \end{bmatrix} \quad (2.23)$$

The ordinary kriging error variance is respectively given by the equation

$$\sigma_{E,OK}^2(\mathbf{u}) = \sigma_X^2 - \sum_{i=1}^{n(\mathbf{u})} \lambda_i c_X(\mathbf{u}, \mathbf{s}_i) - \mu. \quad (2.24)$$

The Lagrange parameter  $\mu$  is always negative  $\mu < 0$ . As such,  $\sigma_{E,OK}(\mathbf{u})$  is always greater than  $\sigma_{E,SK}(\mathbf{u})$  implying greater uncertainty of OK estimations compared to SK, because of the elimination of the requirement for  $E[X(\mathbf{s})] = m$  [15, 23].

As shown in equation (2.5), for stationary random fields the variogram is connected with the covariance. If only the variogram is known the OK weights can be calculated simply by substituting  $c_X$  in equation (2.21) with  $\gamma_X$ . As such, OK weights can be calculated by the equations (2.25) and (2.26)

using only the variogram

$$\sum_{j=1}^{n(\mathbf{u})} \lambda_j \gamma_X(\mathbf{s}_i - \mathbf{s}_j) - \mu = \gamma_X(\mathbf{s}_i - \mathbf{u}), \quad i = 1, \dots, n(\mathbf{u}), \quad (2.25)$$

$$\sum_{i=1}^{n(\mathbf{u})} \lambda_i = 1. \quad (2.26)$$

Similarly, the ordinary kriging error variance is given by the equation

$$\sigma_{E,OK}^2(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i \gamma_X(\mathbf{u}, \mathbf{s}_i) - \mu. \quad (2.27)$$

### Regression Kriging

Regression Kriging (RK) combines a trend function with interpolation of the residuals. In RK the estimate is expressed as

$$\hat{X}(\mathbf{u}) = m_X(\mathbf{u}) + \hat{X}'(\mathbf{u}), \quad (2.28)$$

where  $m_X(\mathbf{u})$  is the trend function, and  $\hat{X}'(\mathbf{u})$  is the interpolated residual by means of OK [55, 14, 27].

The method of regression kriging is used in applications such as the modeling of spatial variability in precipitation [1], modeling spatial distribution of human diseases [54], and mapping of groundwater levels [68]. The advantage of the method relies is its ability to combine a trend model between the dependent variable and auxiliary variables (such as land surface parameters), and allow separate interpretation of the two interpolated components [27].

### 2.2.5 Inverse distance weights

The Inverse distance weighting (IDW) technique is a deterministic predictor [34]. The IDW methods are often applied using a neighborhood of a user-determined radius around the prediction point  $\mathbf{u}$  or a number of nearest neighbors to  $\mathbf{u}$  (i.e. 3 nearest neighbors) to determine a number of  $N$  nearby



data points with known values. This simple linear method of estimation assigns the linear weights  $\lambda_i$ ,  $i = 1, 2, \dots, N$  according to the inverse of the distance  $r_{\mathbf{s}_i, \mathbf{u}}$  of  $\mathbf{s}_i$  from the prediction point  $\mathbf{u}$ . The greater the distance of  $\mathbf{s}_i$  from  $\mathbf{u}$ , the lower the value of the linear weight  $\lambda_i$  that corresponds to the data point  $\mathbf{s}_i$  as shown in equation (2.29)

$$\hat{X}(\mathbf{u}) = \frac{\sum_{i=1}^N \frac{X(\mathbf{s}_i)}{r_{\mathbf{s}_i, \mathbf{u}}}}{\sum_{i=1}^N \frac{1}{r_{\mathbf{s}_i, \mathbf{u}}}}. \quad (2.29)$$

Variations of the method use the inverse of the distance  $r_{\mathbf{s}_i, \mathbf{u}}$  raised to a power  $n$  as shown in equation (2.30) to calculate the values of the linear weights like the *inverse distance squared weights* (IDS) method that uses  $r_{\mathbf{s}_i, \mathbf{u}}^2$  [34]

$$\hat{X}(\mathbf{u}) = \frac{\sum_{i=1}^N \frac{X(\mathbf{s}_i)}{r_{\mathbf{s}_i, \mathbf{u}}^2}}{\sum_{i=1}^N \frac{1}{r_{\mathbf{s}_i, \mathbf{u}}^2}}. \quad (2.30)$$

The benefit of using IDW methods is their computing simplicity. However the assumption that dependence of the value  $X(\mathbf{u})$  on the values of its neighbors is reduced with the square of the distance or some other power leads to a less effective predictor than the kriging predictors. IDW methods also do not give an estimate of the uncertainty of the estimation at each point [39, 28].

## 2.3 Maximum likelihood estimation

Maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model given observation data from an unknown distribution. For a given model, the MLE calculates the values of the parameters that maximizes the probability that the observations will be reproduced by the chosen spatial model [14].

Suppose there is a sample  $x_1, x_2, \dots, x_n$  of  $n$  iid observations, coming from a distribution with an unknown pdf  $f_0(x|\mathbf{k}_0)$ , where  $\mathbf{k}_0$  is the unknown true value of the parameter vector. It is desirable to find an estimator  $\hat{\mathbf{k}}$  which would be as close as possible to the true value  $\mathbf{k}_0$ .

To use the method of maximum likelihood, the joint pdf for all observations is calculated. For an independent and identically distributed sample, this joint density function is

$$f(x_1, x_2, \dots, x_n | \mathbf{k}) = f(x_1 | \mathbf{k}) \times f(x_2 | \mathbf{k}) \times \dots \times f(x_n | \mathbf{k}). \quad (2.31)$$

The likelihood function for iid variables is

$$\mathcal{L}(\mathbf{k}; x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n | \mathbf{k}) = \prod_{i=1}^n f(x_i | \mathbf{k}). \quad (2.32)$$

The method of maximum likelihood estimates  $\mathbf{k}_0$  by finding the parameter vector  $\mathbf{k}$  that maximizes the likelihood  $\mathcal{L}(\mathbf{k}; x_1, x_2, \dots, x_n)$ . Since the logarithm is a monotonically increasing function, maximizing the likelihood is equivalent to maximizing the logarithm of the likelihood function which is given by the equation (2.33) [14].

$$\ln \mathcal{L}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i | \theta) \quad (2.33)$$

### 2.3.1 Correlated random variables

The assumption in our data is that the variables of the random field are correlated, not independent as was the assumption in section 2.3. In such cases, the likelihood function shown in equation (2.32) must use the joint pdf.

If the covariance matrix of a Gaussian random field  $X_{\mathbf{s}}$  is denoted by  $C_X$  and  $E(X_{\mathbf{s}})$  is  $(\mu_1, \dots, \mu_n)$  at  $n$  locations of the field, the joint probability density function of these  $n$  random variables is then given by equation (2.34) [25]

$$f(X_1, \dots, X_n) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(C_X)}} e^{-\frac{1}{2} [(X_{\mathbf{s}} - E(X_{\mathbf{s}}))] C_X^{-1} [(X_{\mathbf{s}} - E(X_{\mathbf{s}}))]^T}. \quad (2.34)$$

In this and other cases where a joint density function exists, the likelihood function is defined as shown in equation (2.35) using the joint pdf

$$\mathcal{L}(\mathbf{k}; x_1, \dots, x_n) = f(x_1, x_2, \dots, x_n | \mathbf{k}). \quad (2.35)$$

## 2.4 Simulation

Uncertainty significantly contributes to the financial risk of mineral resources exploitation. It is introduced by sampling limitations, experimental errors, cost and price fluctuations, and geological factors such as the tectonic movements and the weathering of an area which disrupts the structure and quality features of the deposit. All these factors impact the estimation of total reserves [50]. The best available mathematical tool for exploring uncertainties and the probabilities of different scenarios is geostatistical *Monte Carlo simulation* [14, 15, 53, 46]. Conditional simulations in particular can generate different scenarios (realizations) that reproduce the statistical behavior of the spatial variability and also respect the data [36, 28]. Conditional simulation methods assume a given dataset  $D$  and a set of grid locations  $G$  where the values of the random field  $X$  need to be simulated. The simulation set is denoted by  $S = D \cup G$ .

We use the conditional simulation method which is based on covariance decomposition combined with kriging conditioning (CDKC) [57, pp. 455-459]. CDKC simulates the values of the random field  $X$  at the locations of the simulation set  $S = D \cup G$  using the following steps

1. The covariance model  $C(\mathbf{r})$  is estimated from the data.
2. A vector  $\hat{\mathbf{X}}(S) = \{\mathbf{X}(D) \cup \hat{\mathbf{X}}(G)\}$  is derived by combining the data  $\mathbf{X}(D)$  with the kriging-based estimates  $\hat{\mathbf{X}}(G)$  at the points in  $G$ .
3. The covariance matrix  $\mathbf{C}_X$  is constructed for the points in  $S$  and then the square root decomposition  $\mathbf{A}$  is evaluated, i.e.,  $\mathbf{C}_X = \mathbf{A}^T \mathbf{A}$ .
4. A random vector  $\mathbf{u}$  is generated from the standard (zero mean, unit variance) normal distribution  $N(0, 1)$ . The length of  $\mathbf{u}$  is equal to the number of points in  $S$ .
5. The unconstrained simulation  $\mathbf{X}^u = \mathbf{A} \mathbf{u}$  is generated.

6. Using the unconstrained simulation values  $\mathbf{X}^u(D)$  as data, a second application of kriging generates the interpolated vector  $\hat{\mathbf{X}}^u(G)$ .
7. We then construct the vector  $\hat{\mathbf{X}}^u(S) = \{\mathbf{X}^u(D) \cup \hat{\mathbf{X}}^u(G)\}$
8. Finally, the constrained realization on  $S$  is generated by means of the equation

$$\mathbf{X}^c = \mathbf{X}^u + \hat{\mathbf{X}} - \hat{\mathbf{X}}^u. \quad (2.36)$$

## 2.5 Spatial model validation

Validation methods provide the means for assessing the performance of different spatial models in terms of statistical measures. Validation typically involves methods that measure the predictive performance of the model based on the available data.

Assume the random field  $X_{\mathbf{s}}$  with known values at the locations  $\mathbf{s}_i$ ,  $i = 1, 2, \dots, N$ . In the *Leave-one-out cross-validation* (LOOCV or LVO) method, the value  $x(\mathbf{s}_i)$  is removed from the data set, and  $\hat{x}(\mathbf{s}_i)$  is estimated based on the remaining data. This process is repeated for all  $N$  sampling points. The performance of the predictor is assessed by comparing  $x(\mathbf{s}_i)$  with  $\hat{x}(\mathbf{s}_i)$  for  $i = 1, 2, \dots, N$  according to the validation measures presented in section 2.5.1. It can be shown that the cross-validation (CV) error estimate is an almost unbiased estimate of the true error expected on an independent test set [62].

### 2.5.1 Validation measures

The spatial model's performance is evaluated using certain statistical measures. These measures include the following: the mean error (bias) (ME), the mean absolute error (MAE), the root mean square error (RMSE), Pearson's correlation coefficient ( $\rho$ ) and Spearman's (rank) correlation coefficient ( $r_S$ ).

For the following measures,  $\hat{x}(\mathbf{s}_i)$  and  $x(\mathbf{s}_i)$  are the estimated (based on the  $N - 1$  data that do not include point  $\mathbf{s}_i$ ) and true value of the field at point  $\mathbf{s}_i$ . The quantity  $\overline{x(\mathbf{s}_i)}$  denotes the spatial average of the data and  $\overline{\hat{x}(\mathbf{s}_i)}$  the spatial average of the estimates [1].

**Mean error (bias) (ME)**

The mean error is calculated as follows:

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N [\hat{x}(\mathbf{s}_i) - x(\mathbf{s}_i)]. \quad (2.37)$$

This measure calculates the bias of the predictor. Particularly high positive or negative values of this error imply a systematic error that leads to inaccuracy.

**Mean absolute error (MAE)**

The mean absolute error is calculated as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{x}(\mathbf{s}_i) - x(\mathbf{s}_i)|. \quad (2.38)$$

This measure calculates the accuracy and precision of the predictor. All individual differences are weighted equally by MAE.

**Root mean square error (RMSE)**

The root mean square error is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N [\hat{x}(\mathbf{s}_i) - x(\mathbf{s}_i)]^2}. \quad (2.39)$$

RMSE also calculates the accuracy and precision of the predictor. Since the errors are squared before they are averaged, the RMSE gives a higher weight to large errors. As such, the RMSE is preferable to MAE when large errors are particularly undesirable.

**Pearson's correlation coefficient ( $\rho$ )**

The correlation coefficient,  $\rho$ , is the statistic that is most commonly used to summarize the relationship between two variables. The formula for Pearson's

linear correlation coefficient  $\rho$  is [35]

$$\bar{\rho}_{X,\hat{X}} = \frac{\sum_{i=1}^N [x(\mathbf{s}_i) - \overline{x(\mathbf{s}_i)}] [\hat{x}(\mathbf{s}_i) - \overline{\hat{x}(\mathbf{s}_i)}]}{\sqrt{\sum_{i=1}^N [x(\mathbf{s}_i) - \overline{x(\mathbf{s}_i)}]^2} \sqrt{\sum_{i=1}^N [\hat{x}(\mathbf{s}_i) - \overline{\hat{x}(\mathbf{s}_i)}]^2}}. \quad (2.40)$$

The correlation coefficient  $\rho$  provides a measure of the linear relationship between two variables. This relation can best be illustrated in terms of a scatterplot. If  $\rho = +1$ , the scatterplot is a straight line with a positive slope; if  $\rho = -1$ , the scatterplot is a straight line with a negative slope. For  $|\rho| < 1$  the scatterplot appears as a cloud of points that becomes more diffuse as  $|\rho|$  decreases from 1 to 0. The value of  $\rho$  is often a good indicator of how successful a linear prediction one variable from the other with a linear equation would be [1, 35].

### **Spearman (rank) correlation coefficient ( $r_S$ )**

If the relationship between two variables is not linear, the correlation coefficient  $\rho$  may prove to be a poor summary statistic. Thus, it is useful to supplement the linear correlation coefficient with another measure of the strength of the relationship, the rank correlation coefficient. To calculate the rank correlation coefficient, equation (2.40) is applied *to the ranks* of the data values rather than to the original sample values as shown in equation (2.41).

$$r_S = 1 - \frac{\sum_{i=1}^N (R_{x_i} - R_{\hat{x}_i})^2}{N(N^2 - 1)}, \quad (2.41)$$

where  $R_{x_i}$  is the rank of  $x_i$  among all the other  $x$  values. The rank is calculated by sorting the  $x$  values in ascending order; the rank of a given value is equal to its order of appearance in the sorted list.

Unlike the traditional correlation coefficient, the rank correlation coefficient is not strongly influenced by extreme pairs. Large differences between the two may be due to the location of extreme pairs on the scatterplot. Differences between  $\rho$  and  $r_S$  may also reveal that although the variables are correlated, their relation is not linear [1, 35].

## 2.6 Coal mining terms

In this section, the mining terms used in this thesis are explained briefly. Coal is a combustible black or brownish-black mineral primarily used as fuel for power stations. Coal forms when dead plant matter through lengthy biological and geological processes is converted into peat, which in turn is converted into lignite. As millions of years pass more chemical and physical changes occur and lignite increases in maturity. It is metamorphosed to sub-bituminous coal, after that to bituminous coal, and lastly anthracite, the last of the classifications used when the coal has reached ultimate maturation bibliografia.

Most coal mines are surface mines. The depth at which mining stops, either because the ore is exhausted or because mining becomes unprofitable, is called *pit bottom elevation* [34].

*Long term mine planning* refers to exploitation planning for long periods, usually over five years, such as the planning of a large mine sector. *Medium term* or mid term mine planning refers to shorter periods of time, from six months up to five years. *Continuous mining* methods involve equipment such as bucket wheel excavators and conveyor belts which allow continuous extraction and transport of the ore. *Non-continuous or cyclic mining* methods imply asynchronous extraction and transportation of the ore.

*Waste* is the earth or rock that needs to be extracted, in spite of its zero economic value, to expose the valuable product. Waste is classified as *overburden* if it lies on top of the coal (or other mined product), or *interburden* if it lies between ore seams in multiseam mines. Many coal mines are multiseam mines [34].

*Lower calorific value* (LCV) is the total energy released as heat when coal undergoes complete combustion with oxygen under standard conditions subtracting the energy required to vaporize the water content. LCV depends on several quality parameters (ash content, water content, volatile material content, etc.) and usually exhibits intense variability within the coal deposit [51].

Lignite is lower quality coal in terms of ash content and LCV. There are no universal criteria in the coal mining industry for geologically classifying a seam as lignite [4]. For Mavropigi mine, a seam is characterized as lignite

if its LCV exceeds 900 kcal/kg and the sum of CO<sub>2</sub> and ash content is less than 50% [52]. The coal of Powder River Basin has been characterized as such by the USGS [26].

The *lignite energy content per unit of surface area* (LEC area density in Gcal/m<sup>2</sup>), in this research is defined as follows

$$\text{LEC}(\mathbf{s}_p) = \rho_l \sum_{i=N(\mathbf{s}_p)}^1 d_i E_i, \quad (2.42)$$

where  $E_i$  is the LCV (Gcal/t) of lignite seam  $i$ ,  $\rho_l$  is the lignite density (1.2 t/m<sup>3</sup>), and  $d_i$  is the thickness of seam  $i$  (m).

### 2.6.1 Existing economic indices

Engineers are faced with the problem of optimizing ore production under specified economic and environmental constraints. The mining industry uses several indices that assist in this task. A short synopsis of common existing indices is presented to show the state of the art that this research seeks to supplement.

#### Discounted cash flow

Profitability is commonly measured by means of the *discounted cash flow* method (DCF) which is based on the equation

$$\text{NPV} = -C_0 + \sum_{i=1}^T \frac{Q_i(P_i - C_i)}{(1+r)^T}, i = 1, \dots, T. \quad (2.43)$$

In the above, NPV is the net present value of the mine,  $C_0$  is the investment cost,  $T$  is the expected life of the mine in years,  $i$  the time period in years,  $Q_i$  is the annual production,  $P_i$  is the price per unit ton,  $C_i$  is the production cost per unit ton, and  $r$  incorporates discount effects and risk factors [34, p. 51], [11]. DCF analysis is a simple method for estimating economic profit that provides a single value for the entire mine, but it does not resolve variations across the mine or between the interchanging ore seams and waste.



### Profitability index

The *profitability index* (PIR), refers to the revenue to investment ratio and it is defined by

$$\text{PIR} = \frac{PV}{C_0}, \quad (2.44)$$

where PV is the present value of the mine. A value of  $\text{PIR} > 1$  denotes a potentially profitable investment, whereas  $\text{PIR} < 1$  implies that the investment required exceeds the expected payoff. This index also lacks the ability to account for different mine areas and individual seams [40].

### Stripping ratio

The *stripping ratio* ( $R$ ) is determined by the volume of waste that needs to be disposed to mine a ton of ore [34, p. 389]. The marginal stripping ratio is the maximum allowable stripping ratio beyond which the operation becomes unprofitable. Pit limits can be defined by comparing the stripping ratio for each part of the deposit to the marginal stripping ratio. Hence, the stripping ratio incorporates spatial variability, but it does not account for the profitability of individual seams.

## 2.6.2 pit limit optimization

Mining is a complex operation, with respect to both technological and planning aspects. In *open-pit mining*, blocks are extracted from the ground in order to take the coal or other ore contained in them. This process leaves an increasingly deeper pit until the mining operation ends [34].

The open-pit design problem is to find an optimal ultimate contour of the pit based on estimates of mining costs, the price of coal or other ore, and physical constraints on mining precedence and maximal pit slope. A common method for pit limit optimization is the *Lerchs-Grossman algorithm* [34]. This method is based on block evaluation (taking into account revenue and costs) and technical constraints such as the maximum allowed slope. The algorithm evaluates all combinations of blocks in order to determine the optimal pit. This method and similar ones evaluate blocks, not individual seams, based on block expected revenue and costs. Thus, the application of such

methods in multi-seam mines becomes complicated unless the profitability of individual seams is taken into account during the profitability evaluation of individual blocks.

# Chapter 3

## Spatial Profitability Index

### 3.1 Introduction

This chapter focuses on the development of new geostatistical tools for the analysis of lignite energy resources. The *Spatial Profitability Index (SPI)* is introduced as a novel tool for evaluating the profitability of individual seams in open-pit multiseam mines in order to provide pit limit optimization algorithms with spatial estimates based on economic criteria. Coal price fluctuates significantly with time. For example, the price of coal used for power generation was nearly halved between January 2011 and January 2016 [5]. We use spatial interpolation to estimate the energy content locally and conditional simulations to quantify the uncertainty of the estimates. Using the SPI, it is straightforward to investigate how changes in economic factors impact the estimated reserves and to better assess the costs and revenues of different exploitation scenarios.

The SPI allows evaluating the impact of changes in market prices that could alter the profit expected from mining certain seams. It also incorporates costs due to environmental regulations that can reduce the profitability of deep seams. The SPI also helps to more effectively design the final open pit bottom elevation in different areas of the mine as the grid estimates for the LEC density are based on economic and quality criteria, thus improving the data used by pit limit optimization algorithms for block-evaluation. With regard to medium term mine planning, SPI analysis may suggest can-

celling the exploitation of deep benches or extraction using lower capacity but less costly non-continuous mining methods. We also derive a semi-empirical, explicit but mine-specific, SPI equation that captures the relation between changes of the estimated reserves and economic scenarios.

A case study of the application of SPI in the Mavropigi lignite deposit is investigated in this chapter. Lignite is a carbon-based fuel, mostly used for power generation; it is considered low-quality coal with low calorific value and low volatile content. The world-wide lignite production in 2012 was 0.9 billion tonnes whereas 3% of the global power generation was lignite-based [3, 4]. Since lignite is almost exclusively used for power generation, the most important quality parameter of lignite is its lower calorific value (LCV). Important quality characteristics of lignite include the ash content of the lignite and its sulfur content. These substances produce hazardous and potentially toxic byproducts, the treatment of which increases the cost of the power generation. These quality characteristics can show significant variation within a deposit and thus can significantly affect the profitability of sectors and individual seams of the mine. Changing product prices or environmental regulations over time complicate the estimation of which parts of the mine, seams or sectors, can be mined at a profit.

Matlab 2010b was used for implementation of the algorithms presented in this chapter. The matlab code created to calculate and implement the SPI is included in the appendix.

### 3.1.1 Mavropigi lignite deposit

The Mavropigi mining field is located in the northern part of the west boundary of Ptolemais mining area. The area of the mine covers approximately 11 km<sup>2</sup>. In the NW–SE direction, it extends over approximately 5 km from the Ptolemais power station to the Komanos village and in the NE–SW direction from the former opencast mines (North Field and Komanos) to the mountain front where pre-Tertiary slates and limestone seams occur near the surface. The mining field includes administrative buildings of the opencast mines, industrial facilities and parts of the Ptolemais power station. Inside the planned mining area, the Mavropigi village is located in the southwest of the field.

From the beginning of the mining operations at the end of 2002 until the end of 2014, 63 Mt of lignite were produced. The lignite reserves have a mean LCV of 1.412 Gcal/t. The annual lignite production of the mine is approximately 8 Mt.

The Mavropigi deposit has a multiseam structure. Besides a zone of thin lignite-bearing strata in the south, the thickness of the lignite-bearing series grows towards the northwest, where it mostly ranges between 50 and 150 m. The overlying strata thickness ranges mostly between 10 and 50 m. The mining field is separated by NW–SE striking faults into individual fault blocks. Towards the mountain front it is limited by one or several faults with dips in the direction of the mine. Furthermore, sandy–gravelly seams of the overburden strata locally incorporate consolidated seams [52, 51].

## 3.2 Spatial profitability index

This section introduces and explains the *spatial profitability index (SPI)*. In an open-pit multiseam mine the profitability of deep ore seams that are separated by large interburden waste or lie at the bottom of the deposit depend on the overlying and the underlying seams (if the latter exist). If two lignite seams are separated by a large seam of interburden, the exploitation of the lower seam may not be profitable. The removal of thick interburden seams may be economically feasible if more than one lignite seams are found below the interburden. A higher market price of the produced mineral could change which seams could be mined at a profit and consequently the estimated reserves. In contrast, updated environmental regulations or initially unforeseen technical complications could increase costs thus making unprofitable the exploitation of deeper seams.

