AES/RE/13-37  Appraisal of geostatistical methods and geostatistical prediction of predominantly marine sand inclusions in the Frimmersdorf lignite seam in the Garzweiler open cast mine, Germany

18/09/2013  F.W.G.M. van Beuningan
Title : Appraisal of geostatistical methods and geostatistical prediction of predominantly marine sand inclusions in the Frimmersdorf lignite seam in the Garzweiler open cast mine, Germany

Author : F.W.G.M. van Beuningen

Date : September 2013

Supervisors : Dr. J. Benndorf      Delft University of Technology
               Dr. K.H.A.A. Wolf      Delft University of Technology
               Dr. T. Thielemann     RWE Power AG

TA Report number : AES/RE/13-37

Postal address : Section for Resource Engineering
                 Department of Geoscience & Engineering
                 Delft University of Technology
                 P.O. Box 5028
                 The Netherlands

Telephone : (31) 15 2781328 (secretary)

Telefax : (31) 15 2781189

Copyright ©2013 Section for Resource Engineering

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Section for Resource Engineering.
Abstract

Mining is different from most businesses because knowledge of the product is essentially based on estimates, which by their very nature include a degree of uncertainty (Dominy, Noppé, & Annels, 2004). The risks associated with mining are varied and complex, where the dominant source of risk is the orebody itself (Snowden, Glacken, & Noppé, 2002). The modeling of orebody geology is the basis for all prediction of in-situ grades, mineral resources and recoverable ore reserves, as well as mine design and long term production forecasting. The Lower Rhine Basin is a rift basin located in northwest Germany which hosts several large lignite seams of considerable economical importance. One of these massive lignite deposits - the Garzweiler open cast mine - is containing marine and fluvial sand intrusions which are significantly affecting the reserve estimation and the operational processes. The current model based on a deterministic "best guess" approach fails to properly predict these sand partings within the lignite seams. Probabilistic or stochastic models can provide an alternative.

The aim of this thesis is to elaborate a first appraisal of the possibilities for RWE Power AG to apply geostatistical modeling at its Garzweiler lignite operation. An analysis is made of the available datasets and of the applicable geostatistical methods, including their strengths and shortcomings. This exercise of applying the relevant geostatistical methods to the different datasets provides new insights for predictive modeling. On this basis, recommendations are formulated for geostatistical modeling methods applicable to the unmined area of the Garzweiler mine.

Data collection is not an aim in itself. Purposeful data collection presupposes a theory or a model which gives meaning and significance to raw data by processing them as information that is useful and needed for a specified goal. In the context of this project, this goal is the accurate prediction of the composition of a lignite seam. The different data types are thoroughly described and their strengths and limitations are clearly indicated. From the point of view of geostatistical modeling, the current practice of collecting and processing surveying data has potential for improvement, because of the limited applicability of the existing data. While collecting surveying data demands great effort, production data - digitally stored and therefore easily automated - are readily available at no cost. A continuous, automated input of available production data creates a model with permanently updated geostatistical predictions and therefore a reduction of the empirical error. In addition, drill data provide information for a wider area about not only the coal / sand ratio, but also on the ash content of the lignite. Ideally, this geostatistical model should be complemented with the KOLA data - online analytics of coal composition. The KOLA data themselves can be used for predictive purposes, but above all they possess the unique advantage of being able to validate the predicted ash content. However, practical obstacles prevent an effective use of these data.
The ordinary kriging method is widely in use and in general gives good predictions. These predictions are accompanied by an estimated error. In the context of the stationarity assumption - one of the fundaments of kriging - a significant limitation arises: ordinary kriging is poorly able to cope with sudden “structural” changes. Introducing an automation step will generate a model where continuous input and usage of the available data will result in permanently updated geostatistical predictions and a reduction of the empirical error, particularly in changing environments. The indicator approach, though not providing accurate predictions, does have the virtue of signaling areas with specific features. As opposed to kriging, simulation does not require a normally distributed dataset. The creation of multiple realizations gives a better feeling for possible scenarios of future mining. Simulation applied to ash content data presents promising results.
When I finished high school, with all the options open, I was faced with this question – what next? As some of my friends did, I decided to become a medical doctor. Actually, I was admitted to the University of Amsterdam, so that summer everything seemed clear and well. Deep inside, however, doubts developed and in the end my interest for geology, the adventure associated with working abroad and the challenges of being part of a large project prevailed over a career in a hospital. That same summer, I took the decision to start studying mining- and petroleum engineering at the Delft University of Technology. This turned out to be my best decision so far and eight years later I can say that I had an amazing time.

Over a year ago I was looking for a suitable graduation subject and it was during a student excursion in South-Africa when Jörg Benndorf introduced me to the world of geostatistics. As it is relevant to all kinds of mining operations and is a good mix of theory and practice, this turned out to be the perfect subject matter for my Master of Science thesis. Now I can look back on a time of hard but interesting and pleasant work, feeling some pride to have dealt with significant challenges during this research.

This thesis project has provided me with an invaluable learning experience, for which I owe gratitude to - first of all - the members of my exam committee. Special thanks go to Jörg Benndorf for his valuable advice on every aspect of the challenges I have been facing the past twelve months, for his patience and for all the time he spent in endless calls about problems and progress of my thesis. In fact, Jörg is one of the few speed dials in my mobile phone! One of the main lessons learned during this project is that the social context is of decisive importance for the results of one's efforts. When I first met Thomas Thielemann, my supervisor at RWE Power AG in Cologne, he provided me a warm welcome at the department. I would like to thank Thomas for his lavish support during the whole project and for all the available data he supplied me with. Furthermore, I would like to thank my colleagues at the Kohlequalität and Lagerstätte department of RWE Power AG in Cologne. I appreciate their hospitality and their collaborative spirit, which were of great value both to the project and to myself. Language barriers easily disappear when discussing matters with supportive and professional colleagues like you. Working at a office with an atmosphere as in Cologne is highly motivating and the real German lunches kept me going during the cold winter days.
# Table of content

Preface ................................................................................................................... VII
Abstract .............................................................................................................. IV
List of figures ...................................................................................................... XII
Nomenclature ...................................................................................................... XVII
List of acronyms ................................................................................................ XVII
List of symbols .................................................................................................. XVIII

Chapter 1 .............................................................................................................. 1
Geological setting of the Lower Rhine Basin - with focus on the Garzweiler mine area .................................................................................... 1
  Overview of the geological setting ................................................................ 1
  Stratigraphy and sedimentological setting .................................................... 2
  Structural geology of the regional geology .................................................... 6
Geology of the Garzweiler mine area ................................................................ 8
  Introduction to the Garzweiler opencast mine ............................................ 8
  Structural geology of the Garzweiler mine area .......................................... 8
  Stratigraphy of the Garzweiler mine area ..................................................... 8
  Sand inclusions in the Frimmerdorf lignite seam ........................................ 10

Chapter 2 .............................................................................................................. 11
Lignite - mining and uses of brown coal .......................................................... 11
  Overview of lignite mining areas in Germany and the German Rhineland .......... 11
  Lignite mining - a closer look to the practical side ..................................... 13
  General mining methods for lignite production ........................................ 13
  Mining accuracy and practical conversion of coal percentage to coal recovery ... 13
  Use of the produced lignite and consequences of an increased ash content ...... 14

Master of Science Thesis
Frans van Beuningen
Chapter 3 ...........................................................................................................................................16

"Predicting production" - different models to represent the unknown reality ...........................................16
Current deterministic modeling procedures in place at RWE operations .............................................16
Deterministic vs. stochastic modeling .................................................................................................17

Chapter 4 ...............................................................................................................................................20
Problem statement and thesis description .............................................................................................20
Research background ............................................................................................................................20
Problem statement ...............................................................................................................................21
Thesis definition and objectives ...........................................................................................................21
Thesis outline .........................................................................................................................................22

Chapter 5 ...............................................................................................................................................23
Review of geostatistics and the available geostatistical methods ............................................................23
Introduction ............................................................................................................................................23
The random function model ..................................................................................................................24
Random variable ....................................................................................................................................24
Random function ...................................................................................................................................25
Dependence, linear correlation and stationarity ......................................................................................25
Inference and statistical modeling .........................................................................................................26
Variogram modeling and the theoretical error of prediction ................................................................27
Interpolation methods - Kriging ..............................................................................................................30
Kriging paradigm ....................................................................................................................................30
Other kriging methods ..........................................................................................................................30
Indicator kriging ....................................................................................................................................31
Simulation methods ..............................................................................................................................31
Introduction - estimation vs simulation ...............................................................................................31
Stochastic simulation ............................................................................................................................31
Sequential Gaussian simulation ............................................................................................................32
Outlook to more sophisticated geostatistical methods; Multiple-point statistics .................................32

Master of Science Thesis
Frans van Beuningen
Master of Science Thesis

Frans van Beuningen
Drill data - Ordinary kriging .................................................................87
Drill data - Indicator kriging .................................................................93
Distribution of the empirical error and the relevance of simulation ..........94
Drill data - Sequential Gaussian simulation ..........................................95

Chapter 8 .............................................................................................103
Conclusions, discussion and recommendations ......................................103
Introduction .........................................................................................103
Thesis objectives and the corresponding accomplishments ..................103
Conclusions .........................................................................................104
Discussion ...........................................................................................106
Recommendations and future outlook ....................................................107

References ..........................................................................................109
Appendices .........................................................................................112
Appendix A: The "SNESIM" algorithm ..................................................112
Appendix B: Export predictions and corresponding variances based on the production data .........................................................112
Appendix C: Validation results of the surveying data ...............................112
Appendix A: The "SNESIM" algorithm ..................................................113
List of figures

Figure 1: The Lower Rhine Basin and the Roer Valley Graben (RVG) both forming the Dutch-German Central Graben, a rift that deeply cuts into the Rhenish Massif (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005) ........................................... 2
Figure 2: NNW-SSE-trending facies section of the Cenozoic fill of the LRB (Utescher, et al., 2010) ........................................... 4
Figure 3: Stratigraphic standard section - including grain size profiles, records of mean annual temperature (°C) and mean annual precipitation (mm). The numbering of the horizons refers to the hydrological stratigraphy (Schneider & Thiele, 1965) .......................................................................................................................... 5
Figure 4: The Lower Rhine Basin composed of different blocks, separated by major faults and inclined towards the NE (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005) .......................................................................................................................... 7
Figure 5: The ‘Rosettenschnitt’ cross-section A-B through the LRB. Steeply dipping normal faults separate the structural blocks (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005) .......................................................................................................................... 7
Figure 6: The focus area of this research project (pink) with four southeast-northwest running faults .......... 8
Figure 7: Schematic overview of the stratigraphy in the Rhenish lignite area ; the emphasis of this research is on the Garzweiler II operation, depicted in red at the right side of the figure (Internal RWE data) ................. 9
Figure 8: The Rheinland lignite area produces the most lignite in Germany (RWTH Aachen, 2011) ........ 12
Figure 9: The three operating lignite open cast mines in the Rhenish district (Internal RWE data). ........ 12
Figure 10: A schematic overview of the general mining method for open cast lignite mining (lecture slides RWTH Aachen department of Mining Engineering III) .......................................................................................................................... 13
Figure 11: Different scenarios for the practical conversion of coal percentage to coal recovery ............... 14
Figure 12: Surface refinement using the convergent gridding algorithm (Schlumberger) ......................... 16
Figure 13: Deterministic vs probabilistic model - The deterministic model associates to the unsampled location u a single estimated value, whereas the probabilistic approach provides a probability distribution for the possible values at that location (Goovaerts, 1997). .......................................................................................................................... 18
Figure 14: The main parameters of the variogram; sill, nugget and range .................................................. 27
Figure 15: Based on the known production recoveries - a prediction and corresponding uncertainties (standard deviations) can be given .......................................................................................................................... 28
Figure 16: For a normal distributed set of recovery figures 68% of the predictions will be in the range of the prediction + or - the standard deviation ................................................................. 29
Figure 17: Three possible candidates for a training image. (a) elliptical shapes, (b) fluvial type reservoir, (c) deltaic type reservoir. Only image two can be used as a training image since pattern are stationary over the entire image. .......................................................................................................................... 33
Figure 18: An aerial map of the Garzweiler open cast mine - the red square is indicating the 2 x 2 km focus area (RWE internal data) .......................................................................................................................... 36
Figure 19: The 2 x 2 km focus area is divided into multiple grid cells .......................................................... 37
Figure 21: Spatial distribution of the production data - the 2 x 2 km focus area indicated by the orange square, the missing 2010 data by the red area ................................................................. 39
Figure 20: GPS location of the produced material - the second database required for the production data ............................... 39
Figure 22: The spatial distribution of the material 740: sand partings in the Frimmersdorf lignite seam - the big dots indicate a production block > 7.500 t, the small dots indicate a production block < 7.500 t .................... 40
Figure 23: Every grid cell is containing coal blocks and/or sand partings which make up the production recovery .................................................................................................................. 41
Figure 24: The estimated coal recovery per grid cell based on the production data - divided into multiple regions which are to be discussed later. The graph on the right indicates the average recovery of the corresponding row of grid cells .................................................................................. 43
Figure 25: All grid cells indicate a production recovery (left) or the indicative value 0/1 (right) .................................................. 44
Figure 26: Variogram of the production data - recovery .............................................................................................................. 45
Figure 27: Variogram of the production data - indicator ............................................................................................................. 45
Figure 28: The results for ordinary kriging; left the estimated recoveries, right the corresponding variances ... 46
Figure 29: The results for indicator kriging; left the indicative results, right the corresponding variances ................................. 46
Figure 30: The spatial distribution of all surveying data available ........................................................................................................ 47
Figure 31: A schematic overview of the stratigraphic layers which can be present in the 6C Frimmersdorf lignite seam ................................................................. 48
Figure 32: A schematic view of the creation of new surveying points - on the 3D-average location of two original surveying points .............................................................................................. 49
Figure 33: Wrong input data will lead to wrong ordinary kriging results .................................................................................. 50
Figure 34: A significant shift in vertical distance for the modeled surveying points might lead to the unwanted creation of unrealistic volume ........................................................................ 51
Figure 35: Schematic overview of the modeling of the floor and roof depths of a sand lens .................................................... 51
Figure 36: Ordinary kriging results in the color coding of all the grid cells. The dots simulate the surveying points and the created dummies .................................................................................. 52
Figure 37: All grid cells indicate a coal percentage (left) or the indicative value 0/1 (right) .................................................. 54
Figure 38: The variogram of the surveying data region 2 - applied to all ordinary kriging processes ................................. 53
Figure 39: The variogram of the indicative surveying data region 3 ............................................................................................. 53
Figure 40: Ordinary kriging results (left) and indicator kriging results (right) ................................................................. 54
Figure 41: The positively skewed distribution of the ash content of the drill samples .................................................... 55
Figure 42: The weighted average ash content is currently only slightly affected by the presence of a - mostly - low sand lens ........................................................................................................ 56
Figure 43: The spatial distribution of the drillholes .................................................................................................................... 57
Figure 44: The original ash content dataset is transformed into a Gaussian distributed set .............................................. 57
Figure 45: The ordinary kriging results (left) and indicator kriging results (right) based on the drill data ................................. 58
Figure 46: The Gaussian simulations are transformed back to the original distribution of the weighted average ash content ........................................................................................................ 58
Figure 47: The probability of coal occurrence (left) or sand occurrence (right) in the middle slice ................................. 59
Figure 48: Sequential Gaussian simulated ash content of the whole drill length in the 6C lignite seam for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content .................................................................59
Figure 49: The mined part of the focus area is divided into four regions based on the mining sequence - representing different moments in time ..............................................................................................64
Figure 50: Subdivision of the predictions based on region 2 into multiple mining rows ........................................66
Figure 51: Input production data (left) for the three regions and the corresponding ordinary kriging results (right) ........................................................................................................................................67
Figure 52: The variogram results of the production data for the three regions .......................................................68
Figure 53: The validation results of the ordinary kriging of the production data - divided into the predictions based on the different regions .................................................................69
Figure 54: The validation results of the ordinary kriging of the production data - divided into predictions for multiple rows ........................................................................................................71
Figure 55: The predicted coal production vs the actual coal production - shown for all the grid cells of region 1 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit ..................73
Figure 56: The predicted coal production vs the actual coal production - shown for all the grid cells of region 2 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit ..................73
Figure 57: The predicted coal production vs the actual coal production - shown for all the grid cells of region 3 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit ..................73
Figure 58: For a normal distributed set of recovery figures 68% of the predictions will be in the range of the prediction + or - the standard deviation ..........................................................................................72
Figure 59: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 1 based on the data .................................................................74
Figure 60: The amount and percentage of grid cells within the sigma-borders for the three regions ..................75
Figure 61: The location of outliers - mainly located near the major sand partings in the south ......................75
Figure 62: The general variogram model for the indicator approach of the production data - based on the data of region 2 ........................................................................................................76
Figure 63: The input data (left) and the corresponding indicator kriging results of the three regions (right) ....77
Figure 64: The validation results of the indicator approach of the production data .............................................78
Figure 65: The validation results of the ordinary kriging of the surveying data - divided into the predictions based on the different regions ........................................................................79
Figure 66: Visual explanation of the potential causes of the underestimation of the coal volumes ..........80
Figure 67: The validation results of the ordinary kriging of the surveying data - divided into predictions for multiple rows ........................................................................................................81
Figure 68: The variogram results of the surveying data in region 2 - applied to all three regions .............82
Figure 69: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 1 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit. The dashed blue line indicates a regression other than x = y which might lower the uncertainty ........................................................83
Figure 70: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 2 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit. The dashed blue line indicates a regression other than x = y which might lower the uncertainty ..................83
Figure 71: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 3 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit.

Figure 72: The amount and percentage of grid cells within the sigma-borders for the three regions.

Figure 73: The location of outliers - mainly located near the major sand partings in the south.

Figure 74: The validation results of the indicator approach of the surveying data.

Figure 75: The indicator results based on the surveying data in region 3.

Figure 76: The validation results of the ordinary kriging of the drill data - divided into the predictions in the different regions.

Figure 77: The ordinary kriging results based on the drilling data.

Figure 78: The validation results of the ordinary kriging of the drill data - divided into predictions for multiple rows.

Figure 79: The predicted coal percentage vs the actual coal production - shown for all the grid cells of all the regions separately.

Figure 80: Schematic visualization of the two drill scenarios; 1) the drillhole hitting the sand lens, and 2) the drillhole not hitting the sand lens.

Figure 81: Map of outliers of the drill data - most outliers are located near the sand lens in the southern part of the focus area, however, compared to the other two data types the outliers are more scattered.

Figure 82: The indicator kriging results of the drill data - in the unmined part of the focus area drillholes are containing sand in north-western and south-western area.

Figure 83: The validation results of the indicator approach of the surveying data.

Figure 84: Positively skewed distribution of the practical error of the drill data. Kriging approach is questionable => simulation by-passes this problem.

Figure 85: Histogram of original dataset (top) - transformed to a Gaussian distribution (bottom).

Figure 86: The simulated coal recovery figures vs the actual coal recovery within the four regions.

Figure 87: The simulated coal recovery figures vs the actual coal recovery based on the data in region 1, 2 and 3.

Figure 88: The simulated coal volumes with the minimum recovery vs the actual coal production - shown for all the grid cells of all the regions separately.

Figure 89: The simulated coal volumes with the average recovery vs the actual coal production - shown for all the grid cells of all the regions separately.

Figure 90: The simulated coal volumes with the maximum recovery vs the actual coal production - shown for all the grid cells of all the regions separately.

Figure 91: Sequential Gaussian simulated ash content of the upper horizontal slice for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold.