The SPI aims to provide a quantitative tool for investigating the impact of different choices and external changes on the profitability of individual seams. Using the SPI, the change of estimated reserves according to market price and cost fluctuations can be easily investigated. This knowledge helps engineers to plan the mine and design the pit bottom elevation more effectively. It can also be used to quickly re-evaluate the reserves or the pit-bottom elevation of different mine sectors as economic conditions change. The SPI can also

be used to reevaluate medium-term exploitation plans and select specific subsectors according to the expected product demand, or to reevaluate the profitability of lower benches in light of new economic conditions.

### 3.2.1 Definition

Let the vector  $\mathbf{s} = (\mathbf{s}_p, z)$  denote the position vector within the mine, where  $z$  is the depth along the vertical dimension and  $\mathbf{s}_p = (x, y)$  the map location of the drill hole. Let  $i = 1, \dots, N(\mathbf{s}_p)$  number the lignite seams along a drill hole at  $\mathbf{s}_p$  and  $\Lambda_i$  denote the specific seam;  $i = 1$  corresponds to the top seam. In general,  $\Lambda_i$  at  $\mathbf{s}_p$  may refer to a different seam than  $\Lambda_i$  at  $\mathbf{s}_q$  due to geological discontinuities.

The *spatial profitability index (SPI)* is denoted by  $\delta_i(\mathbf{s}_p)$  for  $\Lambda_i$ ,  $i > 1$ . At  $\mathbf{s}_p$ , the SPI is defined by

$$\delta_i(\mathbf{s}_p) = \frac{\sum_{j=N'(\mathbf{s}_p)}^i P_j(\mathbf{s}_p)}{\sum_{j=N'(\mathbf{s}_p)}^i C_j(\mathbf{s}_p)}, \quad i = N'(\mathbf{s}_p), \dots, 1, \quad (3.1)$$

where  $\sum_{j=N'(\mathbf{s}_p)}^i P_j(\mathbf{s}_p)$  is the expected revenue per unit area from the exploitation of  $\Lambda_i$  and all underlying seams that could be mined at a profit, whereas  $\sum_{j=N'(\mathbf{s}_p)}^i C_j(\mathbf{s}_p)$  is the exploitation cost per unit area for the investigated seam  $\Lambda_i$  and the underlying seams that could be mined at a profit.  $N'(\mathbf{s}_p)$  is the number of seams that can be exploited at a profit in location  $\mathbf{s}_p$ .

To investigate the profitability of the first seam, certain costs should be included in  $C_1$  (costs for the first seam) like the cost of purchasing the land, equipment setup costs, defoliating the area etc.

The extraction of a lignite seam is considered economically profitable if the SPI exceeds a *threshold value*  $\delta_c$ . As such,  $\delta_c$  is the marginal SPI value at which extraction of a seam is considered profitable. Typically  $\delta_c = 1$ , except if the threshold accounts for a risk factor for unforeseen and unaccounted costs. For example, to account for 10% unforeseen costs,  $\delta_c$  could be set to

$\delta_c = 1.1$ , thus requiring expected revenues 10% higher than the expected costs for a seam to be considered profitable.

The expected revenue per unit area for seam  $\Lambda_j$  is given by

$$P_j(\mathbf{s}_p) = T E_j(\mathbf{s}_p), \quad (3.2)$$

where  $T$  is the energy price per Gcal (€/Gcal), and  $E_j(\mathbf{s}_p)$  is the energy content per unit area (Gcal/m<sup>2</sup>) for the seam  $\Lambda_j$ .

Similarly, the costs  $C_j(\mathbf{s}_p)$  are given by

$$C_j(\mathbf{s}_p) = kl_j L_j(\mathbf{s}_p) + kw_j W_j(\mathbf{s}_p) + Ce_j(\mathbf{s}_p), \quad (3.3)$$

where  $L_j(\mathbf{s}_p)$  is the weight per unit area (t/m<sup>2</sup>) of  $\Lambda_j$ ,  $W_j(\mathbf{s}_p)$  is the weight per unit area (t/m<sup>2</sup>) of the overburden —only of the single waste seam on top of  $\Lambda_j(\mathbf{s}_p)$ , not overburden of the mine—  $kl_j$  is the cost (€/t) for the extraction, transfer and storage of the lignite for seam  $\Lambda_j$ ,  $kw_j$  is the cost (€/t) of the extraction and disposal of the single overburden seam of  $\Lambda_j$ , and  $Ce_j$  accounts for additional costs for the entire seam (e.g., extra extraction costs for a seam with pyrite). The coefficients  $kl_j, kw_j$  could have values that depend on the mine sector and the depth of  $\Lambda_j$  if needed. In this study,  $kl_j$  and  $kw_j$  are considered uniform within the same sector for simplicity, i.e.,  $kl_j = kl$  and  $kw_j = kw$  for the entire sector.

We define the *extraction index* as a binary function that selects seams that could be mined at a profit, i.e.,  $I_i(\mathbf{s}_p) = 1$  for seams that can be mined at a profit and  $I_i(\mathbf{s}_p) = 0$  for waste or seams that can't be mined at a profit. The profitability index takes into account only the seams below  $\Lambda_i$ . Nevertheless, seam  $\Lambda_i$  is not extracted if it is unprofitable to extract any of the superjacent seams ( $\Lambda_j, j < i$ ). The extraction index is used to address this issue. Provided that  $I_i(\mathbf{s}_p) = 1$ , the profitability of the seam  $\Lambda_i$  at location  $\mathbf{s}_p$  is determined by the SPI. The extraction index helps to evaluate the total thickness of seams that can be mined at a profit for two-dimensional spatial analysis as well as the estimation of reserves and seams that can't be mined at a profit for three-dimensional analysis and mine planning.

### SPI and extraction index calculation algorithm

The calculation of the SPI,  $\delta$ , and the extraction index,  $I$ , per seam is given by Algorithm 1. For a given drill hole at location  $\mathbf{s}_p$  the algorithm initially sets the number of lignite seams  $N'$  that can be mined at a profit equal to the deepest seam at this location, i.e.,  $N(\mathbf{s}_p)$ .

---

**Algorithm 1** Calculation of the SPI and the extraction index for seams at location  $\mathbf{s}_p$ . For brevity the location  $\mathbf{s}_p$  is not shown explicitly.

---

```

 $N' = N$ 
for  $i = N', N' - 1, \dots, 1$  do (Begin Loop A)
     $\delta_i = \sum_{j=i}^{N'} P_j / \sum_{j=i}^{N'} C_j$ 
    if  $\delta_i < \delta_c$  then
         $N' = i - 1$ 
        for  $k = i, i + 1, \dots, N$  do (Begin Loop B)
             $I_k = 0$ 
        end for (Loop B)
    else if  $\delta_i \geq 1$  then
         $I_i = 1$ 
    end if
end for (Loop A)

```

---

The calculation proceeds sequentially to higher seams, up to the top seam. For a given seam, if  $\delta > \delta_c$ , the respective extraction index is set to one; otherwise, it is set to zero for the current *and all the underlying seams*. In the latter case, the deepest seam that could potentially be mined at a profit becomes the one above the current seam. According to Algorithm 1, a seam with  $\delta > \delta_c$  is not extracted if an overlying seam has  $\delta < \delta_c$ . Then, the extraction index is set to zero for both seams. This condition is fulfilled if the waste removal costs exceed the expected revenue from both seams. For example, assume that the seam  $\Lambda_N$  has an energy content of 10 Gcal/m<sup>2</sup> and is found just five meters below the superimposed seam  $\Lambda_{N-1}$  which has an energy content of 8 Gcal/m<sup>2</sup>. If seam  $\Lambda_{N-1}$  is buried under 100 m of interburden, the SPI may suggest that it is unprofitable to extract 105 m of interburden to exploit 18 Gcal/m<sup>2</sup>.



**Example** The schematic of Fig. 3.1 illustrates the evaluation of the SPI and the extraction index for a lignite drill hole that crosses three seams. The energy content per seam is given in Table 3.1 and the economic parameters in Table 3.2.

$E_1$	$E_2$	$E_3$
9 Gcal/m <sup>2</sup>	16 Gcal/m <sup>2</sup>	5 Gcal/m <sup>2</sup>

Table 3.1: Hypothetical energy content per lignite seam

$T$	$kl$	$kw$	$\rho_l$	$\rho_w$	$\delta_c$	$Ce_1$
7 €/Gcal	2.1 €/t	2.5 €/m <sup>3</sup>	1.2 t/m <sup>3</sup>	1.8 t/m <sup>3</sup>	1.1	4 €/m <sup>2</sup>

Table 3.2: Hypothetical economic parameters of the three lignite seams shown.

The algorithm starts at  $\Lambda_3$  by setting  $N' = 3$ . The expected revenue is  $P_3 = T \cdot E_3 = 35 \text{ €/m}^2$  and the expected cost  $C_3 = (kl \cdot L_3 + kw \cdot W_3) = 106.9 \text{ €/m}^2$ . The respective SPI is  $\delta_3 = 0.33 < 1.1$  and thus  $I_3=0$ . Since  $\Lambda_3$  is deemed unprofitable, we set  $N' = 2$  and calculate the indices for  $\Lambda_2$ . This leads to  $P_2 = 112 \text{ €/m}^2$  and  $C_2 = 32.0 \text{ €/m}^2$ . Hence, the SPI for  $\Lambda_2$  is  $\delta_2 = 3.5 > 1.1$  leading to  $I_2=1$ .

For the first seam, we have  $P_1 = 63 \text{ €/m}^2$  and  $C_1 = 91.5 \text{ €/m}^2$ . Finally the indices are calculated for the first seam as  $\delta_1 = \frac{P_1+P_2}{C_1+C_2} = 1.4 > 1.1$  leading to  $I_1=1$ . In the case of  $I_1 = 0$ , then the whole drill-hole would be unprofitable indicating that the mine should not extend there if possible.

### 3.2.2 SPI and mid-term mine planning

Using the procedure described above, the SPI can be calculated for each seam in a drill-hole. The product prices and mine costs are expected to fluctuate significantly during the mine's lifetime [5]. Similarly, the waste extraction and disposal cost may depend on the specific sector, particularly in a large mine or a mine with varying sulfur content at certain sectors. As such, knowing the SPI for each seam assists in re-evaluating whether it is

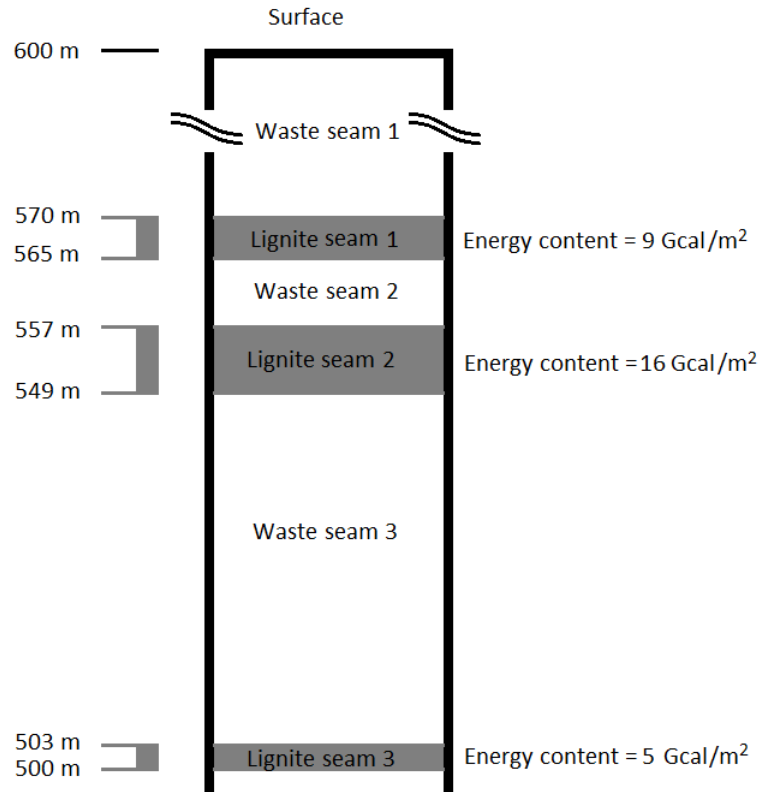


Figure 3.1: Schematic of drill-hole in multiseam lignite mine. The elevation is measured in terms of meters above sea level.

profitable to mine the investigated seam or not, under different economic situations, even over a period of three to five years.

Different values of the coefficients  $kw$  and  $kl$  of equation (3.3) can be evaluated for each sector of the mine to reflect changes in transportation costs and/or in the depth of the lignite-bearing strata or presence of pyrite in specific sectors.

Additionally, a range of values for the parameters  $kw, kl$ , the coefficient  $Ce$  that reflects other seam costs, or for the lignite price  $T$  could be investigated during mine planning. Such an effort could help to quickly re-evaluate the pit limit and the reserves held in a sector, or even the profit expected from the investigated sector in the three-year plan, under different conditions or unpredicted costs.

For example, to take into account a possible increase in waste disposal

costs, different values of  $kw$  could be investigated for a specific sector of the mine that will be exploited within three years. These scenarios could locate areas near deposit outcrops which would not be profitable to mine in case of a waste disposal cost increase, or product price drop. Further investigation would suggest whether these areas should be mined with non-continuous methods or even whether mining the lower seams should be avoided.

### 3.2.3 Supplementing thresholds for SPI

As defined in section 3.2.1, the SPI is a ratio of revenue to costs or  $\delta_i = R_i/C_i$  where  $R_i$  and  $C_i$  respectively are revenue and costs. The Profit  $P_i$  from mining seam  $\Lambda_i$  and all profitable seams below it is  $P_i = R_i - C_i$ . Hence, it is easy to prove [51] that

$$P_i(\mathbf{s}_p) = R_i(\mathbf{s}_p) \frac{\delta_i(\mathbf{s}_p) - 1}{\delta_i(\mathbf{s}_p)}, \quad (3.4)$$

and

$$C_i(\mathbf{s}_p) = \frac{R_i(\mathbf{s}_p)}{\delta_i(\mathbf{s}_p)}. \quad (3.5)$$

Critical threshold  $\delta_c$  has been defined in section 3.2.1 as the marginal value of SPI for a seam to be considered profitable. Instead of using the critical threshold  $\delta_c$ , it is possible to use the SPI with a minimum profit  $P_c$  (€/Gcal) threshold or maximum cost  $C_c$  (€/Gcal) threshold.

A maximum cost threshold  $C_c = T$  where  $T$  (€/Gcal) is the lignite price, means that the cost to extract each unit of LEC (Gcal) must be less than the price  $T$  of each such unit, i.e.  $C_c = T$  is equivalent with  $\delta_c = 1$ . Similarly, if the minimum profit threshold is  $P_c = 0$ , the profit to extract each unit of LEC must be positive which is equivalent with  $\delta_c = 1$ . Thus we can calculate the supplemental thresholds if we substitute revenue  $R_i$  in equations (3.4) and (3.5) with the lignite price  $T$  as shown in equations (3.6) and (3.7)

$$C_c = \frac{T}{\delta_c}, \quad (3.6)$$

$$P_c = T \frac{\delta_c - 1}{\delta_c}. \quad (3.7)$$

Using the SPI with the minimum profit threshold accepts seams with estimated profit per Gcal of energy equal to or greater than the minimum profit threshold  $P_c$  or

$$P_i(\mathbf{s}_p) \geq P_c. \quad (3.8)$$

Using the SPI with the maximum cost threshold accepts seams with estimated cost per Gcal of energy equal to or less than the maximum cost threshold  $C_c$  or

$$C_i(\mathbf{s}_p) \leq C_c. \quad (3.9)$$

A minimum zero profit or a maximum cost equal to the lignite price  $T$  are equivalent with  $\delta_c = 1$ . As such, the same seams would be rejected as not profitable for mining. Greater minimum profit or less maximum cost both are equivalent with  $\delta_c > 1$ .

### 3.3 Data and exploratory analysis

The data used in this part of the research are from the Mavropigi lignite mine in Greece. The lignite reserves of Greece are estimated at 4.9 billion tonnes, 3 billion of which are considered economically exploitable [4]. Most deposits (2.6 billion tonnes) are located in the north of the country. To date, only 30% of the total reserves have been extracted, and it is estimated that the existing reserves will contribute to electric power generation for at least 40 years. Lignite is supplied to seven power stations owned by the Public Power Corporation (PPC), comprising 16 generating units with a total installed capacity of 4802 MW. Greek lignite deposits usually have an average total depth of 150 to 200 m, and they typically comprise alternating seams of lignite and waste [4].

The case study comprises 10280 drill-hole core sample data from 341 vertical drill-holes provided by the PPC. For each lignite core segment the data involve surface coordinates  $\mathbf{s}_p$ , starting and ending depth (above sea level)  $z$  along the core sample, ash water-free content, and water content.

The CO<sub>2</sub> content is provided for 3 933 core samples ( $\approx 38.5\%$  of all samples), whereas the lower calorific value (LCV) is available for 4 416 core samples ( $\approx 43.3\%$ ). Missing data for the LCV are estimated using the statistical method of multiple linear regression [14]. Most of the 10 208 sampled lignite cores were also analyzed for the determination of moisture [52, 51].

In this study, each drill-hole core sample is considered to contain lignite if it has LCV value greater than 900 kcal/kg and combined ash and CO<sub>2</sub> content less than 50%. Continuous lignite core-samples in the same drill-hole are united in lignite seams. Between the lignite seams is the interburden waste [52]. This method of identifying lignite seams and interburden differs from the more complicated methodology used by PPC and doesn't take into account the technical limitations of the equipment [22]. The choice to use a simplified method of seam classification was motivated by the differing fuel quality specifications and equipment used by different power stations and mining companies.

The minimum distance between nearest-neighbor drill holes varies from 22.5m to 610.1m with the average at 138.9m. The locations of the drill holes and the mine surface limits are shown in Fig. 3.2. The map coordinates  $(x_i, y_i)$  of the drill holes are normalized by subtracting the respective mean and dividing by the standard deviation. The normalization is used to avoid numerical instabilities due to very large distance values.

The mine of Mavropigi is divided in five sectors as shown in Fig. 3.2 according to exploitation plans made by the PPC. Each of these sectors have individual values for the parameters of SPI of equation (3.3). The cost parameters for each sector have been estimated by the PPC and are given in Table 3.3. The estimation of the values takes into account the equipment, the equipment's distance from the belt conveyors, the sector's distance from the waste dump area, lignite storage yards and the pit bottom elevation as given by the PPC. The precise estimation of these parameters is beyond the scope of this thesis.

For all sectors it is assumed that lignite density is  $\rho_l = 1.2 \text{ t/m}^3$  and interburden density is  $\rho_w = 1.8 \text{ t/m}^3$ . The selling price of lignite is assumed to be  $T = 15 \text{ €/Gcal}$  in this thesis and critical threshold is set as  $\delta_c = 1.1$ .

The calorific content of the Mavropigi mine is estimated using spatial in-

<i>Sector</i>	<i>kl</i> €/t	<i>kw</i> €/m <sup>3</sup>	<i>Ce<sub>1</sub></i> €/m <sup>2</sup>
S1	2.3	2.6	4
S2	1.3	1.2	4
S3	2.4	2.8	4
S4	1.4	1.2	4
S5	2.2	2.6	4

Table 3.3: Economic parameters of the Mavropigi mine sectors. The selling price of lignite is  $T = 15$  €/Gcal and the assumed critical threshold is  $\delta_c = 1.1$ .

terpolation (based on regression kriging) and conditional simulations (based on CDKC). More precisely, we estimate the *lignite energy content per unit of surface area* (LEC area density in Gcal/m<sup>2</sup>), which is defined as follows

$$\text{LEC}(\mathbf{s}_p) = \rho_l \sum_{i=N(\mathbf{s}_p)}^1 d_i E_i, \quad (3.10)$$

where  $E_i$  is the LCV (Gcal/t) of lignite seam  $i$ ,  $\rho_l$  is the lignite density (1.2 t/m<sup>3</sup>), and  $d_i$  is the thickness of seam  $i$  (m).

The calculations are performed with two data sets: the first includes all lignite seams, whereas the second contains only seams that can be mined at a profit, as determined by means of the extraction index I and the SPI as explained in section 3.2.1. Both data sets comprise values from the 341 drill holes with LCV information. Missing LCV values from the core samples are filled in by means of multiple linear regression [52].

The basic statistics of both data sets are given in Table 3.4. SPI-corrected data are based on a threshold  $\delta_c = 1.1$ . The SPI corrected data have lower LEC area density because the area density depends on the summary of the lignite seams below the investigated area unit as shown in equation (3.10). As less profitable deeper seams are subtracted from the drill-hole, the number of seams  $N'(\mathbf{s}_p)$  that contribute to the LEC area density becomes lower. The SPI ensures that the revenue lost by rejecting those deep seams is less than the cost to mine them. As such, while the revenue decreases, the mining cost decreases more for a net profit increase.

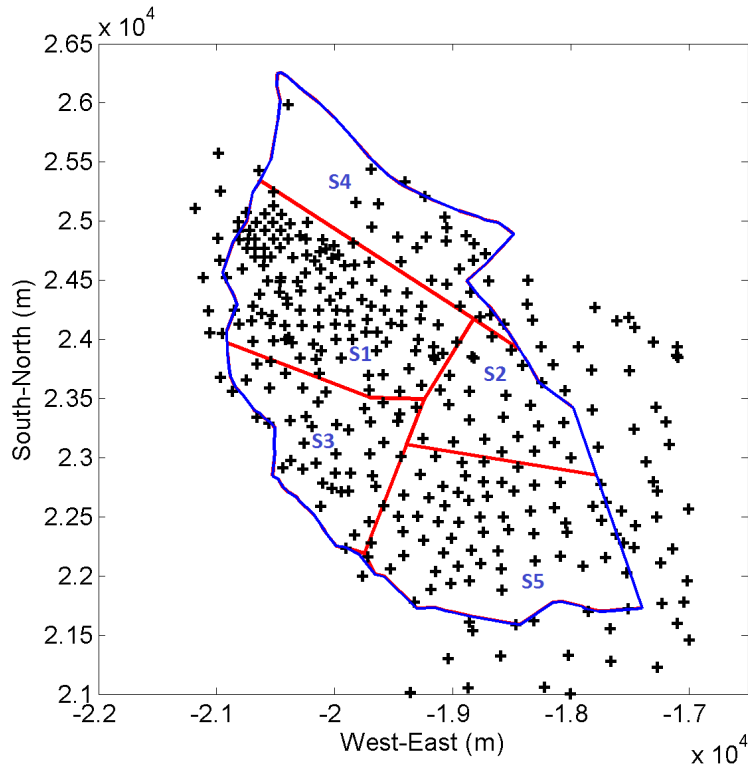


Figure 3.2: Positions of drill holes and mine boundaries (blue online) in Mavropigi mine and the lines dividing the sectors (red lines)

Na balw regression

Statistics	LEC area density	Corrected LEC area density
Mean (Gcal/m <sup>2</sup> )	39.9	37.8
Median (Gcal/m <sup>2</sup> )	35.0	33.7
Standard deviation (Gcal/m <sup>2</sup> )	27.2	28.6
Minimum (Gcal/m <sup>2</sup> )	2.6	0
Maximum (Gcal/m <sup>2</sup> )	122.6	122.6
Skewness	0.7	0.63
Sample size	341	341

Table 3.4: Statistics of lignite energy content (LEC) area density and SPI-corrected LEC area density (Gcal/m<sup>2</sup>) for Mavropigi mine. SPI-corrected data are based on a threshold  $\delta_c = 1.1$ .

### 3.3.1 Spatial data analysis

Each set of data was fit with a linear trend model. The resulting fluctuations were investigated for geometric anisotropy. In the absence of strong anisotropy the isotropic variograms and the variogram models were estimated for both detrended sets. The estimation method of regression kriging was applied to estimate the reserves.