Figure 92: Sequential Gaussian simulated ash content of the middle horizontal slice for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold.
Figure 93: Sequential Gaussian simulated ash content of the whole drill length in the 6C lignite seam for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold .................................. 100

Figure 94: The simulated ash content can offer an indication of the daily expectation for the risk of sand occurrence ................................................................................................................................................. 101
### Nomenclature

#### List of acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>3-dimensional</td>
</tr>
<tr>
<td>LRB</td>
<td>Lower Rhine Basin</td>
</tr>
<tr>
<td>RVG</td>
<td>Roer Valley Graben</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
</tr>
<tr>
<td>LAVA</td>
<td>Lagerstätten- Verwaltung und Auswertung</td>
</tr>
<tr>
<td>cdf</td>
<td>Cumulative distribution function</td>
</tr>
<tr>
<td>pdf</td>
<td>Probability density function</td>
</tr>
<tr>
<td>ccdf</td>
<td>Conditional cumulative distribution function</td>
</tr>
<tr>
<td>RV</td>
<td>Random variable</td>
</tr>
<tr>
<td>RF</td>
<td>Random function</td>
</tr>
<tr>
<td>OK</td>
<td>Ordinary kriging</td>
</tr>
<tr>
<td>IK</td>
<td>Indicator kriging</td>
</tr>
<tr>
<td>SGSIM</td>
<td>Sequential Gaussian simulation</td>
</tr>
<tr>
<td>SNESIM</td>
<td>Single Normal Equation Simulation</td>
</tr>
<tr>
<td>SGeMS</td>
<td>Stanford Geostatistical Modeling Software</td>
</tr>
<tr>
<td>KOLA</td>
<td>Kohle OnLine Analytics</td>
</tr>
<tr>
<td>NAP</td>
<td>Normaal Amsterdams Peil</td>
</tr>
</tbody>
</table>
# List of symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z(u)$</td>
<td>Random variable at location $u$</td>
</tr>
<tr>
<td>$z(u)$</td>
<td>Realized value of random variable $Z$ at location $u$</td>
</tr>
<tr>
<td>$z^*(u)$</td>
<td>Estimate of $z$ at location $u$</td>
</tr>
<tr>
<td>$u$</td>
<td>Location of a random variable</td>
</tr>
<tr>
<td>$\text{Prob}(Z=z)$</td>
<td>Probability of the event $Z$ being $z$</td>
</tr>
<tr>
<td>$F(z)$</td>
<td>Cumulative distribution function of $z$</td>
</tr>
<tr>
<td>$f(z)$</td>
<td>Probability density function of $z$</td>
</tr>
<tr>
<td>$F(z</td>
<td>n)$</td>
</tr>
<tr>
<td>$f(z</td>
<td>n)$</td>
</tr>
<tr>
<td>$A$</td>
<td>Study area</td>
</tr>
<tr>
<td>$\gamma(h)$</td>
<td>Sample variogram value at location $h$</td>
</tr>
<tr>
<td>$h$</td>
<td>Lag vector or lag distance</td>
</tr>
<tr>
<td>$N(h)$</td>
<td>Number of pairs of data locations a vector $h$ apart</td>
</tr>
<tr>
<td>$g(h)$</td>
<td>Variogram model</td>
</tr>
<tr>
<td>$a$</td>
<td>Variogram range</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>$\sigma_m$</td>
<td>Mean standard deviation</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>Weight assigned to datum $z(u)$ interpreted as a realization of the RV $Z(u)$</td>
</tr>
<tr>
<td>$m(u)$</td>
<td>Expected value of the RV $Z(u)$</td>
</tr>
<tr>
<td>$W(u)$</td>
<td>Window of kriging estimations centered on $u$</td>
</tr>
<tr>
<td>$I(u;z_k)$</td>
<td>Indicator RV coding state $z_k$ at location $u$</td>
</tr>
<tr>
<td>$E[A]$</td>
<td>Expected value of event $A$</td>
</tr>
</tbody>
</table>
Chapter 1

Geological setting of the Lower Rhine Basin - with focus on the Garzweiler mine area

Overview of the geological setting
The Lower Rhine Basin (LRB) is a rift basin located in northwest Germany which extends towards the north into the Cenozoic graben systems of The Netherlands and the North Sea (Utescher, et al., 2010) and is strongly correlated with the uplift of the Variscan massifs (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005). It is cut into the north-western headlands of the Rhenish Massif, morphologically forming an 'embayment' (see Figure 1) (Schäfer, 1994). Sedimentation of the coastal marine siliciclastic sediments in the basin began with the Rupelian transgression of the North Sea (Schäfer, Utescher, & Mörs, 2004) The German part of the basin contains about 1,300 m of these Oligocene to Pleistocene siliciclastic sediments with intercalated brown coal layers attaining a thickness of almost 100 m (Hager, 1986). The brown coal consists of lignite of considerable economical importance and has been exploited in open cast mines since the 19th century, at locations where the seams were easily accessible (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005). Since then, the exploitation of the coal by RWE Power AG - formerly Rheinbraun AG - has formed the basis for the regional power supply. These open cast mines - currently active operations as well as closed operations - together with an extensive database of well logs are providing the basis of many studies intensely describing the structural geology and sedimentary facies of the LRB.
Deposition in the LRB began in the Early Oligocene times. The Rupelian transgression led to maximum flooding and the deposition of rich clays in many basins of Central Europe (Vandenberghe & Hardenbol, 1998) and also in the LRB (Vandenberghe, van Laenen, Echelpoel, & Lagrou, 1997). This Rupel Clay forms the base of the stratigraphic record of the LRB, as shown in Figure 2. Subsequently, the Burdigalian to Serravallian part of the strata comprises coastal sands, namely the Morken Sand, Frimmersdorf Sand and Neurath Sand, each representing a sequence with base-level fall. Between the marine sands the Morken-, Frimmersdorf- and Garzweiler brown coal seams are intercalated, all evolved under a rising sea level (Utescher, et al., 2010). From the Lower Miocene onward, the North Sea retreated from the LRB. As the groundwater table was still high and the subsidence of the basin favorable, a thick peat layer formed from wood fragments and leaves of extensive bush forests (Hager, 1986). Rivers drained the lowlands and partially eroded the topmost horizons of the peats. Thus, erosion surfaces on the tops of the coal seams are present in many sedimentary sequences of the LRB (Abraham, 1994).

**Stratigraphy and sedimentological setting**

Figure 1: The Lower Rhine Basin and the Roer Valley Graben (RVG) both forming the Dutch-German Central Graben, a rift that deeply cuts into the Rhenish Massif (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005)
The Upper Miocene in the LRB is characterized by fluvial, lacustrine and paludal conditions. The compaction of the underlying Mid-Miocene brown coal and the corresponding subsidence of different blocks provided additional accommodation space for the stacked fluvial channels and floodplain deposits of a dominantly meandering river regime grading into lacustrine clays and peat bog facies, the so-called Upper Seam (Hager & Prüfert, 1988). This succession is overlain by coarse-grained channel sediments of a braided river system - the Hauptkies Series - marking a distinct northward progradation of the shoreline during the global sea-level stand at the beginning of the Messinian (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005).

The maximum thickness of the Rhenish Main Seam in the eastern part of the central Erft Block is about 100 m (Hager, 1986). Towards the northwest it subdivides and forms three separate smaller seams, the Morken-, Frimmersdorf- and Garzweiler seams (Minnigerode & Richter, 1986). In the northwestern part of the Dutch-German Central Graben - where the North Sea was deeper - , the Oligocene to Miocene sediment fill of the basin is marine, bearing glauconite, and no brown coal was formed (Verbeek, de Leeuw, Parker, & Wong, 2002).

The formation of all geological facies are caused by different sedimentological settings, mainly influenced by the prevailing sea level. The coastal sands of the upper Oligocene to lower Miocene Köln Formation are for example formed during a high-energy coastal marine setting. South of these sands - away from the Cenozoic shoreline - lagoonal and/or back-barrier fine-grained depositional environments formed. The sediments are rich in organic material, contain mainly wood, which today is preserved as brown coal. On the other side - towards the sea in the north - marine clays and sands deposited in environments below the wave base predominate (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005).

Later, the beds of the Miocene Ville formation were deposited. They contain the main lignite seam sequence with the seams Morken I - named "6A" after (Schneider & Thiele, 1965) - Frimmersdorf (6C) and Garzweiler (6E) as well as intercalations with marine and fluvial sediments.
Sediment thickness and stratigraphy in the Lower Rhine Basin are shown in the stratigraphic standard section (see Figure 3). This was compiled from two wells, one in the western part of the Erft Block and the other in the Köln Block, corrected with the help of industry cross-sections through the basin (Schäfer, Utescher, & Mörs, 2004).

Figure 2: NNW-SSE-trending facies section of the Cenozoic fill of the LRB (Utescher, et al., 2010).
Figure 3: Stratigraphic standard section - including grain size profiles, records of mean annual temperature (°C) and mean annual precipitation (mm). The numbering of the horizons refers to the hydrological stratigraphy (Schneider & Thiele, 1965).
Structural geology of the regional geology

The southeastern part of the Dutch-German Central Graben cuts into the north-western headlands of the Rhenish Massif (Hager & Prüfert, 1988), forming the 100 km long and 50 km wide lowland area of the Lower Rhine Embayment in Germany. The rift system consists of a mosaic of NE-ward tilted half grabens and horst blocks. NW-SE trending faults define individually subsiding structural blocks. These blocks are subsiding with different velocities (Zagwijn, Beets, van den Berg, van Montfrans, & van Rooijen, 1985). The Erft Block in the central part of the LRB is the deepest, the Rur Block in the west is relatively shallow. The Erft Fault is separating the Erft Block in the east from the Köln Block, which forms the hanging wall. These tectonic units continue northwards to the Krefeld Bock, the Venlo Block and the Roer Valley Graben, forming the deepest block of the Dutch-German Central Graben system (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005). The Erft and Rur Faults join towards the northwest, to form the Peel Fault, which can be traced towards the northwest along the eastern side of the Dutch-German Central Graben. These faults are inclined westward and have a dip of 70-80°. The LRB developed by extension forming an asymmetric half graben. Close to the surface, the faults are planar in shape. In the shallow subsurface, as far as the Cenozoic sediment fill allows, the basin was formed by SW-NE directed orthogonal extension (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005).

During the Cenozoic, the Lower Rhine Basin subsided at a rate of up to 2 mm per year, predominantly along the northeastern margins of the individual sub-grabens (Quitzow & Vahlensieck, 1955). Nowadays, the subsidence reaches different values across parts of the LRB, due to natural subsidence as well as movements due to groundwater drainage (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005). The present day horizontal extension in an E-W direction is up to 2 mm per year, inferred from GPS measurements. (Campbell, et al., 2002).
Figure 4: The Lower Rhine Basin composed of different blocks, separated by major faults and inclined towards the NE (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005).

Figure 5: The 'Rosettenschnitt' cross-section A-B through the LRB. Steeply dipping normal faults separate the structural blocks (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005).
Geology of the Garzweiler mine area

Introduction to the Garzweiler opencast mine
The Lower Rhine Basin hosts multiple lignite operations (see Figure 9) - all to be discussed in the next chapter. The emphasis here is on one of those operations; the Garzweiler opencast mine. The Garzweiler opencast mine is located to the west of Grevenbroich and is moving in the direction of Erkelenz. The mine mainly touches Rhein-Erft county, Rhein county Neuss and Heinsberg county. The lignite is deposited in three seams which, taken together, are 40 m thick on average. The coal is some 40 to max 210 m below the Earth’s surface (Internal RWE data).

Structural geology of the Garzweiler mine area
Several southeast-northwest running faults are present in the Garzweiler mine, as can be seen in Figure 6. The pink square symbolizes the focus area of this project - to be discussed later - and the yellow lines represent four faults, all with a vertical displacement of up to several meters.

Stratigraphy of the Garzweiler mine area
The stratigraphy present in the Garzweiler mine is schematically depicted in Figure 7. In the context of this research, the following stratigraphic layers are of importance. The three lignite seams present in the Garzweiler mine; 1) 6A Morken lignite seam, 2) 6C Frimmersdorf lignite seam, and 3) 6E Garzweiler lignite seam. These lignite layers are separated by several sand layers; the underlying 5 Morken sand and the intermediate sand layers named 6B Frimmersdorf sand and 6D Neurath sand (Internal RWE data).
Figure 7: Schematic overview of the stratigraphy in the Rhenish lignite area; the emphasis of this research is on the Garzweiler II operation, depicted in red at the right side of the figure (Internal RWE data)
Sand inclusions in the Frimmersdorf lignite seam

The Frimmersdorf lignite seam contains multiple sand intrusions. The shape and size of these sand partings are irregular and both characteristics are showing a large variability. The Garzweiler mine - being the most northern located operation of the Rhenish lignite area - shows a pre-dominantly marine impact. More to the south - e.g. in the Hambach operation - the fluvial impact becomes stronger. However, there is not a common idea about the origin of the sand partings. Several possible scenarios for the origin of these sand partings are shortly described below - of which the first three scenarios are describing a syn-sedimentary process, as opposed to the fourth and last scenario, which describes a post-sedimentary process;

1. **An environment of marine transgression**
   A rising sea level will lead to relatively homogeneous sand bodies in the peat. This marine environment will arise slowly, therefore the sand partings can be created over a longer period of time.

2. **Accidental injection of heterogeneous sand bodies**
   Rough weather and wild sea conditions can accidentally inject a volume of sand within the peat. These events can happen in a short period of time; a daily or even hourly event may suffice. Currently this kind of events are observable at the German coast line near Willemshaven. (Boehnert, 2011) assumes this scenario as the most likely cause of the occurrence of the sand partings, based on the thickness of the sand partings and the present components within.

3. **An environment of marine regression and increasing fluvial impact**
   A decreasing sea level can strengthen the impact of the fluvial conditions. "Crevasse splay" - a situation of a broken embankment causing flooding in the adjacent swampy area - can lead to sand partings within the lignite.

4. **"Coalification"**
   The geochemical process of "coalification" can be simplified as;

   \[ \text{peat} - \text{water} - \text{CO}_2 = \text{lignite} \]

   Here, the incidental \( \text{CO}_2 \) release - comically often described as a "fart of Mother Earth" - can be accompanied by a large volume relocation. This event can remobilize 6C sand or cause an intrusion of 6D sand into the 6C lignite seam.
Overview of lignite mining areas in Germany and the German Rhineland

Germany hosts four main lignite regions, of which the Rhenish lignite area produces the most lignite annually (see Figure 8). The Rhenish lignite mining region covers an area of around 2,500 km$^2$ to the west of Cologne. RWE Power AG, which was created during 2003 by the merger of RWE Rheinbraun with other RWE interests, operates three large open cast mines in the area:

- **Garzweiler open cast mine** - located to the west of Grevenbroich and moving in the direction of Erkelenz. It has an operating surface of ca 30 km$^2$ and an annual lignite production of 35 - 40 Mt. Mining was originally limited to the 66 km$^2$ Garzweiler I area located east of motorway A44, but in 2006 mining in the 48 km$^2$ Garzweiler II sector commenced. Deposited there are 1.3 billion tons of lignite, which are to be extracted by 2045 and account for some 40% of Rhenish lignite output. The focus of this research is on sand partings within the Frimmersdorf lignite seam in the Garzweiler mine.

- **Hambach open cast mine** - 40 km$^2$ lignite open cast mine located between Jülich and Elsdorf and producing 40 - 45 Mt of lignite annually. The remaining mining area contains 2,500 Mt of lignite at a depth up to 450 m.

- **Inden open cast mine** - located between the motorways A44 and A4 and the village Aldenhoven. Inden produces annually 20 - 25 Mt of lignite and has an operating surface of ca. 16 km$^2$. The remaining lignite deposit is estimated around 500 Mt, which with the current production rate will take until 2035 to recover (Source: Internal RWE data).
Figure 8: The Rheinland lignite area produces the most lignite in Germany (RWTH Aachen, 2011)

Figure 9: The three operating lignite open cast mines in the Rhenish district (Internal RWE data).
Lignite mining - a closer look to the practical side

General mining methods for lignite production

Lignite is mined all over the world mainly (approximately 85%) by surface mining. Since the overlying strata consist mostly of unconsolidated rock, the technical methods adopted differ fundamentally from those used in the surface mining of hard coal. The European surface mining technology, designated as such because of the European development of the machines, is characterized by continuously operating excavation, transport and dumping.

In the Rhenish lignite area the lignite occurs as multiple seams strata separated by intermediate sand layers, extended over a large area. Depending on the formation of the deposit and/or on the size and shape of the approved field, either a parallel strip operation, a pivoting operation or a combination of the two will be used as the excavation method. The pivoting operation offers an advantage, since the central mine installations (e.g. belt junction) remain stationary at the base point of output over a long period of time. All three lignite operations described in the beginning of this chapter are pivoting operations.

![Figure 10: A schematic overview of the general mining method for open cast lignite mining (lecture slides RWTH Aachen department of Mining Engineering III)](image)

Mining accuracy and practical conversion of coal percentage to coal recovery

In most European lignite mines - and in the three RWE operations in the Rhenish area as well - operating the bench conveyor system most commonly uses bucket wheel excavators as excavating machines, both in overburden and in coal. These large machines are able to move huge amounts of material and on the other hand - by reason of their structural design and mode of operation - selectively excavate layers of overburden and coal. This is achieved by fine GPS positioning using antennas attached to the excavator combined with an
accurate map of the subsurface. With the operating bucket wheel excavators in the Garzweiler mine a vertical mining accuracy of 30 cm can be obtained. This is important for coal operations in inclined bedding or during separate excavation of intercalations in alternating coal and overburden operations. Therefore, the focus of this research to accurately locate the sand lenses is not the only issue; the practical implementation of the separation of these two materials is of importance as well.

Here, a note must be added. In the context of the obtained mining accuracy the *practical conversion* of a coal percentage in e.g. a drillhole, to an actual coal recovery in that region is important. Depending on the actions of the bucket wheel operator, the volumes of material handled as coal or waste can differ. This has an impact on the coal recovery and the ash content of the produced coal; Figure 11 gives two possible scenarios.

![Figure 11: Different scenarios for the practical conversion of coal percentage to coal recovery](image)

**Figure 11: Different scenarios for the practical conversion of coal percentage to coal recovery**

**Use of the produced lignite and consequences of an increased ash content**

Germany - as one of the largest lignite producing countries in the world - is likely to remain dependent on lignite as its major contributor to the energy mix. In 2012, one-fourth of Germany's electricity production comes from lignite power plants (Bruttoproduktion in Deutschland von 1990 bis 2012 nach Energieträgern, 2012). Because of its low energy density and typically high moisture content, lignite is inefficient to transport and is not traded extensively on the world market compared to coal with higher grades. It is often burned in power stations constructed very close to the lignite operations.
Based on the drill data - introduced and extensively described in chapter 6 - the average ash content in the Garzweiler opencast mine is ca. 6%. If the ash content in the lignite increases, this can have several negative consequences. The annual lignite production will decrease due to the presence of these sand intrusions. This decrease in lignite production will have, in turn, its negative consequences for the financial figures. The increased ash content has also an impact on the operational processes. It requires the utmost attention of the operators at work, expected to switch from lignite production to waste production rapidly by changing material conditions. Furthermore, the sand will increase the abrasion of the bucket wheels cutting through this harder material. The production of material with different components also intensifies the work for the blending operations, to assure a lignite supply with proper compounds to the power plants. Finally, the power plants can experience slagging problems with the excessive presence of polluting components (e.g. ash values) caused by these sand lenses.
Chapter 3

"Predicting production" - different models to represent the unknown reality

Current deterministic modeling procedures in place at RWE operations

Currently, RWE applies two software programs to model the subsurface at their operations; 1) LAVA - "Lagerstätten-Verwaltung und -Auswertung" and 2) Geoframe software by Schlumberger. Both programs are using the process of gridding in order to create a structural model from well data, fault traces and surveying- and seismic interpretation.

Figure 12: Surface refinement using the convergent gridding algorithm (Schlumberger).

The most common gridding algorithm is the convergent gridding algorithm, which is also applied for the operations of RWE. Several other gridding algorithms exist as well, mostly for special (geological) cases. On the basis of Figure 12 a short description of the convergent gridding algorithm is given.

To calculate the horizon of one geological stratum, the measured depth points of one horizon (e.g. from well data or surveying data) are taken. Around every depth point, an evenly distributed raster with 16 points is generated, made of triangles. These 16 points and their depth is weighted reverse proportional to their distance to a measured depth point. After this, the raster density will be doubled (i.e. distances halved) and the same routine is applied. This is repeated as long as the distance between 2 points is less than 8 m. In between these points, the tectonic model is combined and the triangulated results are adopted to the tectonics.
Deterministic vs. stochastic modeling

A model is but a representation of the (unknown) reality. Although that reality is unique, it has many possible representations, depending on the information available and the goal of the study. When building a model, one should use the following guidelines;

1. The model must incorporate all the relevant information.
2. The model must be tractable - one should strike a balance between a congenial but possibly unrealistic model and a more representative model with too many parameters that are difficult to infer from sparse data.
3. The model must be tuned to the goal at hand (Goovaerts, 1997).

There are two different concepts for data-driven geological modeling of the deposit; deterministic modeling and stochastic modeling.

Deterministic modeling - one "best guess" model based on the available data

A deterministic model associates to any unsampled location u a single estimated value, say, \( z^*(u) \) for the unknown value \( z(u) \), without documenting the potential error \( z^*(u) - z(u) \). For all subsequent utilizations, that unique estimated value is taken as the true value; that is, the error is assumed to be nil or negligible. Such implicit disregard for the potential error is justifiable if the estimate \( z^*(u) \) is based on either many data or some knowledge of the physics governing the spatial distribution of the attribute \( z \). Unfortunately, in the mining business such information is limited, and the usually sparse information available does not allow one to ignore the error associated to any estimate \( z^*(u) \), no matter how the variable is estimated (Goovaerts, 1997).

Deterministic models involve data collection and information processing to infer correlations and develop understanding of stratal geometry. The deterministic model inferred fully acknowledges the data - the model contains no random components; consequently, each component and input is determined exactly. Therefore, the better the data acquisition techniques become, the more accurate the resulting model.