#### Trend model

The following linear trend surface is fitted to both data sets

$$m_{LEC}(\mathbf{s}_{p,i}) = a_0 + a_1 \tilde{x}_i + a_2 \tilde{y}_i, \quad i = 1, \dots, N, \quad (3.11)$$

where  $m_{LEC}(\mathbf{s}_{p,i})$  is the mean LEC area density for drill-hole  $i$ ,  $N = 341$  is the total number of drill-holes in the area of the mine, and  $\tilde{x}_i, \tilde{y}_i$  are the normalized coordinates. Table 3.5 lists the optimal parameters of the trend model and the Pearson correlation coefficient ( $R$ ) between the trend and the data. As shown in 3.5, the linear trend model fits both data sets well although the unmodified data set performs slightly better than the SPI corrected data. The difference in the correlation coefficient can be explained because the SPI removes deeper seams from the drill holes altering the spatial distribution of LEC area density.

Parameter	LEC values	SPI-corrected LEC values
$a_0$ (Gcal/m <sup>2</sup> )	39.94	38.60
$a_1$ (Gcal/m <sup>2</sup> )	-18.57	-17.62
$a_2$ (Gcal/m <sup>2</sup> )	0.73	1.80
$R$	0.70	0.66

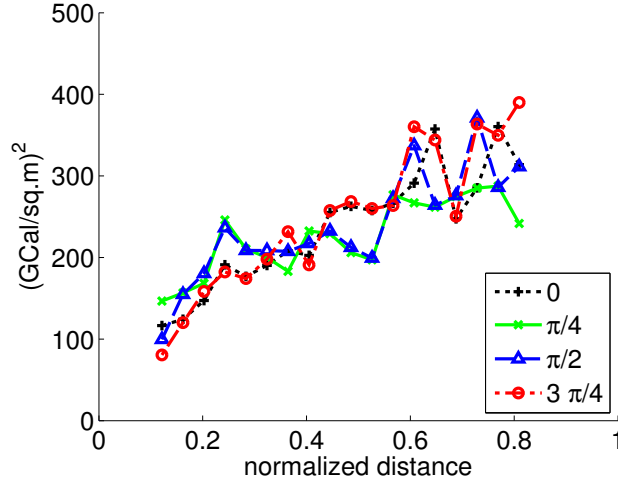
Table 3.5: Optimal parameter values for the linear trend model (3.11).  $R$  is the Pearson correlation coefficient between the data and the trend.

#### Anisotropy estimation

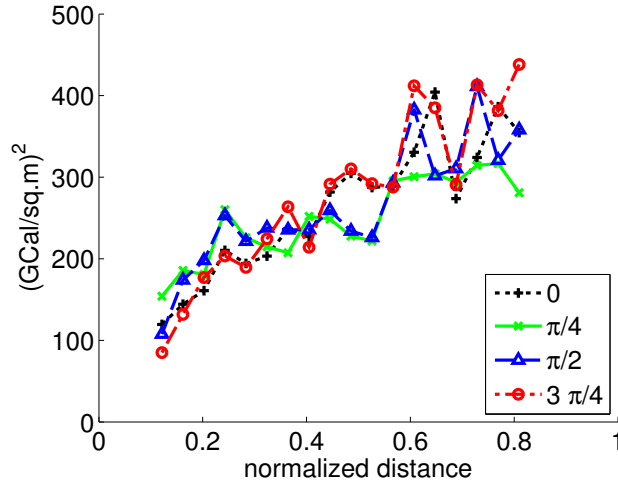
Both sets of detrended data were investigated for geometric anisotropy with the maximum likelihood estimation method (MLE) to estimate the anisotropic ratio and the direction of the anisotropy.



At first directional experimental variograms with tolerance  $\pi/4$  rad were estimated in four directions:  $0, \pi/4, \pi/2, 3\pi/4$ . The variograms for the two data sets are shown in figure 3.3. The directional variograms do not suggest the presence of strong anisotropy.



(a)



(b)

Figure 3.3: Experimental directional variograms (tolerance  $\pi/4$  rad) at the directions  $0, \pi/4, \pi/2, 3\pi/4$  for a) the PPC data and b) the SPI corrected data.

The MLE was performed on the exponential anisotropic model of equation (3.12) to estimate the parameter vector  $\theta = \{C_0, \xi_x, \xi_y, \phi\}$  that would

maximize the likelihood  $\mathcal{L}(\theta; \gamma(\mathbf{r}))$

$$\gamma_x(r) = \sigma_x^2 \left[ 1 - e^{\sqrt{\left(\frac{r_x}{\xi_x}\right)^2 + \left(\frac{r_y}{\xi_y}\right)^2 + \frac{r_x \cdot r_y}{\xi_x \xi_y} \cos(\phi)}} \right]. \quad (3.12)$$

The nugget effect is  $C_0$ , the correlation length towards the horizontal x-axis is  $\xi_x$ , towards the vertical y-axis is  $\xi_y$  and the direction of anisotropy is  $\phi$ . The ratio

$$\rho = \frac{\xi_x}{\xi_y} \quad (3.13)$$

is the *anisotropic ratio*. Table 3.6 shows the estimated anisotropic ratio and the direction of the anisotropy.

Data set	$\rho$	$\phi$
LEC values	1.06	1.44
SPI-corrected LEC values	1.03	1.51

Table 3.6: Anisotropic ratio for both detrended data sets estimated with MLE method.  $\rho$  is the anisotropic ratio estimated with MLE and  $\phi$  is the direction of the anisotropy in radians.

The MLE results are in agreement with investigation of the directional empirical variograms with a tolerance of  $\pi/4$ . It is evident that there's no strong anisotropy in either set and they can be considered isotropic.

### Variogram analysis

The omnidirectional variograms of the LEC area density for both datasets were estimated in the absence of strong anisotropy. The empirical variograms are fitted to the Spartan model of equation (2.10) with nugget effect using the weighted least squares (WLS) method [14, pp. 114-116]. In WLS the objective function  $S$  is equal to the sum of the squared differences between the empirical and model variograms, weighted by the number of pairs used at the respective lag as shown in equation (3.14)

$$S = n_i \sum_{i=1}^N (\gamma(\mathbf{s}_i) - \hat{\gamma}(\mathbf{s}_i))^2, i = 1, \dots, N, \quad (3.14)$$

where  $N$  is the total number of lags (25 in our case),  $\gamma(\mathbf{s}_i)$  is the value of the experimental variogram in distance  $\mathbf{s}_i$ ,  $\hat{\gamma}(\mathbf{s}_i)$  is the value of the variogram model at distance  $\mathbf{s}_i$ ,  $n_i$  is the number of pairs in lag  $i$ .

The WLS emphasizes empirical variogram values that are based on larger number of distance pairs. The resulting variograms (empirical and theoretical) are in good agreement as shown in Fig. 3.4. The optimal Spartan variogram parameters are listed in Table 3.7.

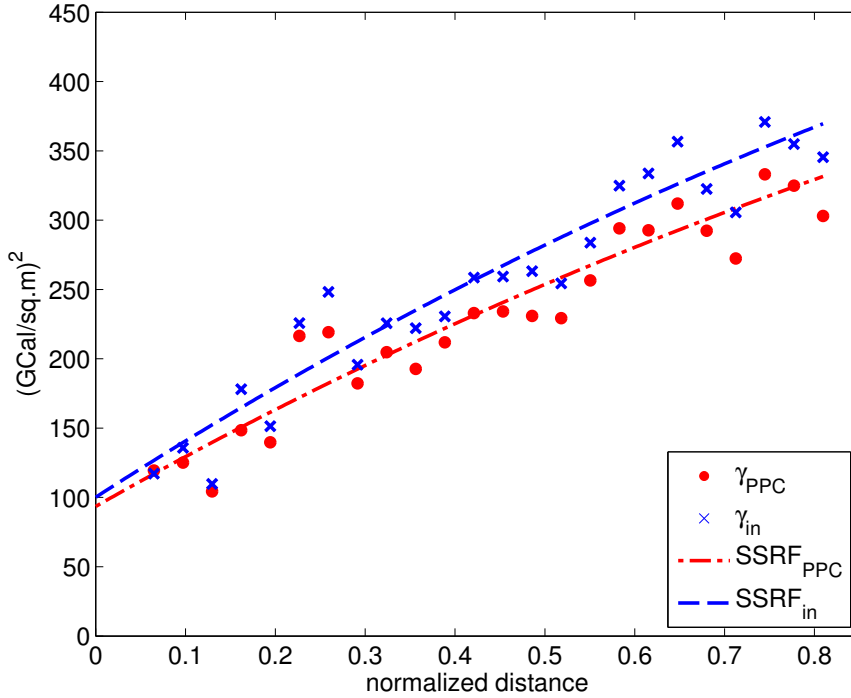


Figure 3.4: Experimental (markers) and optimal SSRF (lines) variogram models for the two LEC area density datasets. The horizontal axis represents normalized distances. The function  $\gamma_{in}(r)$  (crosses for the empirical variogram and dotted line for the SSRF model) corresponds to the SPI-corrected LEC area density, whereas  $\gamma_{PPC}(r)$  (filled circles for the empirical variogram and dash-dot lines for the SSRF model) corresponds to the LEC area density based on the mine's pit bottom elevation as determined by PPC.

SSRF parameters	$\gamma_{PPC}(r)$	$\gamma_{SPI}(r)$
$\eta_0$ (Gcal/m <sup>2</sup> ) <sup>2</sup>	$1.57 \cdot 10^4$	$9.49 \cdot 10^3$
$\eta_1$	2.00	-0.03
$\xi$	1.73	0.93
$C_0$ (Gcal/m <sup>2</sup> ) <sup>2</sup>	93.46	100.12

Table 3.7: Optimal parameters of the SSRF variogram models for LEC area density ( $\gamma_{PPC}(r)$ ) and SPI-corrected LEC area density ( $\gamma_{SPI}(r)$ ) data. The estimation of parameters was based on weighted least squares. The length parameter  $\xi$  is non-dimensional because it is expressed in normalized coordinates.

### 3.4 Regression kriging reserves estimation

The interpolation performance of the spatial models constructed in Section 3.3.1 is tested by means of cross-validation analysis. Then, regression kriging is used to generate maps of energy content and to estimate the total energy content. In these calculations both the primary and the SPI-corrected data are used. The uncertainty of the estimated energy content is investigated using conditional simulation.

#### 3.4.1 Cross-validation analysis of the spatial model

The method of leave-one-out cross-validation, described in Section 2.5 is used to measure the performance of spatial models with respect to interpolation [18]. In this method, we remove one of the LEC area density values at a time, and we use regression kriging with the optimal SSRF variogram model to estimate the missing value. This procedure is repeated for all the sample points, and the estimates are finally compared with the respective measurements at the drill holes. Common validation measures are listed in Table 3.8. The mean error measures the bias of the estimates. The mean absolute error and the root mean square error measure the overall deviation between the estimates and the measured values. Pearson's correlation coefficient  $\rho$  is a measure of linear relation between the estimates and the true values. For nonlinear dependence, Spearman's ranked correlation coefficient  $r_S$  and Kendall's  $\tau$  are also used.

The results of Table 3.8 show that kriging methods perform well for both

Measure	Model 1	Model 2
$\rho$	0.86	0.84
$r_S$	0.86	0.84
$\tau$	0.69	0.66
ME	-0.19	-0.19
MAE	9.54	10.54
MaxAE	62.67	62.65
RMSE	13.97	15.28

Table 3.8: Validation measures for regression kriging. Model 1 refers to the original data with SSRF variogram and Model 2 to SPI-corrected data with SSRF variogram model.  $\rho$  is Pearson's correlation coefficient,  $r_S$  is Spearman's correlation coefficient,  $\tau$  is Kendall's correlation coefficient, ME is the mean error (Gcal/m<sup>2</sup>), MAE is the mean absolute error (Gcal/m<sup>2</sup>), MaxAE is the maximum absolute error (Gcal/m<sup>2</sup>) and RMSE is the root mean square error (Gcal/m<sup>2</sup>).

data sets giving accurate spatial predictions. Kriging on the SPI-corrected data gives slightly less accurate predictions according to LOOCV validation measures. This is caused by the removal of deeper seems by the SPI that changes the value distribution of the data.

### 3.4.2 Estimation of the total energy content

To visualize the distribution of energy content in the mine, we interpolate the drill hole data by means of regression kriging and the spatial model defined in Section 3.3.1. The interpolation grid consists of  $232 \times 245$  rectangular cells with size 19.5 m  $\times$  22.5 m (East-West  $\times$  South-East). The generated maps of LEC area density and kriging error shown in Figs. 3.5a and 3.5b respectively. The map of LEC area density exhibits higher values in the northwest section of the mine and considerably lower values in the southeast and the northern tip of the mine. The uncertainty of the estimates, measured by the kriging error standard deviation, is inversely proportional to the sampling density. The northern tip and the southeastern end of the mine have relatively higher uncertainty than other areas.

The LEC area density prediction for each kriging cell is based on the drill-hole data evaluated by the SPI in order to determine their profitability.

As such, the LEC area density map could be utilized by pit-limit optimization algorithms, such as the Lerchs and Grossman algorithm [34], to better evaluate the profitability of the blocks selected by the algorithm.

In order to estimate the total energy content, we use a coarser cell size of  $140.6 \text{ m} \times 140.2 \text{ m}$ . This choice is guided by the mean distance between neighboring drill holes which is 139 m. Each grid cell has a specified area  $A$ , in the case of our grid  $A = 19\,712 \text{ m}^2$ , and an estimated value  $\hat{X}$  for the LEC area density. To estimate the total energy content of the entire mine or a select sector, it is straightforward to add the energy content of the cells that comprise it, as shown in equation (3.15)

$$LEC = \sum_{i=1}^N A\hat{X}_i, \quad (3.15)$$

where  $N$  is the number of cells in the area of interest.

Thus, the energy content of the Mavropigi mine based on geological lignite data and regression kriging is estimated at 353 Pcal ( $353 \cdot 10^9 \text{ Gcal}$ ). The energy content based on the SPI-corrected data, drops to 346 Pcal. Fig. 3.5c shows the LEC area density difference due to the SPI correction. This map identifies specific areas of the mine where application of the SPI has more impact. In table 3.9, the LEC for each individual sector is given for both datasets.

Sector	LEC <sub>PPC</sub> Pcal	LEC <sub>SPI</sub> Pcal
Sector 1	141	140
Sector 2	26	26
Sector 3	84	82
Sector 4	39	39
Sector 5	64	59

Table 3.9: Lignite energy content for each sector of Mavropigi mine for the original data and the SPI corrected data.

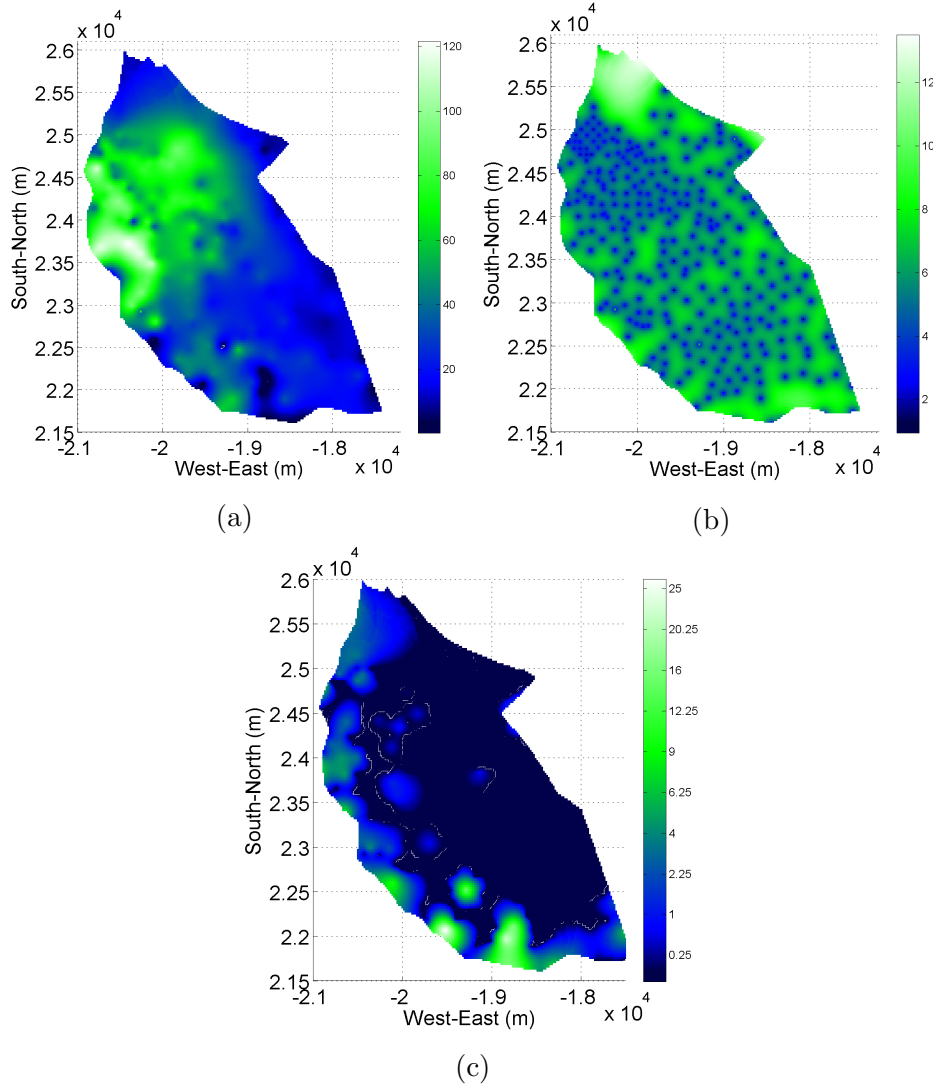


Figure 3.5: Regression kriging interpolation maps using orthogonal grid cells of dimensions  $19.5 \text{ m} \times 22.5 \text{ m}$ . (a) Map of LEC area density in Mavropogi mine based on geological lignite data. (b) Map of the kriging error standard deviation. (c) Difference between LEC area density estimates based on the original versus the SPI-corrected data.

Moments & Quantiles	Dataset 1	Dataset 2
$\bar{X}$ (Pcal)	353	346
$\sigma_X$ (Pcal)	19.0	19.9
$x_{5\%}$ (Pcal)	310	300
$x_{95\%}$ (Pcal)	397	390

Table 3.10: Simulation statistics. Dataset 1 corresponds to data without use of the SPI while dataset 2 are the data corrected by the SPI. The following symbols are used:  $\bar{X}$ : average energy content;  $\sigma_X$ : standard deviation of energy content;  $x_{5\%}$ : 5% percentile;  $x_{95\%}$ : 95% percentile

### 3.4.3 Assessment of energy content uncertainty

To assess the uncertainty of the energy content reserves and to provide respective confidence levels, we use the CDKC simulation method described in section 2.4. Five thousand constrained simulations are generated from the joint Gaussian distribution using (i) the original (without SPI correction) data and (ii) SPI-corrected data for conditioning. Further information about the parameters of the SSRF model is presented in chapter 5.6.

Table 3.10 lists the simulation statistics for both datasets. As evidenced in Table 3.10, the 90% confidence interval of the LEC without SPI correction is estimated at 310-397 Pcal. Similarly, the respective LEC confidence interval after SPI correction is estimated between 300-390 Pcal. Figure 3.6 shows the histograms of LEC obtained from the simulations. The histograms involve all realizations giving LEC values between the 5% and 95% percentiles.

## 3.5 Empirical equation for reserves correction

The SPI determines if the exploitation of a seam is profitable in comparison with the threshold  $\delta_c$  which is typically equal to one. A threshold value of  $\delta_c = 1$  means that the mining of a seam at this location to be considered profitable, its expected revenue must be at least equal to the expected costs to mine it according to revenue and cost equations (3.2) and (3.3). Since the SPI is a revenue to cost ratio, changes in the estimates of reserves are identical



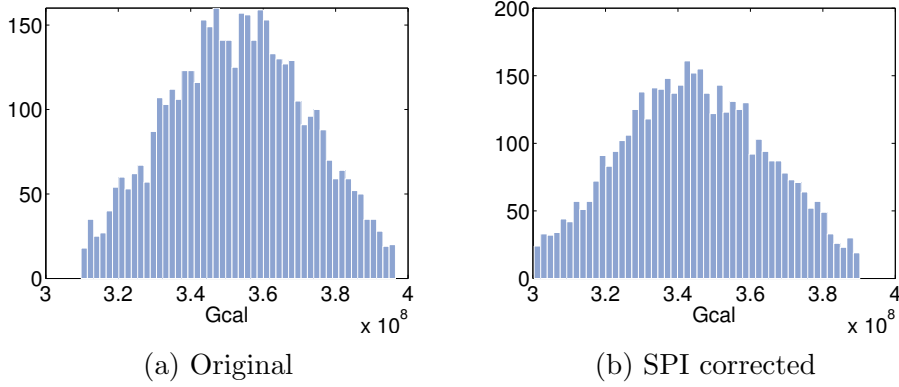


Figure 3.6: Histograms of energy content (Pcal) based on 5 000 Monte Carlo conditional simulations for (a) the original dataset and (b) the SPI-corrected data. LEC values between the 5% and 95% percentile based on the 5000 simulations are shown.

if the revenue per Gcal decreases by 20% or if the mining costs increase by 25% ( $1/1.25 = 0.8$ ) due to price changes, environmental or technical challenges, etc. since in both cases the same seams would be rejected as non-profitable. The impact of financial conditions on the profitable energy reserves can equivalently be investigated by directly evaluating the economic factors or by means of equivalent changes in the threshold  $\delta_c$ . For example,  $\delta_c = 1.25$  represents an increase of the mining costs by 25% or a decrease of the revenue by 20%. Thus, using the SPI with a critical threshold of  $\delta_c = 1.25$  is equivalent to re-evaluating the SPI with the new cost-revenue figures  $T' = 0.8 T$  (or  $k'_w = 1.25 k_w$ ,  $k'_l = 1.25 k_l$ ) and keeping  $\delta'_c = 1$ .

To investigate the dependence of LEC on  $\delta_c$ , we generate nineteen different estimates of the profitable reserves with  $\delta_c$  ranging between 0.35 and 3.05. The difference of the estimated reserves (ERD) with SPI compared to the estimated energy reserves without applying the SPI, is fitted to the sigmoid function defined in equation (3.16)

$$\text{ERD} = \frac{b_1}{1 + e^{-b_2(\delta_c - b_3)}}. \quad (3.16)$$

In equation (3.16), ERD is the estimated reserves difference of the entire mine or sector investigated,  $\delta_c$  is the critical threshold,  $b_1$  is 100% and  $b_2, b_3$  are dimensionless shape parameters. The function of equation (3.16) is just

one of the possible sigmoid functions that could be fit in the ERD results.

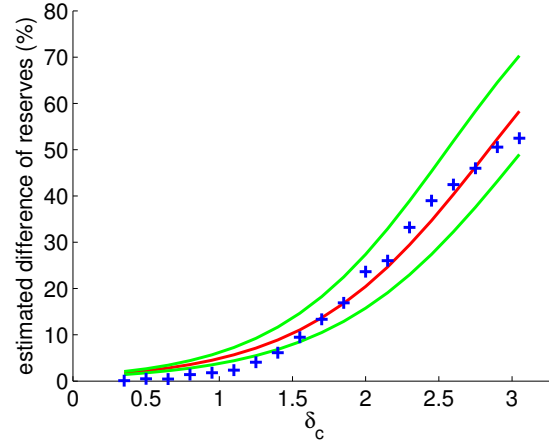
Since the bounds of the sectors are known, it is possible to estimate the ERD for each specific sector or the entire mine. The difference (in %) of the estimated reserves (ERD) compared to the estimated 354 Pcal of energy reserves is shown in Fig. 3.7a. In Fig. 3.7b the ERD of Sector 3 is shown. For the investigated sector, the lignine seams that are profitable to be mined decrease rapidly as  $\delta_c$  increases. An increase of  $\delta_c$  to  $\delta_c = 3$  could render the whole sector unprofitable.

As an example, consider a potential waste dump failure that necessitated the deposition of waste material from sector 3 several kilometers further than initially planned. Assume that this extra distance would increase the sector's cost coefficient  $kw$ , that addresses the cost of removing and dumping the waste, by 4 €/t. Such cost increase would raise  $\delta_c$  to  $\delta_c = 1.9$ . In this hypothetical scenario, nearly 40% of the sector's coal would be in seams that would be unprofitable to mine. It should then be considered in medium term planning whether it would be preferable and technically feasible to raise the pit bottom elevation of the mine abandoning the deeper seams if the waste dump failure could not be addressed.