The modeling of orebody geology is the basis for all prediction of in-situ grades, mineral resources, and recoverable ore reserves, as well as mine design and long term production forecasting. However, most resource evaluation, mine planning, and mineral processing are based on deterministic ‘best guess’ models of the geology. Uncertainty regarding the shape and position of the geological boundaries is neglected.
Stochastic modeling - "predicting probabilities"
Instead of a single estimated value for the unknown $z(u)$, the probabilistic approach provides a set of possible values with the corresponding probabilities of occurrence. Such representation reflects the imperfect knowledge of the unsampled value $z(u)$ and, more generally, of the distribution of $z$ within the area (Goovaerts, 1997). Stochastic modeling is a mathematical method in which ranges of values for each variable are used, in the form of probability distributions.

Probably the best way to construct a single geologically realistic reservoir model is to build it manually using one’s own expert knowledge. The problem with this approach is that it will provide only a single model and that integrating other types of data, such as seismic and production data, requires other expertise. Geostatistics essentially provides a reasonably automated approach to building multiple reservoir models integrating (compromising) between multiple types of information. This automation comes at the price of model simplification. Automated or stochastic methods rely on a model or algorithm that can produce multiple models without much user intervention. This requires the design of a model or algorithm that can...
capture the complexity of all available information. Such model necessarily relies on the principle of repetitivity or stationarity.

**Remarks on deterministic vs. stochastic modeling**

Several remarks should be added to the previous discussion;

- Deterministic and stochastic models may be used together. For example, a deterministic representation of better known large-scale structures can be complemented with a probabilistic modeling of small-scale variability.

- A model is not cast forever. It should be updated whenever the goal of the study changes, additional data becomes available, or the physics of the phenomenon becomes better known.

- When drawing a conclusion from a model, one should always question how much of that conclusion actually originates from the data and how much comes from the model itself. A model based on assumptions that are not supported by the data may generate artificial features (Goovaerts, 1997).
Chapter 4

Problem statement and thesis description

Research background

Mining is different from most businesses because knowledge of the product is essentially based on estimates, which by their very nature include a degree of uncertainty (Dominy, Noppé, & Annels, 2004).

The risks associated with mining are varied and complex, where the dominant source of risk is the orebody itself (Snowden, Glacken, & Noppé, 2002).

One of the most significant contributors to the total risk in the evaluation of a coal-mining project is the uncertainty of the resource tonnage and quality characteristics, often called the resource risk (Sobczyk, 2010).

The depositional and tectonic history of basin coal deposits is a significant determinant of the spatial variability of the thickness and quality characteristics of constituent coal seams and may have a major impact on the accuracy of coal resource estimates and, ultimately, on the investment risk. The thickness of a coal seam and the possible polluting lenses contained determines the total resource tonnage whilst the quality parameters (calorific value, ash content and total sulfur content) determine the coal price (Pardo-Igúzquiña, Dowd, Baltuille, & Chica-Olmo, 2012).

Mineral Resources and their subsequent conversion to Ore Reserves are of key importance to mining companies (Dominy, Noppé, & Annels, 2004). The modeling of orebody geology is the basis for the mine planning, equipment selection, economic evaluation, reporting resources and reserves, geotechnical considerations, geo-hydrological considerations and remediation. Different information from diverse sources and varying data quality can be available about the orebody (Benndorf, 2012).

However, most resource evaluation, mine planning, and mineral processing are based on deterministic “best guess” models of the geology. Uncertainty regarding the shape and position of the geological boundaries is neglected. While grade uncertainty within such fixed geological domains is being tackled with increasing expertise, the uncertainty of the underlying geological domains has not yet been satisfactorily addressed (Osterholt, 2006).
Problem statement

The Lower Rhine Basin (LRB) is a rift basin located in northwest Germany which hosts several large lignite seams of considerable economical importance. The lignite seams have been exploited in open cast mines since the 19th century, at locations where the seams where easily accessible (Schäfer, Utescher, Klett, & Valdivia-Manchego, 2005). Since then, the exploitation of the coal by RWE Power AG - formerly Rheinbraun AG - has formed the basis for the regional power supply.

One of these massive lignite deposits - the Garzweiler open cast mine - is containing marine and fluvial sand intrusions which are significantly affecting:

1. **Reserve estimations**;

   The annual lignite production decreases due to the presence of these sand partings, affecting the financial performance as well.

2. **Mining and processing operations**;

   The possibility of the undesired presence of sand partings in the 6C Frimmersdorf seam will require the utmost attention of the operators at work, expected to switch from lignite production to waste production rapidly by changing material conditions. Furthermore, the sand will increase the abrasion of the bucket wheels cutting through this harder material. The production of material with different components also intensifies the work for the blending operations, to assure a lignite supply with proper compounds to the power plants. Finally, the power plants might experience operational problems with the presence of minor components (e.g. ash values) caused by these sand partings.

The current deterministic approach fails to properly model these sand partings within the lignite seams. The formation processes of these sand partings are very complex and an analytical description using a functional relationship of these processes would require an enormous set of parameters, which would be difficult to infer. Probabilistic or stochastic models provide an alternative. Stochastic models aim to mimic the spatial or spatio-temporal structure of the attribute under consideration without the necessity to fully understand all related geological or physical processes. The modeled randomness in models is not an expression of randomness in the process. Rather it is a measure of our ignorance regarding the complex related processes, which we only partly understand (Benndorf, 2012).

Thesis definition and objectives

This thesis is titled

"Appraisal of geostatistical methods and geostatistical prediction of predominantly marine sand inclusions in the Frimmersdorf lignite seam in the Garzweiler open cast mine, Germany"

and it aims to investigate the opportunities for RWE Power AG to apply geostatistical modeling at its Garzweiler lignite operation. An analysis of both the available data and the relevant geostatistical methods - including their strengths and shortcomings - is made. Applying the most suitable geostatistical methods to the
available data will provide new insights and finally lead to recommendations concerning the most suitable geostatistical modeling methods for the unmined area of the Garzweiler mine.

The thesis’ main objectives are:

1. Analyze the available data and current data acquisition and data processing in place.

2. Analyze and compare the possible geostatistical methods.

3. Give geostatistical predictions of these sand inclusions.

4. Give certainties for the producible coal estimations.

5. Give recommendations for deposit modeling, future exploration and data processing procedures.

**Thesis outline**

The accomplishment of these objectives is set out in this report according to the following outline. The first three chapters are introductory, describing the overall situation and the preconditions of this research. From this general overview the report funnels in the fourth chapter to the problem statement and the thesis definition.

Chapter 5 provides a general introduction to the theory of geostatistics as well as a review of the geostatistical methods relevant to this project. The kriging paradigm is explained, being the pillar of the estimation methods that are currently applied. Furthermore, simulation methods are described in this chapter as well. The final part of chapter 5 gives an outlook of the possibilities of multiple-point statistics; a geostatistical method not used during this project, but most probably of significant importance for any further research.

The available data for the 6C Frimmersdorf lignite seam are presented in chapter 6. This chapter is divided into three sections, describing the main data types and the processing of these data in order to make them available for geostatistical processing. Several preliminary geostatistical steps are undertaken, in order to derive the most suitable methods for the three data types.

Chapter 7 validates the selected geostatistical methods by predicting recoveries for the mined areas of the Garzweiler mine and comparing these predictions with the actual production figures. The possibilities and limitations of all methods and data types are described.

Finally, in chapter 8 conclusions are drawn concerning 1) the available data types and 2) the applied geostatistical methods. Subsequently, the resulting insights and the corresponding paths to achieve them are discussed. Several operational recommendations and a future outlook will conclude this report.
Chapter 5

Review of geostatistics and the available geostatistical methods

Introduction
Earth sciences data are typically distributed in space. Knowledge of an attribute, say, a porosity or - in the context of this research - an ash content within a lignite seam, is thus of little interest unless the location of the measurement is known and accounted for in the data analysis. Geostatistics aims at providing quantitative descriptions of natural variables distributed in space or in time and space (Chilès & Delfiner, 1999). It means to describe spatial patterns and interpolate the value of the attribute of interest at unsampled locations. Model uncertainty about unknown values through generation of alternative images that all honor the data and reproduce aspects of patterns of spatial dependence. Geostatistical analysis includes an introduction of the tools for description, quantitative modeling of spatial continuity and a spatial prediction and uncertainty assessment (Goovaerts, 1997).

A number of simpler interpolation methods - such as inverse distance weighting, bilinear interpolation and nearest-neighbor interpolation - were already well known before geostatistics. Geostatistics goes beyond the interpolation problem by considering the studied phenomenon at unknown locations as a set of correlated random variables, and this estimation at unknown locations is mostly just an intermediate step towards a tougher estimation problem; e.g. an recoverable ore reserves assessment.

In this chapter a review of the available geostatistical methods is given. Firstly, the fundamentals of geostatistics are described, followed by an analysis to the most usual methods and an outlook on potentially more sophisticated geostatistical methods. All other topics related to geostatistics passing by during the course of this thesis are described in this chapter as well.
The random function model

Random variable

The conceptual model at the root of geostatistics, and for that matter of all of statistics and probability theory, is that of a random variable. This is the model that allows making uncertainty assessment about an imperfectly known attribute or value (Remy, Boucher, & Wu, 2009). A random variable is a variable that can take a series of possible outcomes, each with a certain probability or frequency of occurrence. Two types of variables are usually distinguished; a discrete, or categorical, variable and a continuous variable. A random variable is traditionally denoted with a capital letter, say Z. Its possible outcomes are denoted with the corresponding small case letter, say \( z_i \), \( i = 1, \ldots, n \) for a discrete variable with \( n \) outcomes, or \( z \in [z_{\text{min}}, z_{\text{max}}] \) for a continuous variable valued in the interval bounded by a minimum and a maximum value.

In the discrete case, to each outcome \( z_i \) is attached a probability value

\[
p_i = \text{Prob}(Z = z_i) \in [0,1], \quad \text{with } \sum_{i=1}^{n} p_i = 1.
\]

In the continuous case, the distribution of probability values can take the form of

- a cumulative distribution function (cdf), pictured as a cumulative histogram, providing the probability for the RV not to exceed a given threshold value \( z \);
  \[
  F(z) = \text{Prob}(Z \leq z) \in [0,1]
  \]
- a probability density function (pdf), pictured as a histogram, defined as the derivative or the slope of the previous cdf at \( z \)-values where \( F \) is differentiable; \( f(z) = dF(z)/dz \). From such pdf or cdf, probability intervals can be derived:
  \[
  \text{Prob}(z \in [a,b]) = F(b) - F(a) = \int_{a}^{b} f(z)dz
  \]

The key to a probabilistic interpretation of a variable \( z \) is the modeling of the distribution function, cdf or pdf, of the corresponding random variable \( Z \). Note that such modeling does not mean necessarily fitting a parametric function to either the cdf or pdf; a series of classes with attached probability values is a valid model (Deutsch & Journel, 1998). That distribution function should account for all information available; it then provides all that is needed to quantify the uncertainty about the actual outcome of the variable \( z \) (Remy, Boucher, & Wu, 2009).

An indicator random variable is a special kind of random variable associated with the occurrence of an event. The indicator random variable \( I_A \) associated with event \( A \) has value 1 if event \( A \) occurs and has value 0 otherwise. In other words, \( I_A \) maps all outcomes in the set \( A \) to 1 and all outcomes outside \( A \) to 0.
Random function

When the cdf is made specific to a particular information set, e.g. (n) consisting of n neighboring data values $Z(u_\alpha) = z(u_\alpha)$, $\alpha = 1,\ldots,n$, the notation "conditional to (n)" is used, defining the conditional cumulative distribution function (ccdf):

$$F(z|(n), u) = \text{Prob} \{ z(u) \leq z|(n) \}$$

The expression for the cdf models the uncertainty about the unsampled value $z(u)$ prior to using the information set (n); the expression for the ccdf models the posterior uncertainty once the information (n) has been accounted for. The concept of a random function allows such modeling and updating of prior cdf's into posterior ccdf's (Deutsch & Journel, 1998).

A random function (RF) is defined as a set of usually dependent random variables $Z(u)$, one for each location $u$ in the study area $A$, $\{Z(u), \forall u \in A\}$. To any set of $n$ locations $u_k$, $k = 1,\ldots,n$ corresponds a vector of $n$ random variables $\{Z(u_1),\ldots,Z(u_n)\}$ that is characterized by the $n$-variate or $n$-point cdf (Goovaerts, 1997).

The random function model should not be seen as some kind of generator of mineralizations, pollution, or metal prices. A model captures those aspects of the available information deemed important and reliable and allows deductive exercises like prediction through appropriate tools described later in this chapter. A model is never unique and can be judged only on its adequacy to solve the particular problem at hand. A model is not cast forever and should be updated as more information becomes available.

Dependence, linear correlation and stationarity

Several different characteristics are describing the concept of random functions. In this part of the chapter three of these characteristics are shortly described. First, the dependence of two random variables is assessing whether the probability distribution of either one is affected by any knowledge about the other one. Linear correlation is a measure of the ability of a straight line to describe the relation between two variables. Stationarity is a property of the RF model - a property needed for inference. It is not a characteristic of the phenomenon under study. The joint probability distribution of a stationary process does not change when shifted in time or space. Consequently, parameters such as the mean and variance, also do not change over time or position (Goovaerts, 1997). This stationarity property is a constitutive property of the RF model and as such need not (nor can it) be a priori checked or refuted. Nor is the decision to average statistics over a given area $A$ irreversible if more data comes available or if the goal of the study changes (Journel, 1986). Stationarity is a decision made by the user, not a hypothesis that can be proven from data. The stationarity decision allows pooling data over areas that are deemed homogeneous. Exploratory data analysis may indicate the existence of several populations with significant different statistics. This may require the subdivision of the study area into more homogeneous subzones, each being modeled with a different random function. Such subdivision is conditioned by the availability of enough data to infer the geostatistical parameters of each random function (Goovaerts, 1997).
Inference and statistical modeling

Once a random function has been chosen, the next step consists of inferring its parameters from the available data. The focus of this part of the chapter is on the inference process; a statistical process aiming at estimating the parameters of the RF model from the sample information available over the study area. Note that the sample statistics are no parameters of the population itself, since the sample is not considered exhaustive anymore (Goovaerts, 1997).

The use of sample statistics as estimates of population parameters requires that the sample be representative of the underlying area or population. However, data locations are often neither regularly nor randomly distributed over the study area. Several factors may cause specific subareas to be preferentially sampled (Goovaerts, 1997);

- Conditions of accessibility of the sampled area - environment and terrain properties
- Expected attribute values - dense sampling at more critical locations
- Sampling strategy

The sample variogram is computed as

\[ 2\gamma(h) = \frac{1}{N(h)} \sum_{h(u)} [z(u) - z(u + h)]^2 \]

where \(N(h)\) is the number of pairs of data locations a vector \(h\) apart. Variogram inference provides a set of experimental values \(\gamma(h_k)\) for a finite number of lags, \(h_k, k = 1,...,K,\) and directions. Continuous functions must be fitted to these experimental values so as to deduce variogram values for any possible lag \(h\) required by interpolation algorithms, and also to smooth out sample fluctuations (Goovaerts, 1997). Statistical modeling describes in the form of mathematical equations how one or more random variables are related to one or more other variables. To avoid having to test a posteriori the permissibility of a variogram model, a common practice consists of using only linear combinations of basic models that are known to be permissible. The following are the five most frequently used basic models:

- Nugget effect model

\[ g(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1 & \text{otherwise} \end{cases} \]

- Spherical model with range \(a\)

\[ g(h) = \begin{cases} 1.5 \cdot \frac{h}{a} - 0.5 \cdot \left(\frac{h}{a}\right)^3 & \text{if } h \leq a \\ 1 & \text{otherwise} \end{cases} \]

- Exponential model with practical range \(a\)

\[ g(h) = 1 - \exp\left(-\frac{3h}{a}\right) \]
• Gaussian model with practical range $a$

$$g(h) = 1 - \exp\left(-\frac{3h^2}{a^2}\right)$$

• Power model

$$g(h) = h^\omega \text{ with } 0 < \omega < 2$$

The first four models are bounded, which means that a sill is actually or practically reached at a range $a$. For the nugget effect model, the sill is reached as soon as $h > 0$. The spherical model reaches its sill at the actual range $a$. The exponential and Gaussian models reach their sill asymptotically. A practical range $a$ is defined as the distance at which the model value is at 95% of the sill. The power model has no sill (Goovaerts, 1997).

Three types of behavior near the origin are distinguished;

• Parabolic behavior - e.g. Gaussian model
• Linear behavior - e.g. spherical or exponential model
• Discontinuous behavior - e.g. the nugget effect model

The behavior near the origin of the power model changes with the value of the parameter $\omega$ (Goovaerts, 1997).

**Variogram modeling and the theoretical error of prediction**

The main tool for the following traditional geostatistics, and for that matter, the progress of this research, is the variogram. The *variogram* is describing the degree of spatial dependence of the analyzed variables - in this case the ash content - and is measuring the average dissimilarity between these variables at different locations.

![Variogram diagram](image)

*Figure 14: The main parameters of the variogram; sill, nugget and range*
The main parameters of the variogram are shortly described below - see Figure 14 for a visual explanation;

- **Sill**
  - Limit of the variogram tending to infinity lag distances

- **Range**
  - The distance at which the variogram reaches the sill

- **Nugget effect**
  - The sum of geological microstructures and measurement error - any error in the measurement value or the location assigned to the measurement translates to a higher nugget effect. Sparse data may also lead to a higher than expected nugget effect.

Consider a stationary random function Z(u), and any two of its random variables Z(u) and (Z_u+ h) separated by vector h. The relation between these two random variables is characterized by the following 2-point statistics, functions of the separation vector h.

\[ 2\gamma(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [z(u) - z(u + h)]^2 \]

\[ = 2 [C(h) - C(0)], \text{ if } C(h) \text{ exists}, \]

where N(h) is the number of pairs for lag h and where C(0) = Var(Z(u)).

This variance introduced in the variogram is a measure for the uncertainty of the predictions made. In the context of uncertainty measures, the variance is often transferred to the more commonly used standard deviation;

\[ \text{Standard deviation} = \sigma = \sqrt{\text{Variance}} \]

This standard deviation gives a measure for the uncertainty of the predictions, by showing how much dispersion exists from the expected value. A useful property of the standard deviation is that, unlike the variance, it is expressed in the same units as the data. The estimated recovery in a grid cell can be transposed to the next - in this case westward - production grid cell + or - that estimated standard deviation (see Figure 16).
For a normal distributed set of recovery figures - see Figure 16 - about 68% of the predictions are within one standard deviation $\sigma$ away from the predicted recovery; about 95% of the values lie within two standard deviations; and about 99.7% are within three standard deviations. This fact is known as the 3-sigma rule.

![Figure 16: For a normal distributed set of recovery figures 68% of the predictions will be in the range of the prediction $\pm$ the standard deviation](image)

"Error in the average" and the theoretical error of prediction

The foregoing theory about the prediction of the recovery of grid cells forward and the corresponding uncertainties is correct, but it must be placed in a larger context. In the analyzed situation within the Garzweiler operation, the prediction of the investigated parameter - mostly coal recovery or ash content - into the unmined area will take place over the total length of the open cast mine. In general, these geostatistical predictions are done for a particular area or cell, resulting in multiple predictions. The uncertainty in the prediction of one cell is present in all the other cells as well and therefore the chance that these predictions will level out becomes higher. This phenomenon is known as the "error in the average".

The average of multiple predictions each with the same uncertainty ($\sigma$) is given by the following formulas:

\[
\mu = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \sigma_D^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m) \quad \sigma_m^2 = \frac{\sigma_D^2}{n} \rightarrow \sigma_m = \frac{\sigma}{\sqrt{n}}
\]

So the standard deviation of all the $n$ predictions in the Garzweiler mine is the standard deviation of one prediction divided by the square root of the amount of. So - with this information and referring back to Figure 16 - the theoretical error of prediction sets the boundaries in which a certain percentage of the predictions will fall. In chapter 7 this theoretical error of prediction will be compared with the empirical error of prediction, which represents the quality of the prediction.

Now, several geostatistical methods to predict coal recoveries or ash contents for the unmined areas will be described.
Interpolation methods - Kriging

Kriging paradigm
Kriging is a generic name adopted by geostatisticians for a family of generalized least-squares regression algorithms. All kriging estimators are but variants of the basic linear regression estimator $Z^*(u)$ defined as

$$Z^*(u) - m(u) = \sum_{a=1}^{n(u)} \lambda_a(u) [Z(u_a) - m(u_a)]$$

where $\lambda_a(u)$ is the weight assigned to datum $z(u_a)$ interpreted as a realization of the random variable $Z(u_a)$. The quantities $m(u)$ and $m(u_a)$ are the expected values of the random variables $Z(u)$ and $Z(u_a)$. The number of data involved in the estimation as well as their weights may change from one location to another. In practice, only the $n(u)$ data closest to the location $u$ being estimated are retained, i.e., the data within a given neighborhood or window $W(u)$ centered on $u$ (Goovaerts, 1997).