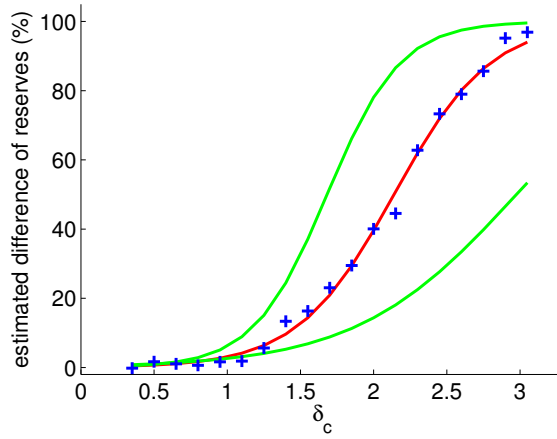
The ERD model parameters for the Mavropigi mine and sector 3 were calculated using the WLS method [14]. The ERD model parameters for the limits of the 90% confidence intervals are calculated using WLS based on 2000 simulations for each  $\delta_c$ . The estimated profitable energy reserves are then fitted with the empirical model (3.16) to determine the optimal parameters  $b_2, b_3$ ). The parameters are shown in Table 3.11. The scaling parameter  $b_1$  is 100% because the percentile of the rejected energy content is investigated. However,  $b_1$  could be set equal to the total LEC (353 Pcal for the entire mine, 84 Pcal for sector 3) if the difference in Pcal is investigated.

As can be seen in figure 3.7, although the parameters of Table 3.11 are the best fit for the equation (3.16) the resulting empirical function still doesn't fit well the reserves difference or the limits for low values of  $\delta_c$ . However the inaccuracy has small impact on the reserves difference as the errors are small (less than 5% of total LEC) in absolute value.

The empirical model allows fast estimates of the ERD due to changes of  $\delta_c$ . For example, if the revenue per Gcal sold is reduced by a third (i.e., if the



(a) Entire Mine



(b) Sector 3

Figure 3.7: Estimated difference of energy reserves versus  $\delta_c$  (markers) plotted against the theoretical model of equation (3.16) (solid red line) for the entire mine (100% is 353 Pcal) and for sector 3 (100% is 82 Pcal). The envelope defining the 90% confidence interval based on 2000 simulations is also shown (green online).

price  $T$  is reduced by about 5 € per Gcal from 15 €/Gcal to 10 €/Gcal), revenues are reduced to  $2/3$ . Hence, a lignite seam is profitable under the new conditions if it brings  $\frac{1}{2/3} = 1.5$  more revenue (i.e.,  $\delta_c = 1.5$ ) than initially planned. Based on equation (3.16) the profitable reserves are reduced by about 9% since  $\delta_c \approx 1.5$ .

Using the ERD it easy to give an estimate of the reserves of the Mavropigi mine depending on the price of the lignite instead of relying on reserves estimates calculated with different prices. Hence, while the deposit of Mavropigi is estimated to have reserves of 300 Pcal to 390 Pcal at the 90% confidence level based on the prices suggested in Table 3.2, if the lignite price drops to  $T=7.5$  €/Gcal ( $\delta_c = 2$ ) the reserves of Mavropigi drop to 240 Pcal to 290 Pcal with 90% confidence.

Parameter	Main	Lower	Upper	Main(S3)	Lower(S3)	Upper(S3)
$b_1$	100%	100%	100%	100%	100%	100%
$b_2$	1.6	1.6	1.7	3.0	1.8	4.0
$b_3$	2.8	3.1	2.6	2.1	3.0	1.7

Table 3.11: Parameters of the ERD sigmoid function model (3.16) used to estimate changes in energy reserves. “Main” corresponds to the ERD parameters of the model based on the SPI-corrected dataset for the different critical thresholds. “Lower” corresponds to the ERD parameters for the function representing the lower limit of the 90% confidence interval, whereas “Upper” corresponds to the ERD parameters for the upper limit. The parameters for Sector 3 are denoted as “Main(S3)”, “Lower(S3)” and “Upper(S3)”. The confidence intervals are based on 2 000 Monte Carlo simulations at each investigated critical threshold.

### 3.5.1 ERD and gross mine profit

As discussed in section 3.2.3, the definition of the SPI makes it possible to exchange the critical threshold  $\delta_c$  with the maximum cost threshold  $C_c$  or the minimum profit threshold  $P_c$  using equations (3.6) and (3.7) respectively. Figure 3.8 shows the ERD according to the 19 different maximum cost thresholds  $C_c$  that correspond to the ERD of the 19 different  $\delta_c$  used in section 3.5, given lignite price  $T = 15$  €/Gcal. As such the maximum cost threshold  $C_c$  ranges between 4.9 €/Gcal (equivalent of  $\delta_c = 3.05$ ) and 42.9 €/Gcal (equivalent of  $\delta_c = 0.35$ ).

To fit a theoretical model to the 19 kriging estimates of section 3.5, first we calculated the maximum cost threshold  $C_c$  that corresponds to each of the 19 thresholds  $\delta_c$ . As explained in section 3.2.3,  $C_c = \frac{T}{\delta_c}$ . Assuming lignite price  $T = 15$  €/Gcal, it is straightforward to calculate the maximum

cost threshold  $C_c$  for each of the 19 SPI thresholds  $\delta_c$ . The kriging estimates of reserves remain unchanged since as explained in section 3.2.3 the same seams would be rejected whether we use  $\delta_c$  or its equivalent  $C_c$ .

The 19 kriging estimates estimated of section 3.5 were fitted with the theoretical model of equation (3.17). Parameter  $b_1$  represents the total LEC in the area investigated regardless of profitability. Parameters  $b_2, b_3$  are shape parameters. The parameters of equation (3.17) for sector 3 and the entire mine are shown in table 3.12. Equation (3.17) allows a fast but accurate estimation for the reserves difference with different maximum cost thresholds (ERDc).

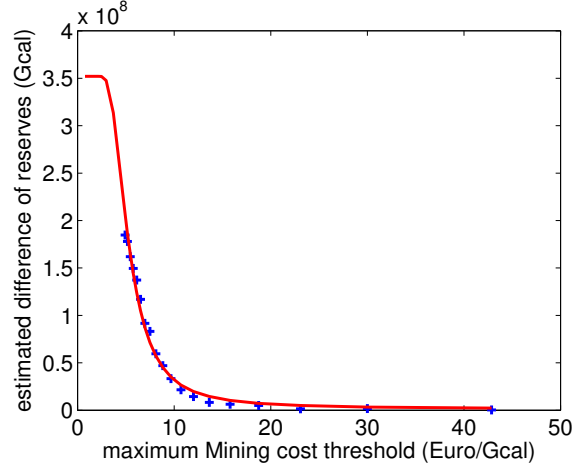
$$\text{ERDc}(C_c) = \frac{b_1}{2} \left[ 1 + \text{erf} \left( \frac{\frac{T}{C_c} - b_3}{b_2} \right) \right] \quad (3.17)$$

Coefficient	Mine	Sector 3
$b_1$ (Pcal)	353	82
$b_2$	1.6	3.0
$b_3$	2.8	2.1

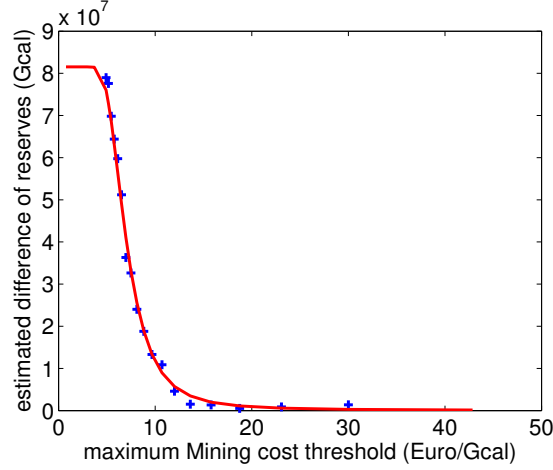
Table 3.12: Coefficients of equation (3.17) for the entire Mavropigi mine and for Sector 3.

The red line in figure 3.8 shows the LEC that would be rejected as unprofitable if the maximum mining cost allowed for mining a seam decreases. For example, if the maximum cost threshold was 10 €/Gcal, the red line in figure 3.8 shows that the seams that would be rejected as not profitable to mine, would hold approximately 30 Pcal of energy content. With increased maximum mining cost  $C_c$  (more costly seams are permitted to be mined) a greater part of the deposit is considered profitable to mine. Similarly with a decrease of  $C_c$ , only the less costly seams fulfil the condition of  $C(\mathbf{s}) \leq C_c$ . Thus, decreasing  $C_c$  results in a smaller part of the deposit considered profitable to mine.

Consider a small increase in the maximum cost threshold from a starting  $C_c = C$  (black vertical line in figure 3.9), equal to  $\delta C$ , to  $C'_c = C + \delta C$  (green vertical line in figure 3.9). As such, a small part of the deposit  $\delta R$ , that with maximum cost threshold  $C_c = C$  was considered unprofitable to mine,



(a) Entire mine



(b) Sector 3

Figure 3.8: Estimated difference of energy reserves versus  $C_c$  (markers) plotted against the theoretical model of equation (3.17) (solid line) for the whole mine and Sector 3, assuming lignite price  $T = 15$  €/Gcal.

becomes profitable (marginally) as the less strict threshold  $C'_c = C + \delta C$  allows more costly seams to be considered reserves. This means that the part of the reserves  $\delta R$  have a mining cost between  $C$  and  $C + \delta C$  since they are only considered profitable if the maximum mining cost threshold increases to  $C'_c = C + \delta C$  and they are not considered profitable if it is just  $C_c = C$ , .

As  $\delta C \rightarrow 0$  so does  $\delta R \rightarrow 0$ . Hence for a very small increase  $dC$  of the

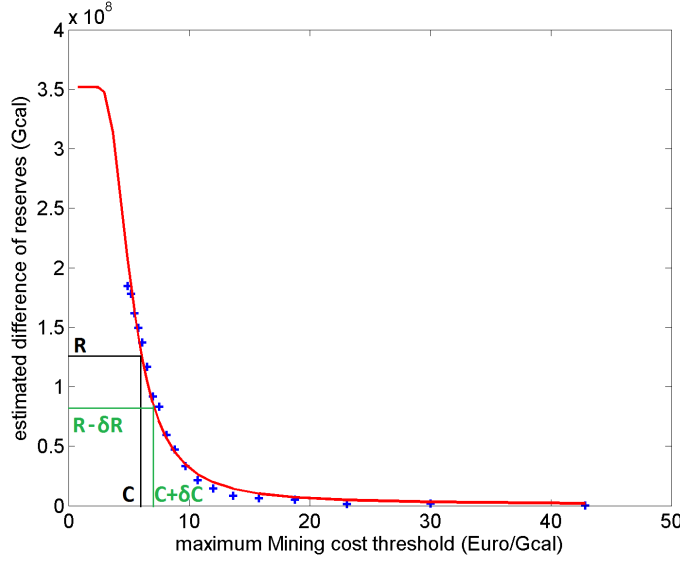


Figure 3.9: Estimated difference  $\delta R$  in the reserves of the Mavropigi mine with an increase of the maximum mining cost threshold from  $C_c = C$  to  $C'_c = C + \delta C$ .

maximum cost threshold from  $C_c = C$  to  $C'_c = C + dC$ , a very small part of the deposit  $dR$  will become profitable. Thus, this part  $dR$  of the deposit has a mining cost between  $C$  and  $C + dC$ , which can be considered to be  $C$  since  $dC \rightarrow 0$ . Hence the red line in figure 3.9 corresponds to the marginal cost threshold  $C$  at which the expected mining costs to mine this  $dR$  part of the deposit equal the expected revenue from this  $dR$ , with the assumed cost parameters  $kw$ ,  $kl$ ,  $Ce_1$  and lignite price  $T$  [51].

Since the ERDc line in Figure 3.9 corresponds to the mining costs, the area below of the ERDc line corresponds to the *total mining costs* for all the lignite seams from Mavropigi mine with given parameters  $kw$ ,  $kl$ ,  $Ce_1$  and  $T$ . Equation (3.18) gives the total mining costs with the given cost parameters.

$$MC = \int_0^{+\infty} dC \cdot \text{ERDc}(C_c) \quad (3.18)$$

In equation (3.18),  $MC$  is the total mining costs for extracting all the lignite within the mine limits. Equation (3.18) can be used with the entire mine or for individual sectors.

In the case of the Mavropigi mine with assumed  $kw$ ,  $kl$ ,  $Ce_1$  and  $T$ , equation (3.18) gives  $MC=2.31 \cdot 10^9$  € for the entire mine and for sector 3,  $MC=0.64 \cdot 10^9$  €. With the mining costs and total LEC estimated and given the price per Gcal, it is straightforward to estimate the total gross profit (MP) with equation (3.19) where MP does not include taxes, depreciation, etc. as

$$MP = (R - ERDc)T - MC \quad (3.19)$$

where  $R$  is the estimated total LEC of the lignite deposit.

In the case of Mavropigi mine equation (3.19) gives  $MP=3.0 \cdot 10^9$  € and for sector 3,  $MP=0.58 \cdot 10^9$  €. It is noted that the actual profit is significantly lower than the total gross profit, since the LEC of the lignite that fulfils the quality criteria of the stations is less than  $R$ , and the taxation, depreciation as well as other costs have not been taken into account. As an example, using a tax rate of 20% on the profits and that about 75% of the lignite would have the quality characteristics required to be sent to the stations, the profit from Mavropigi mine is expected at  $1.3 \cdot 10^9$  €.

### 3.6 SPI-based pit bottom elevation

In this section, the suggested changes to pit bottom elevation by using the SPI will be presented. As explained in section 3.5, increasing the critical threshold  $\delta_c$  causes progressively shallower seams to be rejected by the extraction index. Thus, higher values of  $\delta_c$  raise the pit bottom elevation for seams that can be mined at a profit.

Six different  $\delta_c$  thresholds from the 19 investigated in section 3.5 are presented to illustrate the progressive change in suggested pit bottom elevation. The scenarios selected have thresholds that range between  $\delta_c = 0.35$  to  $\delta_c = 2.6$ .

The maps in Fig. 3.10 show the suggested pit bottom elevation for the different  $\delta_c$  selected. The estimated bottom elevation does not incorporate technical constraints reflecting slope restrictions and the presence of faults. Parts of the mine that are rejected as holding no profitable lignite seams are



in some cases in the interior of the mine area, not the borders, which would complicate their exclusion. However, these maps can serve as a guideline to quickly re-evaluate the profitability of lower benches in case of price changes. Substantial changes to the pit bottom elevation based on the SPI-evaluated seam profitability can be used to locate sectors that should be re-evaluated using pit-limit optimization algorithms. For example, in the case of  $\delta_c = 2.6$  in Fig. 3.10f, large parts of sector 5 of Mavropigi are rejected as non-profitable, suggesting a re-evaluation of pit-limits or even a reconsideration of whether the sector should be mined.

The above maps can be used to assist in designing subsequent phases of the exploitation in case of revenue and cost fluctuations during the course of the mine's life along with technical considerations. For example, a 50% drop in the price of coal during a period of 15 years from the original design, is equivalent to raising the threshold to  $\delta_c = 2$ . In this scenario application of the SPI would suggest a significant raise of the pit bottom elevation in the south-western part of the mine in order to avoid mining seams the exploitation of which is no longer profitable. However, an increase of the price of energy and more efficient mining equipment that would be equivalent to the threshold dropping to  $\delta_c = 0.5$  would make the exploitation of deep seams in the western area of Mavropigi mine profitable.

### 3.7 Summary

In this section, the spatial profitability index (SPI) that locally compares costs and revenue for individual lignite seams was defined. The SPI was illustrated using drill-hole data for the Mavropigi lignite mine. The resulting changes in the reserves due to different economic criteria, both globally and in specific sectors of the mine were investigated using regression kriging. The SPI is herein defined for multiseam geological structures, but it can be extended to energy reserves with continuously varying distributions.

The SPI is a straightforward mathematical tool with spatial resolution abilities that can assist in the medium term and long term planning of energy reserves exploitation conducted with pit-limit optimization algorithms. It enables engineers to investigate changes of estimated energy reserves in

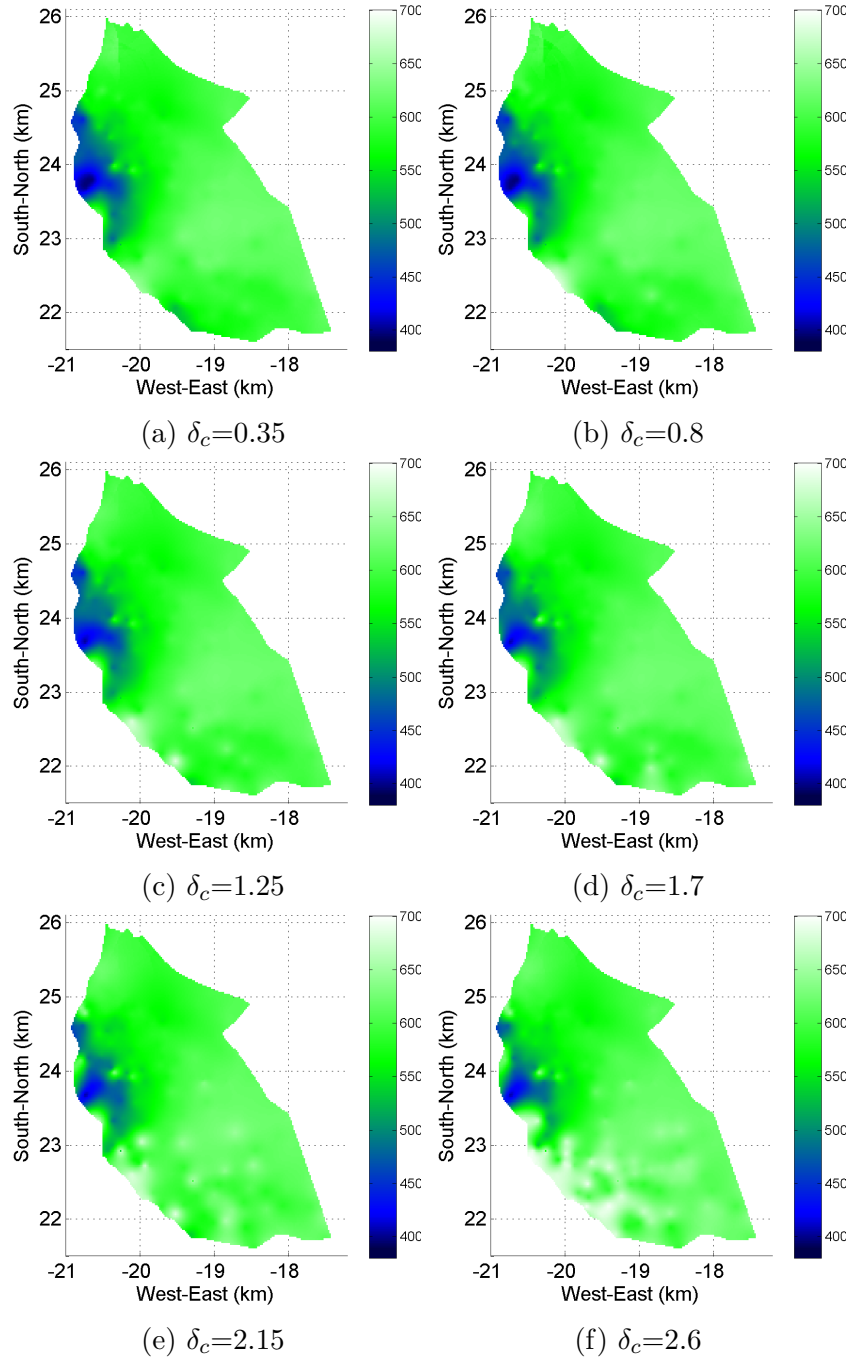


Figure 3.10: Suggested open pit bottom elevation (height above sea level) for various  $\delta_c$  values. The bottom elevation shown does not account for technical constraints.

response to market price trends (such as the coal price drop between 2011 and 2016 [5]) or increased extraction and processing costs. For example, the SPI can help to more effectively determine the total mining cost for sectors or the entire mine.

The SPI allows identifying non-profitable or marginally profitable sectors that can be designated for exploitation using non-continuous, cost-effective methods. With regard to medium term mine planning, the SPI can help to re-evaluate the exploitation of deep benches in response to changes in costs or market prices by providing LEC density estimates based on economic parameters. Thus, the SPI can assist block-evaluation algorithms in determining pit limits for different scenarios. If energy demand is expected to rise or if there is a temporary need to increase revenue, the SPI can be used to determine the most profitable areas for immediate exploitation.

Sigmoid model functions are proposed that can be used to quickly estimate the change of profitable reserves under different economic scenarios and can generate fast, accurate estimations of total mining costs for the entire mine or for specific sectors. The advantage of this approach is that spatial analysis needs to be performed only once to determine the model parameters; following that initial step, the sigmoid functions can be used to investigate different scenarios without repeating the detailed spatial analysis. As such, the energy reserves of the mine can be given at any time depending on current product prices instead of relying on analysis conducted several years ago with possibly very different economic conditions.



# Chapter 4

## Stochastic Local Interaction Model

### 4.1 Introduction

In this chapter, the Stochastic Local Interaction (SLI) model will be reviewed. This new method will be applied to three different case studies and its performance will be compared with that of kriging methods. The first data set investigated is the Mavropigi lignite energy content data from chapter 3. In particular, we use the lignite data generated by application of the Spatial Profitability Index (SPI) as described in Section 3.3. The second case study involves a large dataset from Campbell county, Wyoming, in the Powder River Basin of the USA. The third case study involves a non-Gaussian data set which comprises grayscale pixel values from a digital image of the dwarf planet Pluto. While the third dataset does not involve coal deposits, it is employed in this thesis to investigate the performance of SLI methods in non-gaussian datasets and compare it with the performance of kriging.

### 4.2 Review of SLI theory

The *Stochastic Local Interaction*(SLI) model, employs a local representation which can improve the computational efficiency of spatial prediction for correlated data [29]. It is based on a joint probability density function which

is defined by an energy functional that involves local interactions between neighboring sites. The energy functional represents the "probability cost" for specific spatial configurations (patterns) of field values, and it does not correspond to actual energy levels. This energy functional implements the local interactions by means of kernel functions [2] with locally adaptive kernel bandwidths. The SLI is expressed mathematically by means of the respective *precision matrix* (i.e. the inverse of the covariance matrix). This representation leads to a semi-analytical expression for prediction of the values of random fields. This prediction avoids the computationally costly inversion of the covariance matrix required by kriging methods.

A kernel function is an integrable, non-negative function. For a function  $K(u)$  to be considered a kernel function, it must satisfy the following requirements [2]:

- The function  $K(u)$  must be a non-negative and symmetric function ( $K(-u) = K(u)$  for all values of  $u$ ).
- $\int_{-\infty}^{+\infty} K(u) du = 1$ .
- If  $K(u)$  is a kernel function, so is the function  $K_\lambda(u) = \lambda K(\lambda u)$ ,  $\forall \lambda > 0$ .

Based on the kernel property of  $K_\lambda(u) = \lambda K(\lambda u)$ , if  $\lambda$  is a linear weight, kernels can be used as weighted functions in estimation techniques. Kernel functions are often used in transformations as shown in equation (4.1).

$$T = \int_{t_1}^{t_2} dt K(t)f(t) \quad (4.1)$$

### 4.2.1 Notation

We use the following notation in relation with the SLI model. The sample data are denoted by means of the vector  $\mathbf{X}_s$  which involves the field values at the sample locations  $\mathbf{s}_i$ , where  $i = 1, \dots, N$ . The predictions are denoted by the prediction vector  $\hat{\mathbf{X}}_{sp}$  which involves the unknown field values at the prediction locations  $\mathbf{s}_p$ , where  $p = 1, \dots, P$ .

The pdf  $f_X$  of a Gibbs SRF can be expressed in terms of an energy functional  $H(\mathbf{X}_s; \theta)$ , where  $\theta$  is a set of model parameters as shown in equation (4.2)

$$f_X(\mathbf{X}_s; \theta) = \frac{e^{-H(\mathbf{X}_s; \theta)}}{Z(\theta)}. \quad (4.2)$$

The constant  $Z(\theta)$  is called the *partition function*. It represents the normalization factor of the pdf, obtained by integrating  $e^{-H(\mathbf{X}_s; \theta)}$  over all possible values of the data vector  $\mathbf{X}_s$ .

In order to define the local interactions in the energy functional, we need to employ the *kernel bandwidth* which is related to the range of influence of the kernel functions. The range of influence determines how far the interactions spread around each point. Assume the sampling points  $\mathbf{s}_i, \mathbf{s}_j$  with Euclidean distance  $r_{i,j} = \|\mathbf{s}_i - \mathbf{s}_j\|$ , where  $i, j = 1, \dots, N$ . The kernel bandwidth for each data point of  $\mathbf{s}_i$ , where  $i = 1, \dots, N$  adapts to local variations of the sampling pattern. The kernel bandwidth  $h$  is given by the equation (4.3)

$$h_i = r_{i,k} \cdot \mu. \quad (4.3)$$

The value  $h_i$  represents the kernel bandwidth for the sampling point  $\mathbf{s}_i$ . The model parameter  $\mu$  determines the local bandwidth. The distance  $r_{i,k}$  represents the distance between the sampling point  $\mathbf{s}_i$  and its  $k$ -nearest neighbor. The neighbor order  $k$  is usually chosen as 1 (nearest neighbor) for infinitely supported kernels and as 2 (second nearest neighbor) for compactly supported kernels [29].

If the normalized distance between the sampling points  $\mathbf{s}_i, \mathbf{s}_j$  is defined as  $u_{i,j} = r_{i,j}/h_i$ , then kernel weights associated with each pair of points are defined as shown in equation (4.4) [29].

$$\lambda_{i,j} = K(u_{i,j}) \quad (4.4)$$

Depending on the sampling grid, it is generally possible for the  $k$ -nearest neighbor of  $\mathbf{s}_i$  to be the sample point  $\mathbf{s}_j$ , while the  $k$ -nearest neighbor of  $\mathbf{s}_j$  could be a different sample point  $\mathbf{s}_l$ . In such a case  $h_i \neq h_j$ . As such,

$u_{i,j} \neq u_{j,i}$  and  $\lambda_{i,j} \neq \lambda_{j,i}$ .

Examples of kernel functions are given in Table 4.1 [2]

Name	Function
Triangular	$K(u) = (1 -  u ) \cdot I_{ u  \leq 1}(u)$
Quadratic	$K(u) = (1 - u^2) \cdot I_{ u  \leq 1}(u)$
Tricube	$K(u) = (1 - u^3)^3 \cdot I_{ u  \leq 1}(u)$
Exponential	$K(u) = \exp(- u )$
Gaussian	$K(u) = \exp(-u^2)$

Table 4.1: Examples of Kernel functions [2]. The normalized distance is  $u = r/h$ , where  $r$  is the Euclidean distance and  $h$  is the kernel bandwidth.  $I_A(u)$  is the indicator function of set  $A$ .  $I_A(u) = 1$ , if  $u \in A$  and  $I_A(u) = 0$ , if  $u \notin A$ .

If  $G_X(O)$  is any two-point function that depends on the the locations or values of field  $\mathbf{X}$  ( $O$  could be the distance between two points, the difference between field values etc.) and  $\mathbf{h}$  is the vector of the local bandwidths, we use a *local bandwidth extension* over the network of sampling points [44]

$$\langle G_X(O) \rangle_{\mathbf{h}} = \frac{\sum_{i=1}^N \sum_{j=1}^N K_{i,j} G_X(O)}{\sum_{i=1}^N \sum_{j=1}^N K_{i,j}}. \quad (4.5)$$

This notation will be used in relation with the SLI model in section 4.2.2.

### 4.2.2 The SLI model

Consider a sample  $\mathbf{X}_s$  at the sample locations  $\mathbf{s}_i$ , where  $i = 1, \dots, N$ . The average of the square fluctuations is

$$S_0(\mathbf{X}_s) = \frac{1}{N} \sum_{i=1}^N (x_i - m_X)^2, \quad (4.6)$$

where  $m_X$  is the mean of  $\mathbf{X}$ .

The average of the square gradient in an Euclidean space of dimension  $d$  is defined as

$$S_1(\mathbf{X}_s; \mathbf{h}) = d \langle x_{ij}^2 \rangle_{\mathbf{h}}, \quad (4.7)$$



where the *field increments*  $x_{ij} = x_i - x_j$ ,  $i, j = 1, \dots, N$  refer to the measured sample values at the sampling locations.

The average of the square curvature in an Euclidean space of dimension  $d$  is defined as

$$S_2(\mathbf{X}_s; \mathbf{h}) = c_1 \langle x_{ij}^2 \rangle_{\mathbf{h}} - c_2 \langle x_{ij}^2 \rangle_{\mathbf{h}\sqrt{2}} - c_3 \langle x_{ij}^2 \rangle_{2\mathbf{h}}, \quad (4.8)$$

where the coefficients are given by  $c_1 = 4d(d+2)$ ,  $c_2 = 2d(d-1)$  and  $c_3 = d$ . The values of these coefficients are determined so that the respective terms match the coefficients of the mean square curvature in the continuum case as shown in [33].

For the SLI model, the energy functional  $H_X(\mathbf{X}_s; \theta)$ , of equation (4.2) is expressed as follows

$$H_X(\mathbf{X}_s; \theta) = \frac{1}{2\lambda} [S_0(\mathbf{X}_s) + a_1 S_1(\mathbf{X}_s; \mathbf{h}) + a_2 S_2(\mathbf{X}_s; \mathbf{h}),] \quad (4.9)$$

where  $\theta = (m_X, a_1, a_2, \lambda, \mu, k)$  is the SLI parameter vector,  $\mathbf{h}$  is the vector of the local bandwidths and  $m_X$  is the mean of the sample. The coefficients  $a_1, a_2$  control the relative contributions of the mean square gradient and mean square curvature terms. The coefficient  $\lambda$  controls the overall amplitude of the fluctuations. Finally,  $\mu$  and  $k$  control the bandwidth values  $\mathbf{h}$  as described in section 4.2.1.

The joint pdf of the SLI model is determined by means of a specific energy functional which is given in equation (4.9). This leads to a precision matrix which is explicitly defined in terms of local interactions and thus avoids the computationally costly covariance matrix inversion. The prediction is based on maximizing the joint pdf of the data and the prediction point, which is equivalent to minimizing the corresponding energy functional [29].

The energy functional is permissible if  $H_X(\mathbf{X}_s; \theta) \geq 0$  for all  $\mathbf{X}_s$ . This condition ensures that the precision matrix, and thus its inverse, the covariance matrix, are non-negative definite. As  $S_0$  and  $S_1$  are always non-negative, provided that  $S_2 \geq 0$  a sufficient permissibility condition is  $a_1, a_2, \lambda \geq 0$  [29].

### 4.2.3 Precision matrix formulation

The energy functional of equation (4.9) can be equivalently expressed by means of the *precision matrix*  $\mathbf{J}(\theta)$  as follows:

$$H_X(\mathbf{X}_s; \theta) = \frac{1}{2}(\mathbf{X}_s - m_X)^T \mathbf{J}(\theta)(\mathbf{X}_s - m_X). \quad (4.10)$$

In equation (4.10), the symmetric precision matrix  $\mathbf{J}(\theta)$  is given by the following equation na pw brackets

$$\mathbf{J}(\theta) = \frac{1}{\lambda} \left\{ \frac{\mathbf{I}}{N} + a_1 d\mathbf{J}_1(\mathbf{h}) + a_2 \left[ c_1 \mathbf{J}_2(\mathbf{h}) - c_2 \mathbf{J}_3(\mathbf{h}\sqrt{2}) - c_3 \mathbf{J}_4(2\mathbf{h}) \right] \right\}, \quad (4.11)$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{J}_q$  ( $q = 1, \dots, 4$ ) are *network matrices* that are determined by the sampling pattern, the kernel function, and the kernel bandwidths.

The elements of the network matrices are given by equation (4.12) below

$$[\mathbf{J}(\mathbf{h}_q)]_{i,j} = -u_{i,j}(h_{q,i}) - u_{i,j}(h_{q,j}) + \mathbf{I}_{i,j} \sum_{b=1}^N [u_{i,b}(h_{q,i}) + u_{b,i}(h_{q,b})], \quad (4.12)$$

where  $\mathbf{h}_q = \mathbf{h}$ , for  $q = 1, 2$ ,  $\mathbf{h}_3 = \sqrt{2}\mathbf{h}$  and  $\mathbf{h}_4 = 2\mathbf{h}$  and the weights of the network matrices are given by equation (4.13) below

$$u_{i,j}(h_{q,i}) = \frac{K\left(\frac{\mathbf{s}_i - \mathbf{s}_j}{h_{q,i}}\right)}{\sum_{i=1}^N \sum_{j=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_j}{h_{q,i}}\right)}. \quad (4.13)$$

### 4.2.4 Model estimation

The SLI model of equation (4.9) requires the parameter vector  $\theta = (m_X, a_1, a_2, \lambda, \mu, k)$  described in section 4.2.2. The parameter  $k$  for the  $k$ -nearest neighbor is set to  $k = 1$  for infinitely supported kernels and as  $k = 2$  (second nearest neighbor) for compactly supported kernels. The remaining model parameters need to be estimated from the data.

The expected value  $m_X$  of the random field  $X$  is calculated from the mean value of the sample. The optimal values for the remaining parameters  $a_1, a_2, \mu$  can be estimated by means of maximum likelihood estimation. Alternatively, they can be estimated by minimizing the cross validation functional

$$\Phi(\mathbf{X}_s; \theta) = \sum_{i=1}^N |\hat{x}_i(\theta) - x_i|, \quad (4.14)$$

where  $\hat{x}_i(\theta)$  is the SLI prediction at point  $\mathbf{s}_i$ ,  $i = 1, \dots, N$  using the tested parameter vector  $\theta$  [29]. The parameter  $\lambda$  controls the overall amplitude of the fluctuations as normalizing parameter and has a value of

$$\lambda = \frac{\frac{1}{N} + a_1 d\mathbf{J}_1(\mathbf{h}) + a_2 [c_1 \mathbf{J}_2(\mathbf{h}) - c_2 \mathbf{J}_3(\mathbf{h}\sqrt{2}) - c_3 \mathbf{J}_4(2\mathbf{h})]}{N}, \quad (4.15)$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{J}_q$  ( $q = 1, \dots, 4$ ) are the network matrices defined in section 4.2.3. The coefficients  $c_1, c_2, c_3$  are defined in section 4.2.2.

We chose to use the method of minimizing the validation functional of equation (4.14) because it is faster than MLE. The initial parameters  $a_1, a_2, \mu$  for the starting vector  $\theta_0$  used for the estimation of the SLI model parameter vector  $\theta$  in this chapter were  $a_1 = 1, a_2 = 2, \mu = 1$ . The minimization of the validation functional of equation (4.14) was performed using the function *fmincon* of matlab 2012b.

#### 4.2.5 Prediction with SLI

Assume that the prediction point  $\mathbf{s}_p$  is added to the sampling points. This point is then inserted in the energy functional. Then, the mode (the most probable value) of the joint pdf with the prediction point inserted is determined by finding the value of the predictand that maximizes the joint pdf. The calculations can be carried out analytically and they lead to the following mode prediction equation

$$\hat{x}_p = m_x - \frac{\sum_{i=1}^N J_{i,p}(\theta)(x_i - m_x)}{J_{p,p}}. \quad (4.16)$$

In equation (4.16),  $\hat{x}_p$  represents the prediction,  $(x_i - m_x)$  is the fluctuation at the location  $\mathbf{s}_i$  and  $J_{i,p}$ ,  $J_{p,p}$  represent the elements of the precision matrix. The elements that involve the prediction point are given by the following equations [29]

$$[\mathbf{J}(\mathbf{h}_q)]_{p,p} = \sum_{i=1}^N [u_{i,p}(h_{q;i}) + u_{p,i}(h_{q;p})], \forall p = 1, \dots, P \quad (4.17)$$

$$[\mathbf{J}(\mathbf{h}_q)]_{i,p} = -[u_{i,p}(h_{q;i}) + u_{p,i}(h_{q;p})], \forall p = 1, \dots, P \text{ and } \forall i = 1, \dots, N \quad (4.18)$$

The weights used in the precision matrices are given by the following expressions

$$u_{i,j}(h_{q;i}) = \frac{K\left(\frac{\mathbf{s}_i - \mathbf{s}_j}{h_{q;i}}\right)}{\sum_{i=1}^N \sum_{j=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_j}{h_{q;i}}\right) + \sum_{i=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_p}{h_{q;i}}\right) + \sum_{i=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_p}{h_{q;p}}\right)}, \quad (4.19)$$

and

$$u_{p,j}(h_{q;p}) = \frac{K\left(\frac{\mathbf{s}_p - \mathbf{s}_j}{h_{q;p}}\right)}{\sum_{i=1}^N \sum_{j=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_j}{h_{q;i}}\right) + \sum_{i=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_p}{h_{q;i}}\right) + \sum_{i=1}^N K\left(\frac{\mathbf{s}_i - \mathbf{s}_p}{h_{q;p}}\right)}. \quad (4.20)$$

### 4.3 Case study of Mavropigi mine

In this section SLI is applied to the LEC area density drill-hole data from Mavropigi. The lignite data were evaluated using the Spatial Profitability Index (SPI) with a critical threshold of  $\delta_c = 1.1$ , as explained in chapter 3. The resulting LEC reserves and the validation measures of section 2.5.1 will be compared with the RK estimation of section 3.4. The validation measures for the two methods are obtained with LOOCV and the reserves

are estimated using a grid with  $232 \times 245$  rectangular cells of section 3.4.2. The time needed for to make the prediction map with both methods will also be compared. A linear trend model has been removed from the data used for kriging as explained in section 3.3.1. The SLI methodology does not assume stationarity hence there is no need to remove a trend model from the data set used by the SLI. As such, we use detrended data for kriging and the data without removing trend for SLI.

### 4.3.1 Optimal kernel selection

In order to select the optimal kernel for the SLI model, all kernel functions defined in Table 4.1 were tested by means of LVO cross-validation. The parameters of the resulting SLI models are shown in Table 4.2. The resulting validation measures for the SLI predictions at the 341 locations of the data obtained with each kernel are shown in Table 4.3.

To compare the computational time required (using matlab 2012b) for the prediction by each kernel and the energy content of the mine, the resulting SLI models for all kernels were tested on the prediction grid used for kriging in section 3.4. The prediction grid includes 20 497 prediction locations. The required time and the reserves estimation for each kernel is included in Table 4.3.

Kernel	$a_1$	$a_2$	$\mu$	$\lambda$	$\bar{h}$ (m)	$h_m$ (m)
Triangular	138.99	160.16	1.79	6655	315.48	303.59
Exponential	6.15	15.68	0.50	275.58	69.83	65.35
Gaussian	105.67	178.50	1.17	6423.28	161.85	151.47
Quadratic	125.73	168.65	1.56	6932.66	276.21	265.80
Tricube	43.99	61.28	2.06	2953.00	363.83	350.12

Table 4.2: Parameters of SLI for the tested kernels.  $\bar{h}$  is the mean bandwidth and  $h_m$  is the median of the bandwidth

It is evident from Table 4.3 that all kernels perform similarly well. However, the exponential and Gaussian kernels are faster and have slightly better validation measures. The data are accurately estimated by means of the Gaussian kernel using LVO cross-validation. However, when the SLI is applied to the prediction grid using the Gaussian kernel, a few exception-

Measure	Trian	Expo	Gaussian	Quad	Tric
$\rho$	0.83	0.84	0.84	0.83	0.83
$r_S$	0.83	0.84	0.84	0.83	0.83
$\tau$	0.65	0.66	0.65	0.65	0.65
ME (Gcal/m <sup>2</sup> )	-0.47	-0.66	-0.66	-0.48	-0.48
MAE (Gcal/m <sup>2</sup> )	10.70	10.59	10.55	10.68	10.72
MaxAE (Gcal/m <sup>2</sup> )	76.07	68.72	72.33	82.05	73.36
RMSE (Gcal/m <sup>2</sup> )	15.71	15.34	15.35	15.71	15.70
Time (s)	11.08	6.68	9.80	12.01	12.80
Reserves (Pcal)	349	355	405	346	349

Table 4.3: Validation measures of SLI for different kernel functions, using LVO cross-validation. Time needed and reserves estimated are included.  $\rho$  is Pearson’s correlation coefficient,  $r_S$  is Spearman’s correlation coefficient,  $\tau$  is Kendall’s correlation coefficient, ME is the mean error, MAE is the mean absolute error, MaxAE is the maximum absolute error and RMSE is the root mean square error. Trian represents the triangular kernel function, Expo the exponential kernel function, Quad the quadratic kernel function and Tric the tricubic kernel function.

ally high values occur at some of the prediction points. Out of the 20 497 prediction points, 57 have estimated LEC density values that exceed 367.73 Gcal/m<sup>2</sup> which is three times higher than the maximum LEC density value present in the data (122.57 Gcal/m<sup>2</sup>). This happens because the Gaussian kernel is more sensitive to rapid data value fluctuations introduced by the faults or the SPI. These few outliers give a higher than expected estimate of reserves. winsorizing, Robust Kriging, Cressie

Based on the fact that the exponential kernel gives better maximum absolute error and is faster than the Gaussian kernel and exhibits more stable performance at the grid locations, we chose chose to use the exponential kernel in the comparisons between the SLI and kriging.

### 4.3.2 Method comparisons

The exponential kernel (defined in Table 4.1) with the optimal parameters shown in Table 4.2 was chosen for the application of the SLI model. Using this kernel, the reserves are estimated at 355 Pcal with the SLI compared to the 341 Pcal estimate obtained by kriging as discussed in section 3.4.

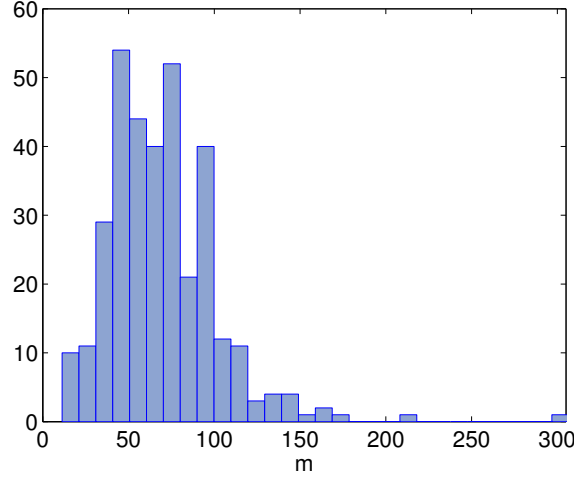


Figure 4.1: Histogram of the SLI bandwidth for the exponential kernel, obtained for the Mavropigi mine data set using  $k=1$  for the  $k$ -nearest neighbor.

The bandwidth histogram is shown in figure 4.1. The isolated high values are depended on the few isolated drill-holes of Mavropigi mine, as shown in figure 3.2. The minimum distance between nearest-neighbor drill holes varies from 22.5m to 610.1m in Mavropigi mine as explained in section 3.3. The SLI parameter  $\mu$  adjusts the kernel bandwidth as shown in equation (4.3). With the parameter  $\mu = 0.5$  and  $k = 1$  for the  $k$ -nearest neighbor, the kernel bandwidth is half the distance from the investigated drill hole to its nearest-neighbor.

The maps created by means of the two estimation methods are shown in figure 4.2. The validation measures are compared in Table 4.4. The validation measures, maps and predictions of reserves are very similar for both methods.

As shown in the Table 3.4, the LEC area density of the corrected by the SPI data of Mavropigi is between 0 and 122.6 Gcal/m<sup>2</sup> with a mean value of 37.8 Gcal/m<sup>2</sup>. The very low values of the mean error shows that neither method shows a significant bias. The mean absolute error and the root mean square error are also significantly lower than the mean LEC area density value. The value of Kendal's  $\tau$  rank correlation coefficient is satisfactory. Both Pearson's correlation coefficient  $\rho$  and Spearman's rank correlation coefficient  $r_S$  are at 84%, which is considered very strong correlation.

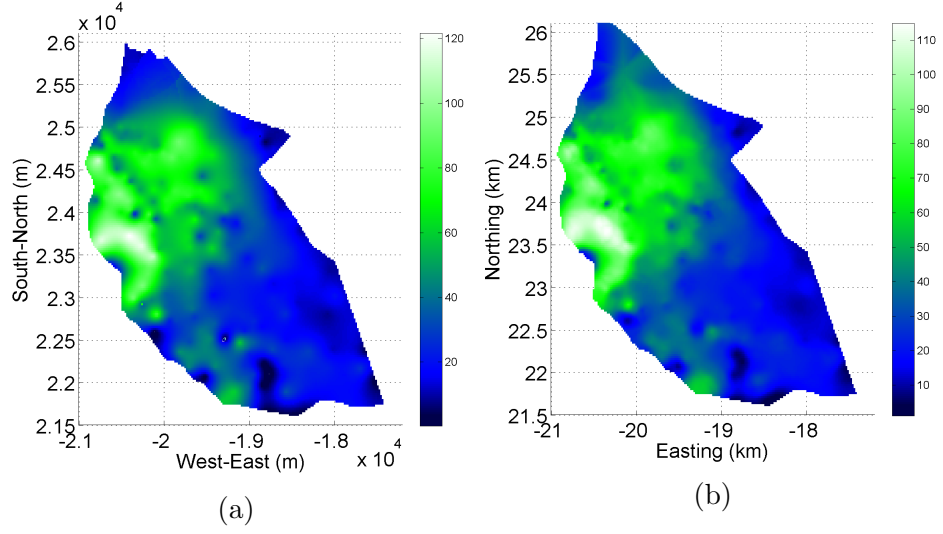


Figure 4.2: Interpolation maps of the LEC area density in Mavropigi mine using SPI with a critical threshold  $\delta_c = 1.1$ . (a) Kriging map (b) SLI map.

Furthermore, SLI requires half the computational time required by kriging for the estimation. Kriging uses a user-defined neighborhood radius of 640m for each grid location to create the required covariance matrix, while SLI uses the entire mine to make the precision matrix.

Measure	SLI	Kriging
$\varrho$	0.84	0.84
$r_S$	0.84	0.84
$\tau$	0.66	0.66
ME (Gcal/m <sup>2</sup> )	-0.66	-0.19
MAE (Gcal/m <sup>2</sup> )	10.59	10.54
MaxAE (Gcal/m <sup>2</sup> )	68.72	62.65
RMSE (Gcal/m <sup>2</sup> )	15.34	15.28
Time (s)	6.7	13.6

Table 4.4: Validation measures for the estimations with SLI compared to Regression Kriging.  $\varrho$  is Pearson's correlation coefficient,  $r_S$  is Spearman's correlation coefficient,  $\tau$  is Kendall's correlation coefficient, ME is the mean error, MAE is the mean absolute error, MaxAE is the maximum absolute error and RMSE is the root mean square error.





Figure 4.3: Power River Basin location in southeast Montana and northeast Wyoming, in the USA. The Campbell county is surrounded by the red line.

## 4.4 Coal reserves estimation in Campbell county

The second case study examined is a large dataset from Campbell county, Wyoming, in the Powder River Basin of the USA. The Powder River Basin, shown in figure 4.3, is a geologic structural basin in southeast Montana and northeast Wyoming known for its rich coal deposits. In 2007, the Powder River Basin produced 396 million tonnes of coal. Campbell county lies entirely inside the basin and covers 13,400 km<sup>2</sup>. The estimated coal deposits in the county amount to about 117 Gt. The Black Thunder Coal Mine and the North Antelope Rochelle Mine, the two largest coal mines in the world are located in Campbell county. North Antelope Rochelle Mine produced 101 Mt of coal in 2013 and Black Thunder Coal Mine produced 92Mt of coal in the same year [3].