All the kriging methods share the same objective of minimizing the error variance under the unbiasedness of the estimator. The random function $Z(u)$ is decomposed into a residual component $R(u)$ and a trend component $m(u)$. The expected value of the random variable $Z$ at location $u$ is the value of the trend component at that location: $E[Z(u)] = m(u)$. According to the model used for the trend $m(u)$ three kriging variants can be distinguished (Goovaerts, 1997);

- **Simple kriging** - considers the mean $m(u)$ to be known and constant throughout the study area.
- **Ordinary kriging** - accounts for local fluctuations of the mean by limiting the domain of stationarity of the mean to the local neighborhood $W(u)$. The mean is unknown. Ordinary kriging is usually preferred to simple kriging because it requires neither knowledge nor stationarity of the mean over the entire study area.
- **Universal kriging / Kriging with a trend model** - considers that the unknown local mean $m(u')$ smoothly varies within each local neighborhood $W(u)$, hence over the entire study area.

Other kriging methods
Based on this main kriging algorithm several other variants of kriging exist, all with their own characteristics and applications. Below several general kriging methods are listed shortly;

- **Block kriging** - allows for estimation of attribute values linearly averaged over supports (segment/area/volume) much larger than data supports. Beware that arithmetic averages of data are not the same as physical averages if the attribute does not average linearly in space like for example ore grade. In contrary, the term *point kriging* refers to estimation on point support.
• **Factorial kriging** - is a kriging method for filtering low- or high frequency components from the spatial variation of the attribute. The objective is no longer to estimate the values of the attribute, but rather to understand the origins of that value.

• **Dual kriging** - another presentation of kriging whereby estimates are expressed as linear combinations of covariance values instead of data values. This provides insight into filtering properties of kriging and reduces computational cost of kriging.

**Indicator kriging**

The indicator approach introduced in the beginning of this chapter can also be applied in the context of local estimation. The linear estimator as described in the foregoing part of this section can be expressed in terms of indicator random variables as well, defined as

\[
[I(u; z_k)]^* - E[I(u; z_k)] = \sum_{a=1}^{n(u)} \lambda_a(u; z_k) [Z(a, z_k) - E[I(u; z_k)]]
\]

where \(\lambda_a(u; z_k)\) is the weight assigned to the indicator datum \(i(u; z_k)\) interpreted as a realization of the indicator RV \(I(u; z_k)\). As with kriging estimators of the \(z\)-attribute values, two indicator kriging (IK) variants are distinguished; simple indicator kriging - for which the indicator mean is known and constant throughout study area \(A\) - and ordinary indicator kriging, accounting for local fluctuations of the indicator mean by limiting the domain of stationarity of that mean to the local neighborhood \(W(u)\) (Goovaerts, 1997).

**Simulation methods**

**Introduction - estimation vs simulation**

Let \(\{z^*(u), u \in A\}\) be the set of kriging estimates of attribute \(z\) over the study area \(A\). Each estimate \(z^*(u)\) taken separately, i.e. independently of neighboring estimates \(z^*(u')\), is "best" in the least-squares sense because the local error variance is minimum. The map of such best estimates, however, may not be best as a whole. Interpolation methods tend to smooth out local details of the spatial variation of the attribute, and this smoothing is not uniform; smoothing is minimal close to the data locations and increases as the location being estimated gets farther away from data locations. Instead of a map of local best estimates, stochastic simulation generates a map or a realization of \(z\)-values, which reproduces statistics deemed most consequential for the problem at hand (Goovaerts, 1997).

**Stochastic simulation**

**Stochastic simulation** is the process of generation of multiple realizations of the joint distribution of attribute values in space. Traditional (2-point) simulation algorithms aim at reproducing a prior covariance \(C(h)\) model, or equivalently a variogram model, that is a statistical relation between any two values \(z(u)\) and \(z(u+h)\) in space. The missing information about what should be the relation in space of three or more values taken jointly is then necessarily provided by the simulation algorithm retained. Multiple-point structures imposed
implicitly by the algorithm are most likely of high entropy nature, i.e. minimizing organization (Journel & Zhang, 2006). The covariance-based simulation algorithms widely used in practice stem from essentially two classes; the first class is anchored on the properties of the multivariate Gaussian RF model, the second class builds on the interpretation of an indicator expected value as a conditional probability (Remy, Boucher, & Wu, 2009). However, for the remainder of this part of the chapter, the focus is only on sequential Gaussian simulation.

**Sequential Gaussian simulation**

**Sequential Gaussian simulation** is based on two important theoretical properties of the multivariate normal (or Gaussian) distribution; first, the conditional distribution of an unknown variable at a particular location, given a set of known values at nearby locations, is normal. Second, the mean and variance of this conditional distribution are given by the kriging estimate of the unknown value and its associated error variance (Desbarats, 1996).

The corresponding sequential Gaussian simulation algorithm relies on the idea of simulating each grid cell facies or petrophysical property sequentially along a random path, where the simulation of cells later in the process is constrained by cells earlier simulated, along with well and seismic data (Remy, Boucher, & Wu, 2009). The algorithm proceeds as follow;

1. Start with a set of conditioning data values at scattered locations over the field to be simulated.
2. Select at random a point on the grid discretizing the field where there is not yet any simulated or conditioning data value.
3. Using both conditioning data and values already simulated from the surrounding area, calculate the kriging estimate and corresponding error variance. These are the mean and variance of the conditional distribution of the unknown value at the point given the set of known values from the surrounding area.
4. Select at random a value from this conditional distribution.
5. Add this value to the set of already simulated values.
6. Return to step 2 and repeat these steps recursively until all points of the discretized field have been assigned simulated values.

This process will lead to fields having greater spatial variability than fields of kriged values.

**Outlook to more sophisticated geostatistical methods; Multiple-point statistics**

**Introduction**

While outcrop models can provide important information on reservoir architecture and heterogeneity, it is not entirely clear how such information can be used exhaustively in geostatistical reservoir modeling. Traditional, variogram-based geostatistics is inadequate in that regard since the variogram is too limiting in capturing geological heterogeneity from outcrops. A new field, termed multiple-point geostatistics does not rely on variogram models; Instead, it allows capturing structure from so-called “training images”. Multiple-point geostatistics borrows multiple-point patterns from the training image, then anchors them to subsurface well-log, seismic and production data. Nevertheless, multiple-point geostatistics does not escape from the same...
principles as traditional variogram-based geostatistics: it is still a stochastic method, hence relies on the often forgotten principles of stationarity and ergodicity. These principles dictate that the training image used in multiple-point geostatistics cannot be chosen arbitrarily, and that not all outcrops might be suitable training image models (Caers & Zhang, 2002).

Variogram-free geostatistics - Training image
One often refers to exhaustive datasets, such as outcrop analogs, to extract the required variogram parameters such as nugget, range and direction of anisotropy. But if such an abundance of information is available from outcrops, why restrict oneself to mere two-point correlations. Reproduction of curvilinear features from outcrops would require at least correlations between three spatial locations at a time and possible more, namely: multiple-point correlations or multiple-point statistics. Multiple-point geostatistics relies on the concept of training images (Caers & Zhang, 2002).

A training image is essentially a database of geological patterns, from which multiple-point statistics, including the variogram, can be borrowed. The training image replaces the variogram in multiple-point geostatistics as a measure for geological heterogeneity, it contains multiple-point information and, more importantly, is much more intuitive since one can observe, prior to any geostatistical estimation/simulation, what patterns will be reproduced in a set of multiple reservoir models (Caers & Zhang, 2002).

Figure 17: Three possible candidates for a training image. (a) elliptical shapes, (b) fluvial type reservoir, (c) deltaic type reservoir. Only image two can be used as a training image since pattern are stationary over the entire image.

Geostatistical principles - Stationarity and ergodicity
Figure 17 shows three possible candidates for a training image: a set of elliptical shapes, a fluvial channel reservoir and a fan-like deposit. Under the principle of stationarity, only training image 2 can be used as a training image. Recall that training images are essentially databases to store patterns. If patterns are to be extracted from it, we will need to rely, as any statistical method does, on the stationarity of that pattern over the entire domain covered by that image. In other words, enough repetitivity is required to estimate or extract a set of multiple-point statistics from the training image. It is clear that training image 1 does not follow the principle of stationarity in the sense that patterns are changing over the training image. Consistency of patterns/texture is required. Training image 3 appears to be reasonably stationary in terms of its pattern but a
closer inspection reveals that channels are thinner in the SE corner than in the NW, also the angle at which channel features occur varies from 0° azimuth to 90°, hence non-stationarity is present in the anisotropy of local patterns.

Geostatistical simulation algorithms reproduce input statistics such as a variogram and histogram under certain “ergodic fluctuations”. These fluctuations are due to the limited, finite extent of the spatial domain being simulated. Ergodic fluctuations are not due to some geostatistical/ mathematical trick that allows sample statistics to float due to their uncertainty. Simulation on an infinitely large domain will result in statistics of a realization that exactly match the model statistics. Therefore, when simulating on a finite domain, some statistics have smaller variations than other. For example the variogram statistics of a realization for small lag distances typically show less variability than the variogram for larger lag distances. The size of the training image is an important factor. More importantly, the relative size of the training image with respect to the largest feature/pattern to be reproduced in the actual reservoir model needs to be carefully considered before any simulation takes place. This size consideration is required under the principle of ergodicity. When estimating parameters from a finite domain or when simulating reservoir models on a finite region, any statistics, including multiple-point statistics tend to fluctuate and that fluctuation becomes larger as the distances over which these statistics are calculated becomes larger. For example, reproduction of large scale patterns (long correlations) such as channels (training image 2), would require a large training image, at least double the reservoir size in the direction of the channels continuity is recommended. A small training image would lead to large fluctuations of large range correlations, hence channels will tend to break up if the training image is taken the same size or smaller than the actual reservoir size. If smaller features are to be generated, such as in training image 1, a much smaller training image is required. In such case the training image can have a size much less than the actual reservoir (Caers & Zhang, 2002).

The "SNESIM" algorithm is a sequential simulation algorithm reproducing the patterns of the training image and at the same time honors any well- or seismic data. See appendix A for more information about this algorithm.

**Introduction to the geostatistical software program SGeMS**

The geostatistical methods introduced in this chapter are applied for the modeling of the sand partings within the 6C Frimmersdorf lignite seam in the Garzweiler operation. For these geostatistical calculations the software program SGeMS is used. The Stanford Geostatistical Modeling Software (SGeMS) is an open-source computer package for solving problems involving spatially related variables, offering a wide variety of geostatistical algorithms (Remy, Boucher, & Wu, 2009).
Chapter 6

The available data of the 6C Frimmersdorf lignite seam - analyzing and processing the data in order to make it applicable for geostatistical modeling

First things first
Data collection should never be an aim in itself. Data collected without any clear purpose are solely bytes on a computer or drawers filled with files. Purposeful data collection presupposes a theory or a model which gives meaning and significance to raw data by processing them as information that is useful and needed for a specified goal. In the context of this project, this goal is the accurate prediction of the composition of a lignite seam.