The data comprise the total coal seam thickness at 12,951 locations as shown in fig. 4.4. The normal probability plot of the reserves is shown in fig. 4.5 and were collected by the United States geological society (USGS). The median distance from one drill hole to its nearest neighbor is 462 m. However, as evidenced in fig. 4.4, this distance is shorter in some areas,

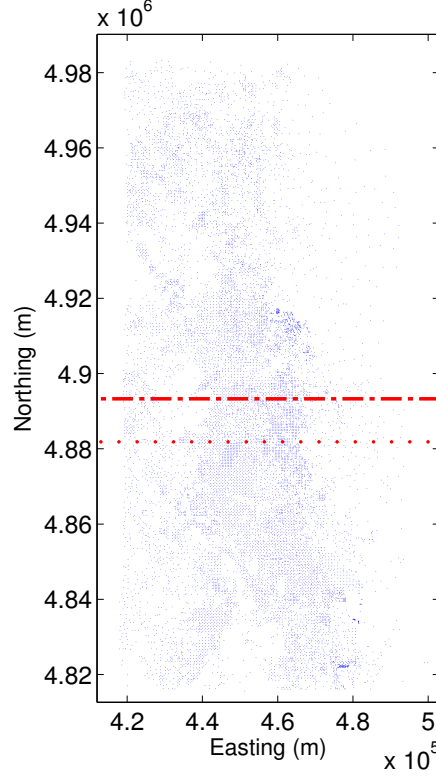


Figure 4.4: Drill hole locations in Campbell county, Wyoming. Locations above the dotted line belong to area 1 and locations below the dashed line belong to area 2. There is a 7% overlap between the two areas.

which coincide with the areas of the mining activity. The normal probability plot of the Campbell county data is presented in fig. 4.5. As evidenced, the distribution of the data (blue dots) deviate from the normal distribution (red line) below the 10% quantile and above the 90% quantile. The data from Campbell county are considered sufficiently close to the normal distribution.

#### 4.4.1 Coal reserves estimation with SLI

The SLI model was used in the dataset of the Campbell county to estimate the total coal seam thickness over a dense square grid composed of 1338 cells by 653 cells, with each cell having a side length of 125 m. Consequently, the value of the total coal seam thickness is estimated at 873 714 locations.

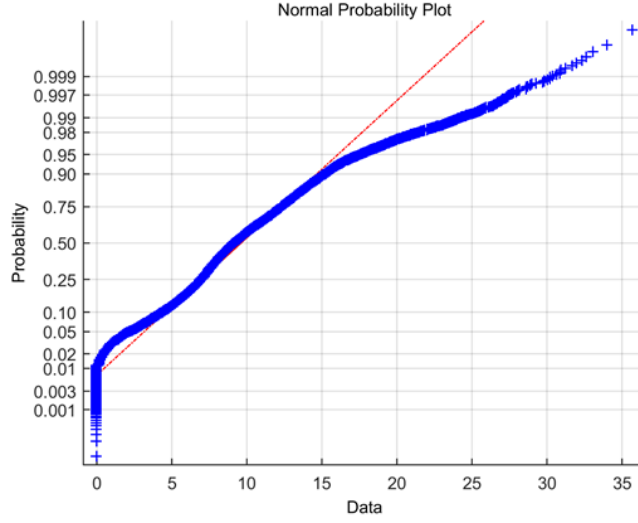


Figure 4.5: Normal probability plot of the coal data. The data (blue dots) are sorted along the X-axis according to value. The Y-axis represents the quantiles. The red line represents the theoretical normal distribution.

Due to the size of the grid and the dataset, the data for Campbell county were broken up in two neighborhoods, Northern and Southern, as shown in figure 4.4. The two neighborhoods overlap over a 7% of the entire county area. The average of the two predictions was used as the estimation for the locations that are contained in both areas.

SLI was tested using the kernel function of equation (4.21). The prediction map of coal thickness is shown in figure 4.7. The reserves are estimated at 117 Gt of coal. The SLI parameters for the two neighborhoods are shown in Table 4.5. The histograms of the kernel bandwidths for both neighborhoods are shown in figure 4.6. In both cases, the mode of the distribution of the bandwidths has value close to the median distance from one drill hole to its nearest neighbor.

$$K(u) = \exp(-u^{0.7}). \quad (4.21)$$

model parameters	Area 1	Area 2
$\lambda$	1.95	0.94
$a_1$	5.51	7.42
$a_2$	16.25	16.35
$\mu$	0.50	0.50

Table 4.5: SLI parameters of equation 4.9 for the two neighborhoods of the Campbell county

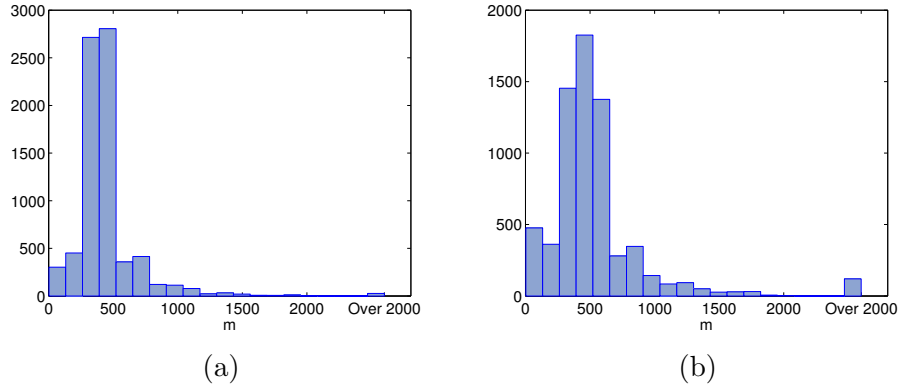


Figure 4.6: Histogram of the bandwidth for Campbell county (a) Area 1 (b) Area 2

#### 4.4.2 Coal prediction with kriging

The Campbell county data set of coal thickness at 12 951 locations was modeled with regression kriging, explained in section 2.2.4. The coordinates of the locations were normalized as in section 3.3. The Campbell county lies in a geological basin as noted in 4.4. Coal seam thickness is expected to be affected by the shape of the basin over long distances which can be modelled with a trend model. A quadratic trend model following equation (4.22) was removed from the data. The trend coefficients are shown in Table 4.6. The SSRF variogram model was fitted to the experimental variogram of the fluctuations using the WLS method [14]. The experimental variogram and the SSRF model are presented in figure 4.8. The parameters of the variogram are given in Table 4.7.

$$m_E(\mathbf{s}_{p,i}) = a_0 + a_1 \tilde{x}_i + a_2 \tilde{y}_i + a_3 \tilde{x}_i^2 + a_4 \tilde{y}_i^2 + a_5 \tilde{x}_i \tilde{y}_i, \quad i = 1, \dots, N \quad (4.22)$$

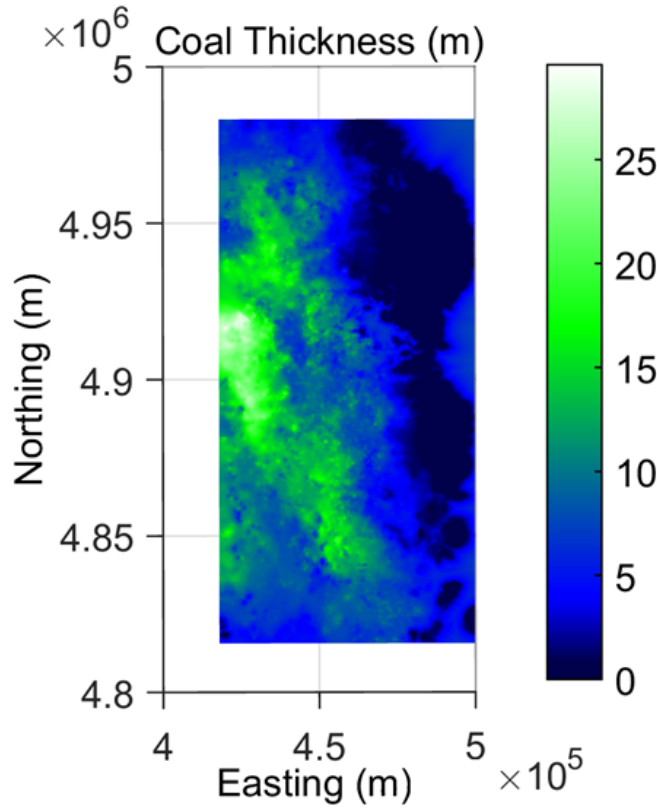


Figure 4.7: SLI map for coal seam thickness (m) in Campbell county, Wyoming in the USA.

$a_0$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$R$
11.48	-4.75	-0.41	-2.82	-0.70	-1.89	0.59

Table 4.6: Optimal parameter values  $\alpha_0, \dots, \alpha_5$  for the quadratic trend model of equation (4.22).  $R$  is the Pearson correlation coefficient between the data and the trend.

Using the variogram and trend parameters estimated and a kriging neighborhood of 7138 m, the kriging estimation map of figure 4.9 was made. The choice of the kriging neighborhood radius was motivated by the correlation length of the SSRF variogram of Table 4.7 and cross-validation performance using different ranges. The grid used is the same with the SLI interpolation grid in 4.4.1, with 1338 by 653 cells. The estimated coal reserves are 109 Gt. The reserves are estimated by multiplying the prediction of the cell total coal

thickness by the cell area ( $15\,625\text{ m}^2$ ) for all cells.

$\eta_0\text{ (m}^2\text{)}$	$\eta_1$	$\xi\text{ (m)}$	$C_0\text{ (m}^2\text{)}$
43.47	-1.84	6\,726	4.49

Table 4.7: SSRF variogram model parameters

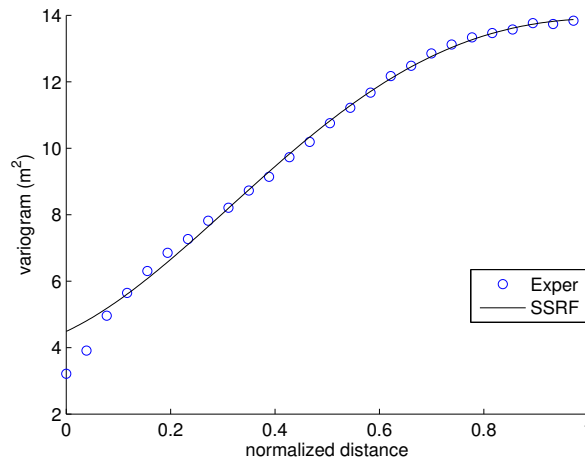


Figure 4.8: SSRF variogram model (equation (2.10)) for coal seam thickness (m) in Campbell county (line) and experimental variogram values (circles). The model parameters are in Table 4.7.

### 4.4.3 Method comparisons

Using LOOCV on the data, the validation measures presented in section 2.5.1 for SLI and kriging are calculated and the results are shown in Table 4.8. The validation measures, maps and predictions of reserves are very close for both methods as in the case of the Mavropigi mine of section 4.3.2.

The values of the thickness field are estimated at nearly  $900 \cdot 10^3$  points, which requires significant computational time. SLI requires approximately an hour and half for the prediction while kriging requires over two and a half hours. While SLI requires approximately one hour less than kriging it has similarly good validation measures as shown in Table 4.8. Regression kriging also requires detrending of the data, variogram estimation and fitting as well as investigation for the optimal kriging neighborhood. It should be

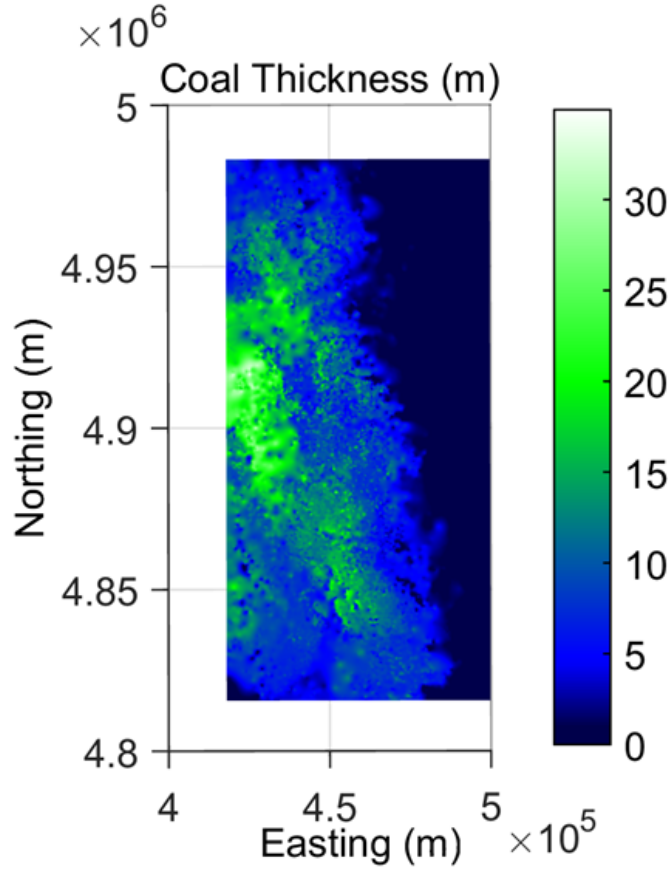


Figure 4.9: Regression kriging interpolation map for coal seam thickness (m) in Campbell county, Wyoming in the USA.

noted that the time requirements for the estimation of the optimal kriging neighborhood, trend model and variography are not included in Table 4.8.

As shown in figure 4.5, the dataset of Campbell county coal seam thickness are closer to the normal distribution than the data from the Mavropigi mine. Since kriging is the best linear unbiased estimator for Gaussian data [14, 39, 23], the close performance of the SLI in shorted computational time shows the significance of the method.

Measure	SLI	Kriging
$\varrho$	0.80	0.77
$r_S$	0.72	0.74
MAE (m)	2.3	2.0
RMSE (m)	3.1	3.0
Time (h)	1.4	2.6

Table 4.8: Leave-one-out cross-validation measures for the estimations with SLI compared to Regression Kriging for the coal reserves of Campbell county.  $\varrho$  is Pearson’s correlation coefficient,  $r_S$  is Spearman’s correlation coefficient MAE is the mean absolute error and RMSE is the root mean square error. Time is the time required for the prediction.

## 4.5 Grayscale image of Pluto

The third case investigated does not follow the normal distribution. The data are part of a grayscale digital image of the dwarf planet Pluto. The full image (fig. 4.10) is 88 by 88 pixels (7744 pixels). The data correspond to discrete values of intensity ranging from  $x_s = 0$  to  $x_s = 255$ . The histogram of these values diverges significantly from the normal distribution evidenced in the histogram of figure 4.11. This data set was investigated to compare SLI prediction with kriging on non-Gaussian datasets of discrete values.



Figure 4.10: Image of the dwarf planet Pluto, 88 by 88 pixels.

We selected 1535 points (20%) from the original data as the training set. 818 of these were randomly selected, while the other half are points that the intensity Laplacian  $L_s = \nabla^2 x_s$  has a magnitude that exceeds the threshold value of  $L_c = 30$ . Values of  $L_s < 0$  are set to 0 and values of  $L_s > 255$  are set



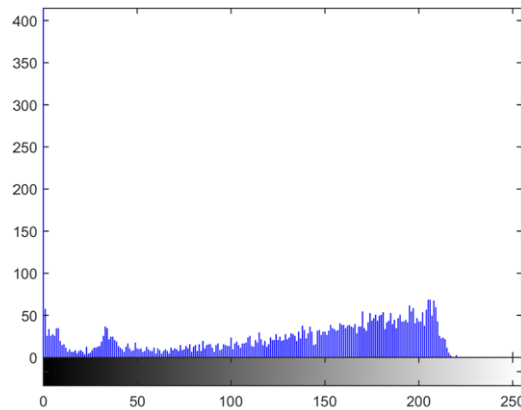


Figure 4.11: Histogram of the Pluto intensity values.

equal to 255 before the threshold is tested. These points have significantly different values from their neighbors. The threshold was chosen so that close to half the training set would be comprised by these points designated as significant with edge detection. The remaining 727 data points of the training set were randomly selected. The resulting training data set image is shown in figure 4.12 and the histogram of the Laplacian is shown in figure 4.13.



Figure 4.12: Image of the Training set, containing 20% of the pixels from the original Pluto image.

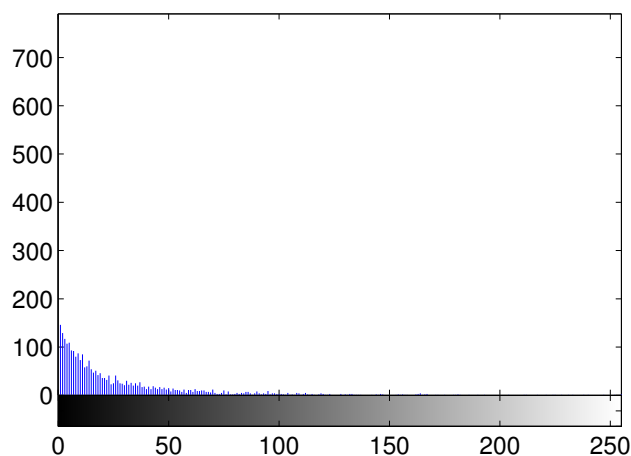


Figure 4.13: Histogram of the Laplacian  $L_s = \nabla^2 x_s$  of the intensity values  $x_s$  of the entire Pluto image of figure 4.10. Values of  $L_s < 0$  are set to 0 and values of  $L_s > 255$  are set equal to 255.

### 4.5.1 Grayscale prediction with SLI

In order to estimate the missing values of the Pluto image using the SLI model, the tricube kernel (table 4.1) was chosen. The model was applied on the missing data image of figure 4.12. The SLI model parameters are given in table 4.9 and the histogram of the bandwidths is given in figure 4.14. The reconstructed image based on the 1 560 training points is shown in figure 4.15.

$\lambda$	$a_1$	$a_2$	$\mu$
1 712.4	59.7	191.0	1.65

Table 4.9: Parameters of the SLI model used for the Pluto image reconstruction.

### 4.5.2 Grayscale prediction with kriging

Ordinary kriging was used in the training set from the digital Pluto image of 1 560 pixels of section 4.5 to estimate the 6 184 missing values. The same training set used for reconstruction with SLI was used. The SSRF variogram model was fit on the experimental variogram of the 1 560 values of the

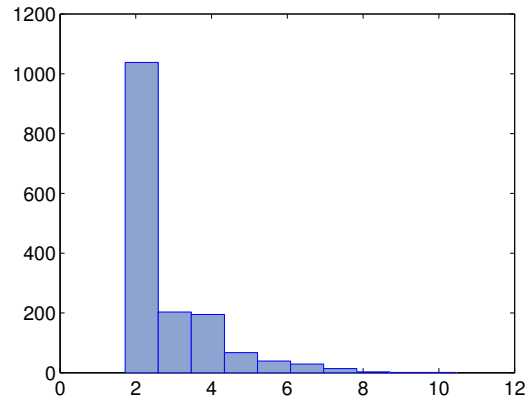


Figure 4.14: Histogram of bandwidths for the Pluto image.



Figure 4.15: Pluto image reconstructed by the SLI model with the parameters of Table 4.2 and the tricube kernel.

dataset using the WLS method. The experimental variogram and the SSRF variogram model fitted to the data are shown in figure 4.16. The parameters of the variogram are given in Table 4.10. The kriging neighborhood was chosen at a 29 pixels. The recreated image from the 1 560 data points and the OK is shown in fig 4.17.

$\eta_0$	$\eta_1$	$\xi$
69 481	-1.94	1.1332

Table 4.10: SSRF variogram model parameters fit on Pluto image training set.

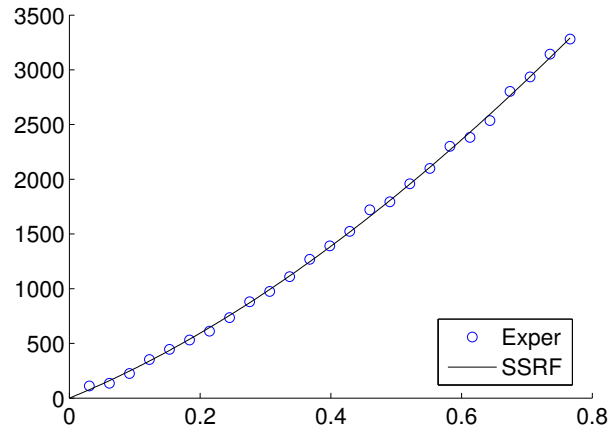


Figure 4.16: SSRF variogram model (equation (2.10), line) and experimental variogram (circles) for pixel grayscale intensity for the training set of Pluto. The model parameters are in Table 4.10.

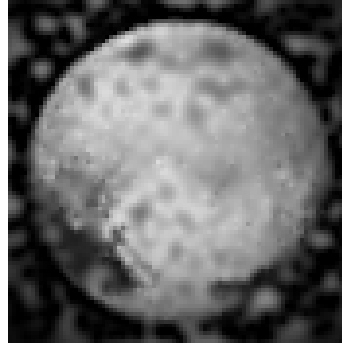


Figure 4.17: Pluto image recreated by ordinary kriging

### 4.5.3 Method comparisons

The validation measures for the predicted values of the test set (6 184 pixels) are shown in Table 4.11. SLI gives evidently better predictions for the missing data of this non-Gaussian dataset as can be seen from Table 4.11 but also from the visualization of the reconstructions in figures 4.17 and 4.15. SLI also requires less computational time, 74 seconds instead of 146 for ordinary kriging.

Measure	SLI	Kriging
$\rho$	0.99	0.85
MAE (m)	7.6	36.4
RMSE (m)	13.0	45.8
Time (s)	74	146

Table 4.11: Validation measures for the predictions of the 6 184 missing points with the SLI model compared to ordinary kriging for the reconstruction of the Pluto image based on the 1 560 data points.  $\rho$  is Pearson's correlation coefficient, MAE is the mean absolute error and RMSE is the root mean square error. Time refers to the computational time required for the reconstruction.

## 4.6 Summary

We have compared the recently proposed SLI method with kriging using diverse data including data from coal mines. Based on our studies, we conclude that the SLI method yields similar, and at times better, results to kriging. At the same time, the computational time required by SLI is considerably less than that required by kriging. This gives SLI an advantage over kriging, especially if many exploitation scenarios need to be generated and studied.

If the distribution of the data values is close to the normal distribution (like the data of Campbell county in section 4.4), SLI performs nearly as well as kriging. If the distribution deviates from the normal distribution (like the data from Mavropigi lignite mine), kriging estimations become less accurate. In the case of Mavropigi mine, SLI gives as reliable estimations as kriging. Finally, in cases where the distribution of the data is not even approximately close to the normal distribution (like the data from the Pluto image of section 4.5), SLI performs significantly better than kriging methods.

SLI models are faster than kriging methods since the precision matrix is generated by construction and in contrast with kriging the inversion of the covariance matrix is not required, unless one is interested in estimating the prediction uncertainty.



# Chapter 5

## Simulations

### 5.1 Introduction

For the Mavropigi mine in Northern Greece, we have estimated the lignite energy content (LEC) in chapter 3 using regression kriging over drill-hole data evaluated with the Spatial Profitability Index (SPI). To assess the uncertainty of the estimation for the reserves estimated in section 3.4, we used a conditional simulation method. Conditional simulation methods assume a given dataset  $D$  and a set of grid locations  $G$  where the values of the random field  $X$  need to be simulated. The method used is based on covariance decomposition combined with kriging conditioning (CDKC) [57]. CDKC is explained in section 2.4. The results of the uncertainty assessment for the reserves are discussed in 3.4.3. This chapter presents different approaches investigated than the one used in chapter 3.

The LEC area density is strictly non-negative. Nevertheless, the CDKC generates a few negative values in each realization, which have no physical meaning. Different transformations of data or the resulting simulation results were tested in an attempt to address this problem. This chapter explains the transformations tested for the original, not-SPI corrected data and includes alternative transformations for the data-set. The method that was eventually used in chapter 3 which was using the unmodified simulation results, is explained in section 3.4.3.

## 5.2 Tested transformations

The simulation method used for the conditional simulation of the Mavropigi drill-hole data was the method of covariance decomposition with kriging conditioning (CDKC). This method is presented in Section 2.4. Using the original (without SPI correction) data for conditioning, CDKC simulation generates a few negative values in each realization, which have no physical meaning. To address this problem different transformations of data or the resulting simulation results were tested.

Setting the negative LEC values of the simulation artificially to zero adds a positive bias in the estimated reserves. To address this problem, the data were transformed by applying two symmetrizing transformations; the natural logarithm and the square root of the LEC area density were tested. Both transformations overcome the problem of negative LEC area density values as the reverse transformation of the LEC area density of either gives only positive values.

After applying the transformations on the data, a linear trend is removed to obtain the fluctuations as shown in equation (5.1)

$$m_{LEC}(\mathbf{s}_{p,i}) = a_0 + a_1 \tilde{x}_i + a_2 \tilde{y}_i, \quad i = 1, \dots, N, \quad (5.1)$$

where  $m_{LEC}(\mathbf{s}_{p,i})$  is the mean LEC area density for drill-hole  $i$ ,  $N = 341$  is the total number of drill-holes in the area of the mine, and  $\tilde{x}_i, \tilde{y}_i$  are the normalized coordinates. The coefficients of the trend for all three transformations are given in Table 5.1.

Parameter	$X$	$\sqrt{X}$	$\ln(X)$
$a_0$	39.94	5.93	3.30
$a_1$	-18.57	-1.57	-0.58
$a_2$	0.73	0.08	38.60
$R$	0.70	0.73	0.73

Table 5.1: Optimal parameter values for the linear trend model (5.1) for the non-transformed and transformed data.  $X$ : Non-transformed data (lignite energy content);  $\sqrt{X}$  Square root transform of the data;  $\ln(X)$ : Logarithmic transform of the data.  $R$  is the Pearson correlation coefficient between the data and the trend.



### 5.2.1 Regression kriging on the transformed data sets

The transformations of section 5.2 produce two data sets of transformed LEC area density data. The CDKC method is explained in section 2.4. The first step of CDKC is to estimate the covariance model  $C(r)$  from the data set  $D$ . The exponential variogram model presented in section 2.2.2 was used for both transformations and for the original, non-transformed, data set for comparison. The distances are normalized as explained in section 3.3 and the variogram model is fitted on the experimental variogram using WLS. Table 5.2 lists the parameters of the exponential variogram models for the non-transformed and the transformed data.

Variogram parameters	$X$	$\sqrt{X}$	$\ln(X)$
Correlated variance $\sigma^2$	345.82	2.06	0.26
Correlation length $\xi$	0.81	0.81	0.81
Nugget variance $C_0$	53.91	0.28	0.016

Table 5.2: Exponential variogram model parameters for the non-transformed and transformed data.  $X$ : Non-transformed data (lignite energy content);  $\sqrt{X}$  Square root transform of the data;  $\ln(X)$ : Logarithmic transform of the data.

The second step is to use kriging on the grid locations (set  $G$ ) to obtain the kriging estimates vector  $\hat{\mathbf{X}}(G)$  to form the vector  $\hat{\mathbf{X}}(S) = \{\mathbf{X}(D) \cup \hat{\mathbf{X}}(G)\}$  of the simulation set  $S$ . Regression Kriging (RK) was used for all three data sets. RK was performed on a grid with cell size of 140.6 m  $\times$  140.2 m. This cell size choice is guided by the mean distance between neighboring drill holes which is 139 m as explained in section 3.4.2. The estimated LEC area density for each data set is obtained by (i) adding the trend to the predicted fluctuations and (ii) reversing the symmetrizing transformation.

The performance of RK on the transformed data sets was validated before employing them for simulations. Leave-one-out cross-validation was used for the validation measures presented in Table 5.3. The best correlation coefficients are obtained using the square-root transformation which yields an estimate of 347 Pcal. The other two data sets, however, yield similar estimates.

Statistic	$X$	$\sqrt{X}$	$\ln(X)$
Energy content (Pcal)	353	347	340
Pearson's $\rho$	0.855	0.859	0.818
Spearman's $r_S$	0.861	0.863	0.857
Kendall's $\tau$	0.682	0.682	0.673

Table 5.3: Energy content for Mavropigi mine estimated with Regression Kriging and cross-validation correlation coefficients for the three data sets of section 5.2.  $X$ : Non-transformed data (lignite energy content);  $\sqrt{X}$  Square root transform of the data;  $\ln(X)$ : Logarithmic transform of the data.

### 5.3 Conditional simulations

The transformations of section 5.2 perform similarly well when used in RK, as explained in section 5.2.1. As such,  $N_S = 5\,000$  conditional simulations of the fluctuations are generated with CDKC [57] for each transformation to assess the uncertainty of the LEC reserves prediction for each.

As explained in section 2.4, the simulation set  $S$  is  $S = D \cup G$ , with  $D$  the data set and  $G$  the set of the grid locations where the values of the LEC area density need to be simulated. The selected grid has cell size of  $140.6 \text{ m} \times 140.2 \text{ m}$ . In section 5.2.1 we have obtained the vector  $\hat{\mathbf{X}}(S)$  for all three data sets.

The next step of CDKC is to construct the covariance matrix  $\mathbf{C}_X$  for the points in  $S$ . The square root decomposition  $\mathbf{A}$  of  $\mathbf{C}_X$  was evaluated, i.e.,  $\mathbf{C}_X = \mathbf{A}^T \mathbf{A}$ .

We generated 5 000 random vectors  $\mathbf{u}_i, i = 1, \dots, 5\,000$  from the standard (zero mean, unit variance) normal distribution  $N(0, 1)$ . The length of  $\mathbf{u}_i$  was equal to the number of points in  $S$ . From these vectors  $u$ , 5 000 unconstrained simulations were generated as  $\mathbf{X}^u = \mathbf{A} \mathbf{u}$ .

We used the unconstrained simulation values  $\mathbf{X}^u(D)$  as data for a second application of kriging, which generated 5 000 interpolated vectors  $\hat{\mathbf{X}}^u(G)$ . The vectors  $\hat{\mathbf{X}}^u(S) = \{\mathbf{X}^u(D) \cup \hat{\mathbf{X}}^u(G)\}$  was subsequently constructed. Each constrained realization on  $S$  was generated by the equation

$$\mathbf{X}^c = \mathbf{X}^u + \hat{\mathbf{X}} - \hat{\mathbf{X}}^u.$$

For the two transformations, the estimated LEC area density for each simulation is obtained by (i) adding the trend to the simulated fluctuations

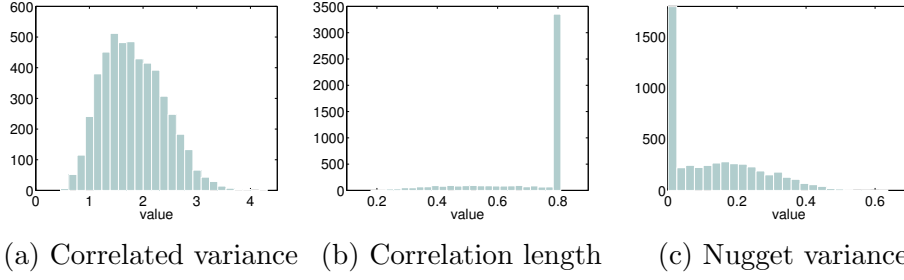


Figure 5.1: Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the square root transformation of LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS.

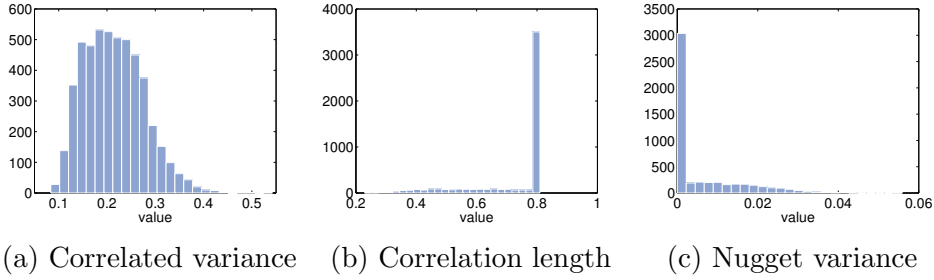


Figure 5.2: Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the logarithmic transformation of LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS.

and (ii) reversing the symmetrizing transformation. For the non-transformed data, the estimated LEC area density for each simulation is obtained by (i) adding the trend to the simulated fluctuations and (ii) changing negative values of LEC area density to zero.

The exponential variogram model was used for kriging. For each simulated realization, the parameters of the variogram model were obtained by fitting the model to the respective experimental variogram using WLS. The histograms of the exponential variogram parameters obtained from the simulated realizations are shown in Figs. 5.1–5.3.

These parameters were used in RK interpolation to estimate the energy content for each of the 5 000 simulations. For all three datasets, the correlation lengths are similar for the majority of the simulations. Zero nugget

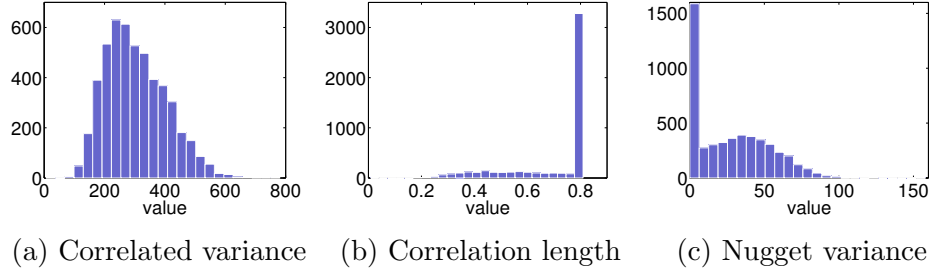


Figure 5.3: Histograms of exponential variogram parameters based on 5 000 Monte Carlo conditional simulations of the non-trasformed LEC data. For each simulation, the model parameters were obtained by fitting the model to the experimental variogram using WLS.

variance is estimated for many of the simulated states, especially for the logarithmically transformed data.

## 5.4 Simulated energy content reserves

The RK interpolations of section 5.3 were used to estimate the energy content for each of the 5 000 simulations. To obtain the reserves from the grid for each realization for the three data sets, the following procedure was used. Each grid cell has a specified area  $A$ , in the case of our grid  $A = 19712\text{m}^2$ , and a simulated value  $\hat{X}$  for the LEC area density. Multiplying the area with the simulated value of the LEC area density, the LEC for each cell is obtained. The sum of all the cells gives the LEC reserves for the mine for each realization. CDCK was also used on the non-transformed data and the negative LEC area density values were changed to zero.

The normal probability plots of the resulting LEC reserves are shown in figure 5.4. It can be seen from figure 5.4b that the reverse transformation of the logarithm can give significantly higher reserves estimates than the other two methods. Subfigure 5.4c presents the normal probability plot of the simulations before the artificial changing of negative results to zero.

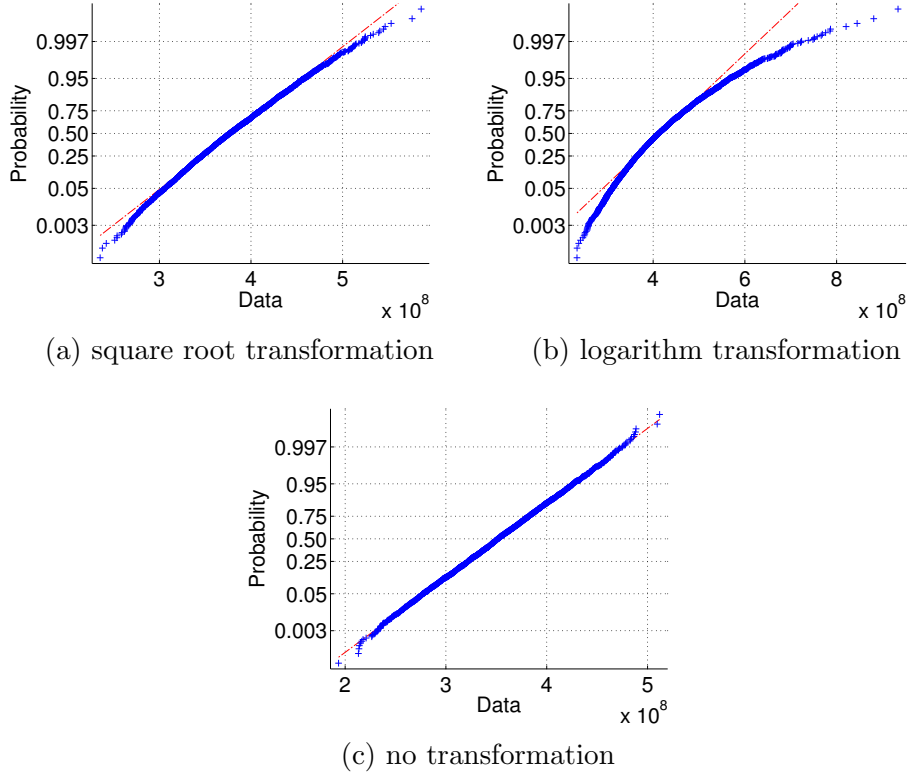


Figure 5.4: Normal probability plots for the LEC reserves estimations. 5000 simulations were conducted using CDKC simulation for the three different data sets.

## 5.5 Analysis of reserves simulation results

The statistics of the LEC energy content reserves based on 5 000 conditional simulations for each of the three data sets introduced in section 5.2 are presented in Table 5.4. The table gives the comparison of the results for the transformed data and non-transformed data after setting any simulated negative cell values to zero. The total mine energy content based on the square root and the logarithm transformations does not follow the Gaussian distribution as evidenced in the histograms of Figure 5.5 and Figure 5.4. Hence, in addition to the mean and standard deviation of the energy content per realization we also report the 2.5% percentile  $x_{2.5\%}$  and the 97.5% percentile  $x_{97.5\%}$ . The difference  $\bar{X} - \hat{X}_{OK}$  (bias) compares the simulation average of the energy content with the RK estimate (given in Table 5.3).

Moments & Quantiles	$X$	$\sqrt{X}$	$\ln(X)$
$\bar{X}$ (Pcal)	366	379	423
$\sigma_X$ (Pcal)	41.6	46.7	80.7
$x_{2.5\%}$ (Pcal)	287	293	296
$x_{97.5\%}$ (Pcal)	449	474	607
bias (Pcal)	13	32	83

Table 5.4: Simulation statistics.  $\bar{X}$ : average energy content;  $\sigma_X$ : Standard deviation of energy content;  $x_{2.5\%}$ : 2.5% percentile;  $x_{97.5\%}$ : 97.5% percentile; bias =  $\bar{X} - \hat{X}_{OK}$ .

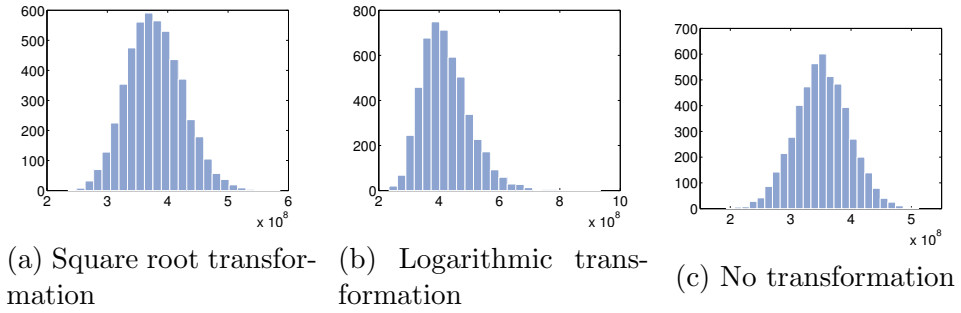


Figure 5.5: Histograms of energy content (cal) based on 5000 Monte Carlo conditional simulations.

For the square root transformation, any negative predictions would be turned to positive when the results are reversed to obtain the LEC area density estimation. As such, the more negative the value  $\hat{X}_{s_d}$  at location  $s_d$ , the higher the value of  $\hat{X}_{s_d}^2$ . This leads to positive bias and also gives false high values.

For the logarithm transformation negative LEC area density estimates would produce low estimates close to zero. However, high values of the estimated logarithm would give very high values for LEC area density after the transformation is reversed because of the exponent. This leads to very high estimates for total LEC reserves in a significant number of cases, as shown in figure 5.5b and high bias. RK also performs slightly worse with this transformation as shown in Table 5.3.

Adjusting the negative values to zero gives low bias (Table 5.3) compared to the transformations but requires artificial changing of simulation results.

## 5.6 Summary

Starting with the drill-hole LEC area density data from Mavropigi mine, we considered three different data sets for reserves estimation and uncertainty assessment: The non-transformed data, the square root of the data and the natural logarithm of the data. The validation measures of Table 5.3 show that RK performs comparatively well with all three data sets.

The CDKC simulations of the original data contain a few negative values for the LEC area density in each realization which have no physical meaning. These values can be changed to zero, but this produces a positive bias as shown in Table 5.4.

The CDKC of the two tested transformations produce only positive values for the LEC area density, overcoming the problem of the negative values without artificially changing the data. However, as shown in Table 5.4 the mine's average energy content given by the simulations of both transformations is significantly higher than the energy content given by RK of the data. Hence, none of these transformations were used to investigate reserves estimates for Mavropigi mine and assess the uncertainty.

The histograms and results of the 5 000 simulations without any transformation shown in chapter 3 have no bias nor produce any abnormally low values because of the small number of negative numbers. Hence to assess the uncertainty, it was decided to use the simulations made with the non-transformed data without changing negative values. Once this decision was made, it was followed with the SPI-corrected data.





# Chapter 6

## Conclusions

### 6.1 Conclusions

In this dissertation we introduced the spatial profitability index (SPI). SPI is a straightforward mathematical tool that locally compares costs and revenue for individual lignite seams. As shown with the drill-hole data from the Mavropigi mine, the SPI can assist in the medium term and long term planning of energy reserves exploitation. It enables engineers to investigate changes of estimated energy reserves in response to market price trends or increased costs. For example, the SPI can help to more effectively determine mine limits or total mining costs. In the medium term, SPI can assist in re-evaluating mine limits to accommodate new prices and costs. If energy demand is expected to rise or if there is a temporary need to increase revenue, the SPI can be used to determine the most profitable areas for immediate exploitation. The SPI also allows identifying non-profitable or marginally profitable sectors that can be designated for exploitation using non-continuous, lower capacity methods in the long term mine planing.

Sigmoid functions are proposed to provide quick estimates of the change of profitable reserves under different economic scenarios. These functions could be used to generate fast and accurate estimates of total mining costs for the mine or sector investigated. The advantage of this approach is that spatial analysis needs to be performed only once to determine the model parameters. Following that initial step, the sigmoid functions can be used to investigate

different scenarios without repeating the detailed spatial analysis. As such, the reserves of the mine can be given at any time depending on current prices and costs instead of relying on analysis performed several years ago with different economic situations.

The Stochastic Local Interaction (SLI) model employs a local representation to improve the computational efficiency of predictions. It is based on a joint probability density function defined by an energy functional which involves local interactions between the data points. This is achieved by means of kernel functions with locally adaptive kernel bandwidths. SLI is expressed by a respective precision matrix. This representation leads to a semi-analytical expression for prediction, which avoids the computationally costly inversion of the covariance matrix required by kriging methods.

SLI models performed reliably in all the datasets investigated in this research. Their prediction accuracy is comparable to kriging methods when used in lignite mine data from Mavropigi mine or coal thickness data from Campbell county. The SLI method calculates the prediction faster than kriging. If the distribution of the data values is close to the normal distribution, SLI performs as well as ordinary kriging, which is the best linear estimator for Gaussian data. However, SLI is faster than kriging. If the data probability distribution deviates from the normal distribution (like the data from Mavropigi lignite mine), SLI still gives very reliable estimate while kriging predictions become less accurate than SLI predictions. Finally, in cases where the distribution of the data is not even approximately close to the normal distribution (like the data from the Pluto image), SLI gives significantly better predictions than kriging with faster computation time.

## 6.2 Progress towards the objectives

In section 1.2 the objectives and goals of this that this dissertation sought to address were presented. We present here a short synopsis of how these goals and objectives were answered in this research.

The coal reserves estimates for Campbell county and Mavropigi mine are reliable according to the validation measures used. The estimation uncertainty is sufficiently assessed with conditional simulations. The SLI and

kriging maps of energy content area density and the coal seam thickness are informative, giving an accurate spatial representation of those properties. In combination with the SPI, kriging maps created for Mavropigi mine can be used reliably to assist the engineer to address economic concerns or mine exploitation issues. We believe this research fulfilled the objective of proving that geostatistical tools are sufficient to analyze the spatial variability and the estimation of uncertainties related to mine exploitation.

SLI models proved to be faster predictors than kriging methods and performed as well or even better in cases that the dataset diverges from the normal distribution. As such, this research provides evidence that SLI models are a reliable and faster alternative to kriging, especially in addressing non-Gaussian datasets.

An objective of this research was to provide simple tools that can assist in defining the pit limits and especially to adjust them under evolving economic conditions. SPI was shown to provide pit bottom elevation suggestions quickly for a range of revenue to cost ratios. Although the suggested pit bottom does not take into account technical and mechanical limitations, suggestive solutions under the current economic conditions or near-future predictions would assist in making more informed decisions regarding the mid-term mine planing.

Another goal of this research was to provide tools for the quick and efficient re-evaluation of the reserves and expected revenue of each sector or for the entire mine under different economic conditions. The ERD and the resulting graphs, based on the SPI have been shown to provide an accurate estimate of reserves changes as well as changes in total mining cost for the entire mine or the sector investigated. By virtue of being represented with a graph, ERD is very easy to use.

## 6.3 Suggestions for future studies

In future research, the SPI can be extended to single-seam deposits or different multi-seam deposits (like uranium or magnesite). With modifications suitable for the more selective exploitation methods used in underground mining, the SPI could be applied in underground mining, directing the ex-

exploitation and helping with the medium term mine planning. Using accurate 3D predictions for blocks could further motivate the use of the SPI for underground mining or to assist in defining the benches for open-pit mines.

Challenges in using the SPI in underground mining include the different paths to physically approach the block targeted for extraction. Hence, the different possible extraction costs will depend on the direction of the exploitation. However, by assessing these different costs to revenue ratios, the SPI could provide assistance in the long term underground mine planning as well as medium term mine planning.

The promising SLI models method could also be improved with further research. At the time, SLI does not provide a map of prediction variance in space similar to the kriging variance maps.

A method of conditional simulation based on the decomposition of the precision matrix created by SLI could be developed in future studies. Such a method would also require the precision matrix to be constructed in a way that ensures it is positive definite for all sampling configurations.

# Publications based on the dissertation research

- [1] A. Pavlides, M. Galetakis, D. Hristopoulos, and C. Roumpos. Evaluation of reserves using geostatistical tools to assist with long-term and mid-term mine planning. In *Mineral Resources and mine development*, pages 459–472. RWTH Aachen University, May 2015.
- [2] A. Pavlides, D. Hristopoulos, Z. Agioutantis, and C. Roumpos. Evaluation of multilayer deposit layers using a profitability index. In *Sustainable Development in the Minerals Industry - From Primary Production to Sustainable Supply Chains*, pages 317–324, June 2011.
- [3] A. Pavlides, D. Hristopoulos, M. Galetakis, and C. Roumpos. Geostatistical analysis of the calorific value and energy content of the Mavropigi multi-seam lignite deposit in northern Greece. In *Mineral Resources and mine development*, pages 199–209. RWTH Aachen University, May 2013.
- [4] A. Pavlides, D. Hristopoulos, C. Roumpos, and Z. Agioutantis. Spatial modeling of lignite energy reserves for exploitation planning and quality control. *Energy*, 93:1906–1917, dec 2015.



# Bibliography

- [1] V.D. Agou. Geostatistical analysis of precipitation on the island of crete. Master's thesis, Technical University of Crete, 2016.
- [2] N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, aug 1992.
- [3] Anonymous. EIA - U.S. Energy Information Administration. <http://www.eia.gov/coal/>, 2013. retrieved: March, 2016.
- [4] Anonymous. EURACOAL - European Association for Coal and Lignite. <http://www.euracoal.org>, 2014. retrieved: March, 2016.
- [5] Anonymous. Indexmundi. <http://www.indexmundi.com/commodities/>, May 2016. retrieved: May, 2016.
- [6] P. B. Appiah and J. R. Sturgul. Optimal mine operating strategy for a fluctuating price. *International Journal of Surface Mining, Reclamation and Environment*, 6(3):121–128, jan 1992.
- [7] M Armstrong. Common problems seen in variograms. *Mathematical Geology*, 16(3):305–313, 1984.
- [8] J. Benndorf. Application of efficient methods of conditional simulation for optimising coal blending strategies in large continuous open pit mining operations. *International Journal of Coal Geology*, 112:141–153, Jun 2013.
- [9] D. Bertermann, H. Klug, L. Morper-Busch, and C. Bialas. Modelling vS-GPs (very shallow geothermal potentials) in selected CSAs (case study areas). *Energy*, pages 226—244, July 2014.

- [10] O. Bertoli, A. Paul, Z. Casley, and D. Dunn. Geostatistical drillhole spacing analysis for coal resource classification in the Bowen basin, Queensland. *International Journal of Coal Geology*, 112:107–113, Jun 2013.
- [11] R. Bhappu and J. Guzman. Mineral investment decision making. *Engineering and Mining Journal*, 7:36–38, 1995.
- [12] Z. Bian, H. I. Inyang, J. L. Daniels, F. Otto, and S. Struthers. Environmental issues from coal mining and their solutions. *Mining Science and Technology (China)*, 20(2):215–223, mar 2010.
- [13] M. Cellura, G. Cirrincione, A. Marvuglia, and A. Miraoui. Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. *Renewable Energy*, 33(6):1251–1266, 2008.
- [14] J. P. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. Wiley series in probability and statistics. Wiley, New York, 2nd edition, 2012.
- [15] G. Christakos. *Random Field Models in Earth Sciences*. Academic Press, San Diego, 1992.
- [16] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.
- [17] N. Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990.
- [18] N. Cressie. *Spatial Statistics*. John Wiley and Sons, New York, 1993.
- [19] F. Dai, Q. Zhou, Zhiqiang Lv, X. Wang, and G. Liu. Spatial prediction of soil organic matter content integrating artificial neural network and ordinary kriging in Tibetan Plateau. *Ecological Indicators*, 45:184–194, 2014.
- [20] C. V. Deutsch. *Geostatistical Reservoir Modeling*. Oxford, New York, 2002.



- [21] S. Elogne, D. Hristopulos, and E. Varouchakis. An application of Spartan spatial random fields in environmental mapping: focus on automatic mapping capabilities. *Stochastic Environmental Research and Risk Assessment*, 22(5):633–646, 2008.
- [22] M. Galetakis. *Determination of the quality of lignite mined by continuous operation methods for multiseam deposits*. PhD thesis, School of Mineral Resources Engineering, Chania, Greece, 1996. (In Greek).
- [23] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Applied geostatistics series. Oxford University Press, New York, 1997.
- [24] H.J. Greene. *Evaluation of corrosion protection methods for Aluminum metal matrix composites*. PhD thesis, University of Southern California, Los Angeles, California, U.S.A., 1992.
- [25] A. Gut. *An Intermediate Course in Probability*. Springer New York, 2009.
- [26] J. E. Haacke and D. C. Scott. Drill hole data for coal beds in the Powder River Basin, Montana and Wyoming. Technical report, 2012.
- [27] T. Hengl, G.B.M. Heuvelink, and D.G. Rossiter. About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33(10):1301–1315, 2007.
- [28] M. Hohn and J. Britton. A geostatistical case study in West Virginia: All coals are not the same. *International Journal of Coal Geology*, 112:125–133, Jun 2013.
- [29] D. T. Hristopulos. Stochastic local interaction (SLI) model: Bridging machine learning and geostatistics. *Computers & Geosciences*, 85:26–37, dec 2015.
- [30] D. T. Hristopulos and S. Elogne. Analytic properties and covariance functions of a new class of generalized Gibbs random fields. *IEEE Transactions on Information Theory*, 53(12):4667–4679, 2007.

- [31] D.T Hristopulos. Spartan Gibbs random field models for geostatistical applications. *SIAM Journal on Scientific Computing*, 24(6):2125–2162, 2003.
- [32] D.T. Hristopulos. Applied geostatistics. Technical report, Technical University of Crete, 2012.
- [33] D.T. Hristopulos and S.N. Elogne. Computationally efficient spatial interpolators based on spartan spatial random fields. *IEEE Transactions on Signal Processing*, 57(9):3475–3487, sep 2009.
- [34] W. Hustrulid, M. Kuchta, and R. Martin. *Open Pit Mine Planning and Design*. CRC Press, 2013.
- [35] E.H. Isaaks and R.M. Srivastava. *Applied Geostatistics*. Oxford University Press, 1989.
- [36] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics*. New York : The Blackburn Press, 2003. Mining. Statistical Mathematics (BNB/PRECIS).
- [37] C. Karacan, R. Olea, and G. Goodman. Geostatistical modeling of the gas emission zone and its in-place gas content for Pittsburgh-seam mines using sequential gaussian simulation. *International Journal of Coal Geology*, 90-91:50–71, Feb 2012.
- [38] B. Z. Kebaili and A. Chebbi. Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. *Journal of Hydrology*, 365(1-2):56–73, 2009.
- [39] D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.
- [40] J. Law and J. Smullen. *profitability index*. Oxford University Press.
- [41] H. Liu, J. Shi, and E. Erdem. Prediction of wind speed time series using modified Taylor kriging method. *Energy*, 35(12):4870–4879, 2010.

- [42] J. A. Luppens, D. C. Scott, J. E. Haacke, L. M. Osmonson, and P. E. Pierce. Coal geology and assessment of coal resources and reserves in the Powder River Basin, Wyoming and Montana. Technical report, 2015.
- [43] Armstrong M. *Geostatistics*, volume 2. Springer Science, 1989.
- [44] E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, jan 1964.
- [45] R. Olea. A six-step practical approach to semivariogram modeling. *Stochastic Environmental Research and Risk Assessment*, 20(5):307–318, 2006.
- [46] R. Olea and J. Luppens. Sequential simulation approach to modeling of multi-seam coal deposits with an application to the assessment of a Louisiana lignite. *Natural Resources Research*, 21(4):443–459, Dec 2012.
- [47] R. Olea, J. Luppens, and S. Tewalt. Methodology for quantifying uncertainty in coal assessments with an application to a Texas lignite deposit. *International Journal of Coal Geology*, 85(1):78–90, Jan 2011.
- [48] R.A. Olea. *Geostatistics for Engineers and Earth Scientists*. Springer US, 1999.
- [49] A. Papoulis and S.U. Pillai. *Probability, Random Variables and Stochastic Process*. McGraw-Hill Inc., New York, fourth edition, 2002.
- [50] E. Pardo-Igúzquiza, P.A. Dowd, J.M. Baltuille, and M. Chica-Olmo. Geostatistical modelling of a coal seam for resource risk assessment. *International Journal of Coal Geology*, 112:134–140, Jun 2013.
- [51] A. Pavlides, M. Galetakis, D. Hristopoulos, and C. Roumpos. Evaluation of reserves using geostatistical tools to assist with long-term and mid-term mine planning. In *Mineral Resources and mine development*, pages 459–472. RWTH Aachen University, May 2015.
- [52] A. Pavlides, D. Hristopoulos, M. Galetakis, and C. Roumpos. Geostatistical analysis of the calorific value and energy content of the Mavropigi multi-seam lignite deposit in northern Greece. In *Mineral Resources*

- and mine development*, pages 199–209. RWTH Aachen University, May 2013.
- [53] A. Pavlides, D. Hristopoulos, C. Roumpos, and Z. Agioutantis. Spatial modeling of lignite energy reserves for exploitation planning and quality control. *Energy*, 93:1906–1917, dec 2015.
  - [54] D.R.J. Pleydell, F. Raoul, F. Tourneux, Danson F.M., A.J. Graham, P.S. Craig, and P. Giraudoux. Modelling the spatial distribution of echinococcus multilocularis infection in foxes. *Acta Tropica*, 91(3):253–265, 2004.
  - [55] J. Rivoirard. On the structural link between variables in kriging with external drift. *Mathematical Geology*, 34(7):797–808, 2002.
  - [56] K. Saikia and B.C. Sarkar. Coal exploration modelling using geostatistics in Jharia coalfield, India. *International Journal of Coal Geology*, 112:36–52, Jun 2013.
  - [57] O. Schabenberger and C.A. Gotway. *Statistical methods for spatial data analysis*. Chapman and Hall. CRC Texts in Statistical Science. CRC Press, 2004.
  - [58] T. Shoji. Statistical and geostatistical analysis of wind: a case study of direction statistics. *Computers & Geosciences*, 32(8):1025–1039, 2006.
  - [59] D. S. Simonton and S. King. Hydrogen sulfide formation and potential health consequences in coal mining regions. *Water Quality, Exposure and Health*, 5(2):85–92, 2013.
  - [60] R. M. Srivastava. Geostatistics: a toolkit for data analysis, spatial prediction and risk management in the coal industry. *International Journal of Coal Geology*, 112:2–13, 2013.
  - [61] M.L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York, 1999.
  - [62] V. Sudhir and S. Richard. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(91):1471–2105, 2006.

- [63] A. Tercan, B. Ünver, M. Hindistan, G. Ertunç, F. Atalay, S. Ünal, and Y. Killoğlu. Seam modeling and resource estimation in the coalfields of western Anatolia. *International Journal of Coal Geology*, 112:94–106, Jun 2013.
- [64] A.E. Tercan and B. Sohrabian. Multivariate geostatistical simulation of coal quality data by independent components. *International Journal of Coal Geology*, 112:53–66, Jun 2013.
- [65] E. Vanmarcke. *Random Fields: Analysis and Synthesis*. World Scientific, Princeton University, USA, 2010.
- [66] E. A. Varouchakis and D. T. Hristopulos. Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables. *Advances in Water Resources*, 52:34–49, 2013.
- [67] E.A. Varouchakis. *Geostatistical Analysis and Space-Time Models of Aquifer Levels: Application to Mires Hydrological Basin in the Prefecture of Crete*. PhD thesis, Technical University of Crete, 2012.
- [68] E.A. Varouchakis, D.T. Hristopulos, and G.P. Karatzas. Improving kriging of groundwater level data using nonlinear normalizing transformations – a field application. *Hydrological Sciences Journal*, 57(7):1–16, 2012.
- [69] D. Yang, Z. Dong, T. Reindl, P. Jirutitijaroen, and W. Walsh. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. *Solar Energy*, 103:550–562, May 2014.
- [70] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, and W. Walsh. Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging. *Renewable Energy*, 60:235–245, Dec 2013.
- [71] P. L. Younger. Environmental impacts of coal mining and associated wastes: a geochemical perspective. *Geological Society, London, Special Publications*, 236(1):169–209, 2004.

- [72] M. Žuković and D.T. Hristopulos. Spartan random processes in time series modeling. *Physica A-statistical Mechanics and Its Applications*, 387:3995–4001, 2008.

# Appendices





# Appendix A

## Appendix: SPI code

In this appendix the programs made to calculate SPI in matlab 2010b and 2012b are given.

### Seam LCV estimation

This program takes the drill-hole coal core data for each drill-hole and combines them into seams. Each seam's average LCV is calculated from the core data that correspond to it. The program Takes as input a matlab cell from the mfile "Data.mat". This matlab cell contains the drill-hole core data. Each column of the cell has a different drill-hole. As such, CGEO{1,5} corresponds to the 5th drill-hole.

The format of each drill-hole in the matlab cell of the data should follow the format given in Table A.1.

```
% Requires:
% CGEO to input cell. exei #### X Y Z 0 0 0 0 0;
% #### orofh(m) dapedo(m) 0 ygras, CO2, tefra, K8D, paxos

% GEWT format:
% X Y Z 0 0 ...
% 1.Orof 2.Dap 3.Paxos 4.Apo 8.alas 5.Ygras 6.Tefr+CO2 7.K8D ...
```

ID	X	Y	Z	0	0	0	0	0
ID-core	ceiling	floor	0	W%	CO <sub>2</sub> %	AWF%	LCV	thickness
ID-core	ceiling	floor	0	W%	CO <sub>2</sub> %	AWF%	LCV	thickness
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table A.1: drill-hole format. ID and ID-core are the ID numbers assigned to the drill-hole and the drill-hole core data. Floor and ceiling correspond to the height above sea level for the floor and ceiling of the core in question. W%, AWF% and CO<sub>2</sub>% are the water content, ash water free content and CO<sub>2</sub> content for the core. LCV is the Lower Calorific Value (in kcal/kg) of the core. X, Y, Z are the cardinal coordinates of the drill-hole head.

```

8.Pyknot
% 0... sto telos ths gewtrhshs

% GEWT to cell tw n geotrhsewn. Sthn prwth grammh exei ta ...
    cell olwn (steira + lign)
% kai sthn deuterh mono ton lignith

load Data.mat

ngeo=length(CGEO);
GEWT=cell(2,ngeo);
ADel(1,8)=0;

for Geot=1:ngeo
    %Geot h trexousa geotrhsh
    % Daped, orof, h orofh kai dapedo sta samples ths geotrhshs
    % string o ari8mos tw n lignitikwn strwmatwn
    % B to mazemen o cell ths geotrhshs
    A=CGEO{Geot};
    [n m]=size(A);

    Daped=A(2:n-1,3); Orof=A(3:n,2); Diaf=Daped-Orof; % ...
        Diaf(i) = 0 an to dapedo tou i+1 sample einai iso me ...
            thn orofh tou i+2
    d=find(Diaf~=0);
    strlig=length(d); % ta strwmata lignith pera apo to prwto
    B(strlig+3,8)=0;
    B(1,:)=A(1,2:9);
    lin=2; % 3ekinw apo thn 2rh grammh tou B, giati h prwth ...

```

```

    einai ta ths geot
B(2,1)=A(2,2); B(2,2)=A(2,3); B(2,3)=A(2,9); ...
    B(2,5)=A(2,5); B(2,6)=A(2,6)+A(2,7); B(2,7)=A(2,8); ...
    B(2:strlig+2,8)=1.20;
for i=2:n-1
    if Diaf(i-1)==0; % moorw na to balw <0.1 gia na ...
        enwnei strwmata me mono 10 pontous diafora.
        B(lin,2)=A(i+1,3);
        B(lin,3)=B(lin,3)+A(i+1,9);
        B(lin,5)=(B(lin,5)*(B(lin,3)-A(i+1,9))+A(i+1,5)*A(i+1,9))/ ...
            B(lin,3); % zygizw ka8e fora me to paxos tou ...
            strwmatos. Gia na to kanw auto
        % pollaplasiazw thn prohgoumenh ygrasia me to ...
            paxos PRIN
        % pros8esw to neo strwma (e3ou kai afairw to neo ...
            paxos) kai meta
        % pros8etw thn nea ygrasia x to paxos tou ...
            strwmatos ths
        B(lin,6)=(B(lin,6)*(B(lin,3)-A(i+1,9)) + ...
            (A(i+1,6)+A(i+1,7)) *A(i+1,9)) / B(lin,3);
        B(lin,7)=(B(lin,7)*(B(lin,3)-A(i+1,9)) + ...
            A(i+1,8)*A(i+1,9))/ B(lin,3);
    else
        lin=lin+1;
        B(lin,1)=A(i+1,2); B(lin,2)=A(i+1,3);
        B(lin,3)=A(i+1,9);
        B(lin,5)=A(i+1,5);
        B(lin,6)=A(i+1,6)+A(i+1,7);
        B(lin,7)=A(i+1,8);
    end
end

B(2:strlig+2,3)=B(2:strlig+2,1)-B(2:strlig+2,2);
% Edw teliwnei to B ths ka8e gewtrhshs

C((strlig+1)*2+2,8)=0; % Gia na balw kai ta steira, pou ...
    einai panw apo ka8e lignitiko
C(1,:)=B(1,:); % Ta ths gewtrhshs
C(2,1)=B(1,3); % To prwto steiro 3ekina sthn epifania
C(2,2)=B(2,1); % To dapedo tou prwtou steirou sthn orofh ...
    ths prwths lignitikhs
C(2,3)=C(2,1)-C(2,2); C(2,8)=1.8; C(2,7)=-40;

```

```

for i=3:2:(strlig+1)*2+1
    C(i,:)=B((i+1)/2,:);
end

for i=4:2:(strlig+1)*2+1
    C(i,1)=B(i/2,2);
    C(i,2)=B(i/2+1,1);
    C(i,3)=C(i,1)-C(i,2);
    C(i,8)=1.8;
    C(i,7)=-40;
end

ADe1=[ADe1;C];
GEWT{1,Geot}=C;
GEWT{2,Geot}=B;
clear A B C n m Daped Orof Diaf d strlig lin i

end

%% Fixing ADe1 gia to DedRep
% Exw mono ta lignitika sto ADe1 kai exw mia extra grammh me ...
% 0 sthn arxh.

ADe1(1,:)=[];

degeot=(sum(ADe1')==0)'; % ekei pou degeot=1 allazei geotrhsh

save Amyst GEWT ADe1

```

## SPI calculation

The following algorithm Calculates the SPI for drill-holes, provided with the initial pit limits and regions for sectors. The drill-hole core data needed are provided by the previous program.

```

% Deikths ekmetalleushmothtas gia esoda/e3oda
% O ari8mhths einai esoda/sq.meter kai o paranomasths ...
    e3oda/sq.meter.
%
% mat files pou xreiazontai (Apo alla programmata)
% Pairnei to texniko Peras apo arxeio Texper.mat
% Pairnei apo to arxeio Amyst.mat to Adel, Geot, GEWT
% Pairnei ta oria twn perioxwn kai tis times twn parametrwn ...
    apo to arxeio Regions.mat
%
%
% To Adel exei ola ta dedomena
% Geot einai oi 8eseis twn gewtrhsewn
% deikded einai 1 opou teliwnei gewtrhsh sto Adel
% GEWT einai to cell twn gewtrhsewn. Sthn prwth gramh exei ...
    tis gewtrhseis
% me ola ta dedomena kai sthn deuterh gramh ths gewtrhseis ...
    me mono ta
% lignitika strwmata kai sthn trith ta steira strwmata
%
% INPUT:
% To Krisimo Orio, ka8aros ari8mos
%
% OUTPUT:
% DEIK cell, opou sthn prwth grammh o deikths ...
    ekmetalleushmothtas
% sth deuterh o indicator
% dcris to crisimo orio

% GEWT CELL:
% X Y Z 0 ...
% 1.Orof 2.Dap 3.Paxos 4.therm per 5.Ygras 6.Tefr 7.K8D 8.Pykn

% Bazw ena epipleon kostos, to COSTEXT, to opoio gia thn wra ...
    einai 0. To
% kostos auto dinetai apo allou kai perilambanei "loipa" ...
    kosth sto ka8e strwma lignith.
% PX mporei na einai pio akribh metafora apo thn perioxh ktl.
% Ena kostos gia mia perioxh, mporei na spasei omoiomorfa ...
    stis gewtrhseis ths perioxhs
% kai sta strwmata ths ka8e gewtrhshs analoga me to pws to ...
    8eloume.
% PX ena kostos 150Eurw/sq.m se mia perioxh me 10 gewtrhseis ...

```

```

    mporei na spasei san
% 150 Eurw/gewtrhsh ths perioxhs. An h ka8e gewtrhsh exei 5 ...
    strwmata, 30 eurw sto strwma

function [DEIK]=Deikthsprwtostr(crisor)

% Parametroi se tonous h kwh kai eurw
% kwaste, klign h parametros/tn gia e3ory3h, metafora ...
    apo8esh/apo8hkeush
% T to poso poulietai h MWh – to poso kostizei.
% nsta8m h apodosh tou sta8mou. P kanei ta gcal/sqm se MWh/sqm

t0=tic;

load Amyst.mat
load Texper % Tex peras DEH
kwaste=2.5/1.8; klign=2.5/1.2; T=15; nRetent=1; P=1; % ...
    nRetent=0.35; P=4.184/3.6;
Cextra=4; % Kostos agoras gia prwto strwma 4 Euro/m^2
kdesulf=0/1.2; kbenefi=0/1.2; %Desulfurization, beneficiation ...
    parameters PER TON

n=length(deikded);
texper(1:n,1)= PER(:,3); %texper to texniko peras an den ...
    8elw, to bazw 0. Paei ana gewtrhsh. An 8elw epifaneia, to ...
    bazw apo ta x,y tw n gewtrhsewn
%texper(1:n,1)= 0;

%% COSTEX
% Ta kosth
% Mia gewtrhsh ana sthlh

COSTEX=GEWT(2,:);
for i=1:n
    [gr st]=size(COSTEX{i});
    COSTEX{i}(:,8)=[];
    COSTEX{i}(2:gr,4:7)=0;
    COSTEX{i}(2,4)=Cextra;
    COSTEX{i}(2,5)=kwaste;
    COSTEX{i}(2,6)=klign;
    COSTEX{i}(2:gr,7)=kdesulf; %Cost by seam
    COSTEX{i}(1,5)=0; % Region generic

```

```

    geoloc(i,:) = COSTEX{i}(1,1:2);
end

clear gr st

% ***** Regions *****
% DREG to cell pou leei poies gewtrhseis anhkoun sthn ka8e ...
    region

load Regions.mat
[gee nreg] = size(Region);
if nreg == 0; DREG = []; end
for reg = 1:nreg
    xv = [Region{1,reg}(:,1); Region{1,reg}(1,1)]; ...
    yv = [Region{1,reg}(:,2); Region{1,reg}(1,2)];
    x = geoloc(:,1); y = geoloc(:,2);
    in = inpolygon(x,y,xv,yv); DREG{1,reg} = find(in==1);
end

clear geoloc gee

for reg = 1:nreg
    for i = 1:length(DREG{reg})
        igeo = DREG{reg}(i);
        [gr st] = size(COSTEX{igeo});
        COSTEX{igeo}(2,4) = Region{2,reg}(1); %Cextra;
        COSTEX{igeo}(2,5) = Region{2,reg}(2); %kwaste;
        COSTEX{igeo}(2,6) = Region{2,reg}(3); %klign;
        COSTEX{igeo}(2:gr,7) = Region{2,reg}(4); %kdesulf;
        COSTEX{igeo}(1,5) = reg; %Se poia region
    end
end

clear gr st igeo

%% Crisimo Orio

dcris(1:n,1) = crisor;

%% Cell by cell
% E3etazw indicator kai deikth ekmetalleushmothtas ana gewtrhsh
% Genika gia ka8e strwma Kerdos/sq.m = ari*(d-1)/d

```

```

DEIK=cell(2,n);

for cel= 1:n

    [Gram Sth1]=size(GEWT{2,cel});
    LIG=GEWT{2,cel};
    STE=GEWT{3,cel};
    CostEX=COSTEX{1,cel};

% Prwth Gramh 8esh gewtrhshs, kai 8elw strwmata≥1 gia deikth

    if Gram<2
        DEIK{1,cel}=nan; DEIK{2,cel}=nan;
    end

    if Gram≥2
        kwaste=CostEX(2,5); klign=CostEX(2,6); ...
        kdesulf=CostEX(2,7); % Parameters by location. ...
        DESULF is wrong! Needs to be done by lignite seam
        DEIK{1,cel}=nan(Gram-1,1);
        DEIK{2,cel}=nan(Gram-1,1);
        indi=nan(Gram-1,1);
        Gramt=Gram;

% Bazw Gramt gia to strwma panw apo to opoio koita o ...
    deikths, epeidh den
% allazei o ari8mos twn loop pou kanei to for

        for i=Gram:-1:2
            para=kwaste*sum( ...
                STE(i:Gramt,3).*STE(i:Gramt,8)/100 ...
            )+klign*sum( ...
                LIG(i:Gramt,3).*LIG(i:Gramt,8)/100 ) + ...
            sum(CostEX(i:Gramt,4)/100); %e3oda bazw/100 ...
            giati to exw sto 8ermiko periexomeno.
            para=para +kdesulf*sum( LIG(i:Gramt,3) ) ...
            +kbenefi*sum( LIG(i:Gramt,3) ); % DESULF ...
            needs to be done by lignite seam
            ari=nRetent*T*P*sum(LIG(i:Gramt,4)); % esoda. ...
            Sto 8ermiko periexomeno exw diairesei me 100
            d=ari/para;

            if d≥dcris(cel)

```



```

        indi(i-1,1)=1;
        DEIK{1,cel}(i-1,1)=d;
    end
    if d<dcris(cel)
        indi(i-1:Gram-1,1)=0;
        Gramt=i-1;
        DEIK{1,cel}(i-1,1)=d;
    end
    if LIG(i,1)<texper(cel); % petaw strwmata katw ...
        apo texniko peras. An den 8elw, bazw texper=0;
        indi(i-1:Gram-1,1)=0;
        Gramt=i-1;
    end

    DEIK{2,cel}=indi;
    DEIK{1,cel}(indi==0,1)=0;
end
end

clear indi d Gram Gramt LIG STE

end
%% Finale

save Deikths DEIK COSTEX kwaste klign T nRetent kdesulf ...
    kbenefi Region DREG

Tol=toc(t0)

```