Introduction
This chapter describes the following data types available for creating the geological model of the Garzweiler deposit;

- Production data
- Surveying data
- Drill data
- KOLA data

First, several preconditions for the modeling process are described, in order to use all the data as efficient as possible. For the first three data types, a description of the acquisition of the data is given. Furthermore, the data are processed to make it applicable for the geostatistical methods and several initial geostatistical steps are taken. A description of the KOLA data - including the potential benefits and the current obstacles preventing the implementing of the data - is subsequently given.
Focus area in the Garzweiler mine and the subdivision into grid cells

The whole Garzweiler mine - i.e. Garzweiler I and II together - comprises of over 100 m$^2$. However, during this project the attention is on a 2 x 2 km focus area. This exact focus area is chosen for covering an area in which both a produced and an unproduced part of the 6C Frimmersdorf lignite seam lies, as well as an area containing many sand intrusions and a part with less sand intrusions in this seam. See the figure below for a map of the Garzweiler mine; the red square is indicating the focus area. The focus area is large enough to contain a sufficient amount of data available for the geostatistical modeling and realistic predictions of the occurrence of sand partings in the Frimmersdorf lignite seam. On the other hand, even though the datasets are composed out of thousands of data points, the created model is still tractable.

![Focus area in the Garzweiler mine and the subdivision into grid cells](image)

Figure 18: An aerial map of the Garzweiler open cast mine - the red square is indicating the 2 x 2 km focus area (RWE internal data).

To pinpoint exact locations within the focus area, the easting and northing coordinates used in the operations of RWE, are applied in this project as well. The easting coordinates - i.e. the direction from west to east - are all located between the coordinates 32000 and 34000. In northing direction - i.e. the direction from south to north - the focus area is between the coordinates 59000 and 61000. This can be seen in Figure 19.
The 2 x 2 km focus area is divided into multiple grid cells. In both the easting and northing direction these grid cells are 50 m long, so it gives 40 x 40 = 1,600 grid cells in the whole focus area. This grid is applied for all the data modeling and the geostatistical estimations, see Figure 19. This subdivision into cells of 50 x 50 m is done to pool data to be able to work more efficient with all the different data points. However, if the size of the grid cells is enlarged, the resolution of the modeling will reduce. For this project the applied sizes meet those requirements - when new geostatistical modeling methods are actually implemented, these sizes can be adopted to achieve a higher resolution.

RECOMMENDATION:
Reduce the dimensions of the grid cells in order to increase the resolution of the geostatistical modeling results.
Data types and their own strengths - spatial occurrence and unused data

First, a difference between the drill data on one hand and the surveying- and production data on the other hand must be made. The production data as well as the surveying data can only be gathered of areas where the Frimmersdorf lignite seam is actually excavated. This in contrast to the drill data, which can be acquired from unmined areas outside the borders of the current mine pit. The consequences of this difference are described in this chapter as well.

Furthermore, a distinction between the surveying data and drill data on one hand and the production data on the other hand must be made. Namely, the first two data types are currently used as input data for the modeling process, where the production data are currently not used. This additional data type is omni-present in the mined areas of the focus area and can therefore be a valuable addition to the geostatistical modeling techniques.

Analysis of the three available data types

For all the data types a description of the raw data is given, followed by the processing steps in order to make them applicable for the geostatistical modeling. Several preliminary geostatistical steps are taken as well, to see and compare the possible results of these techniques.

Production data - "always present, never used for geological modeling and hence great potential"

Combining two raw databases to create a new useful dataset

The production data - both of the lignite and the waste material - are currently not used for any geological modeling purposes. In order to do so, the information about the produced material is to be combined with the original location of that material. Therefore, the production data used in the context of this research are composed out of two databases. First, the production data of the excavator 285 producing in the Frimmersdorf lignite seam. This database contains a lot of data;

- **Shift time**: The month, the day and the time in minutes the material was produced.
- **Material type**: Roughly a distinction between coal and sand - to be discussed later.
- **Tonnage (t) and volume (m$^3$)** of the produced material - and therefore the corresponding specific gravity (t/m$^3$) of the material.
- And furthermore some specifications about the material (e.g. water content %) and the excavator (e.g. relative position to the production face), which is currently not used in the processing of the data.

The second database which is used to create an applicable production dataset is the GPS location of the excavator. On top of the wheel of the excavator a GPS device is located, measuring its position every two seconds. This GPS location is vertically shifted downwards to simulate the exact position of the material which is produced (see Figure 20). Since the timeframe in which the material is produced, as well as the actual location of the bucket wheel during this interval, the production data are complete and can be used in the processing steps.
An important strength of this data type is the fact that it in principle contains information about the material type on every location within the mined part of the focus area. However, unfortunately the production data are only available for the years 2011 and 2012. The GPS data of 2010 have been stored for 2 years and are now deleted causing a gap in the data of the focus area (see Figure 21). In total it gives over 30,000 production data points.

**Unique time codes**

Raw datasets for both the GPS locations and the production data supplied by RWE Power contain time data in different formats. In order to correspond these datasets correctly, an unique time code for every data point is created, in the form of mm-dd-time.

*For example, the time codes 11170935 - 11171134 are indicating the production shift starting November 17th at 9:35 AM and ending on the same day at 11:34 AM.*

![Figure 20: GPS location of the produced material - the second database required for the production data](image)

![Figure 21: Spatial distribution of the production data - the 2 x 2 km focus area indicated by the orange square, the missing 2010 data by the red area](image)
Material types of the production data

The database containing all the production figures is supplying information about the material which is produced in a particular shift. The different types of material which are present in the analyzed dataset are listed below:

- **410** Sand general - part of the 6B or 6D sand seam
- **610** Clastic sediments with less than 30% clay
- **730** Intermediate sand layer in the Garzweiler lignite seam
- **740** Intermediate sand layer in the Frimmersdorf lignite seam
- **980** Frimmersdorf coal - low ash content - low critical constituents
- **982** Frimmersdorf coal - high ash content
- **984** Frimmersdorf coal - high critical constituents

Material type **740: Intermediate sand layer in the Frimmersdorf lignite seam** is notable, since it by itself already indicates when sand partings were produced where Frimmersdorf coal was expected. Modeling the spatial distribution of this material type indicates therefore the presence of sand partings. See Figure 22 - in which a difference is made between produced material < 7,500 t (small dot) and > 7,500 t (large dot). It is clear that the sand partings are located predominantly in the southern part of the mined area of the focus area. This material is also used in the determination of the production recovery to estimate the sand presence within the lignite seam.

![Figure 22: The spatial distribution of the material 740: sand partings in the Frimmersdorf lignite seam - the big dots indicate a production block > 7.500 t, the small dots indicate a production block < 7.500 t](image-url)
Clean up the production dataset

For every production shift the minimum and maximum GPS locations of the excavator are estimated. Since the production database contains over 30,000 data points, this is an immense work to do manually. However, Excel 2007 can supply these estimates of the minimum and maximum GPS values using an array-formula. These minimum and maximum easting, northing and depth values are creating a 3D-block representing the mined production block. The volume of this generated production block is compared with the raw data of the actual produced volume. If the volume of the generated block is much larger than the original volume, the excavator has been moving without producing. The exact location of the produced material can therefore never been determined and that actual production shift is removed from the dataset. In several cases, the excavator was just producing for a short period, which generates no useful data since the generated production block will be very small. Therefore, all the shifts with a tonnage smaller than 100 t are removed from the dataset as well. Finally, a failing GPS system can cause the minimum and maximum value of a certain dimension to be the same within a shift. Since properly pinpointing the exact location of the produced material is impossible, these data points are removed as well. After foregoing clean-up of the dataset, and only focusing on the 2 x 2 km area, the total amount of useful shifts is around 2,000 and 1,600 for 2011 and 2012, respectively.

Introduction of the grid cells and estimation of the coal recovery

The 50 x 50 m grid cells introduced in the first part of this chapter are applied to order the production data. These grid cells and the created production blocks are to be combined to generate an insight in the ratio sand partings vs. lignite in every grid cell. For every dimension - X, Y and Z - the ratio of production block in a particular grid cell is estimated.
Now for every grid cell the contained production blocks are known, as well as the type of material of these blocks. Together, this gives a sand vs. coal ratio - see Figure 23. However, in order to only focus on the Frimmersdorf lignite seam and the containing sand partings, this ratio must be sharpened to the following formula;

\[
\text{Production recovery} = \frac{\text{total coal}}{\text{total coal} + \text{sand partings}}
\]

In RWE material type codes; \[\text{Production recovery} = \frac{(980 + 982 + 984)}{(980 + 982 + 984 + 740)}\]

For every grid cell containing production blocks this production recovery is estimated. The results are depicted in Figure 24 - an overview of all the mined grid cells and their estimated production recoveries. The mined area is divided into 4 regions - to be discussed in chapter 7. The graph on the right shows the average recovery of the cells in the corresponding northing row. It can be seen that the production recoveries in the northern part of the focus area are significantly higher than in the southern part.

**Improvements of the creation of the production dataset**

The current approach to convert the two databases - production data and GPS data - into one production dataset serves its purpose; a preliminary appraisal of geostatistical modeling based on the production data. However, in order to use these data on a large scale, an automation step is required. The fact that all the raw data are yet digitally available assures a high potential for automation. This automation step needs to make the processing in Excel redundant, since for larger and more continuous processes this program will not be able to work efficiently.

Furthermore, an enhanced accuracy can be achieved by creating production blocks which more accurately follow the 3D movement of the excavator. Currently, simple rectangular blocks - depending on the minimum and maximum X/Y and Z dimensions - are used. However, all the GPS data within a particular time slot of a shift should be used to create a more realistic 3D volume of the produced material.

**RECOMMENDATION:**

*Improve the approach of combining the production dataset and the GPS location of the excavator in order to more realistically simulate the 3D movement of the bucket wheel and the volumes produced*
Figure 24: The estimated coal recovery per grid cell based on the production data - divided into multiple regions which are to be discussed later. The graph on the right indicates the average recovery of the corresponding row of grid cells.
**Geostatistical modeling of the production data**

**Available geostatistical modeling methods**

The combined production dataset will be used in the geostatistical modeling of the 6C Frimmersdorf lignite seam. The fact that the production data have a high spatial density makes it well applicable for two geostatistical modeling methods: **ordinary kriging** and **indicator kriging**. Ordinary kriging will estimate the expected recoveries of the mining blocks to be produced in the near future, based on the actual production recoveries. Indicator kriging will supply the probabilities for a 100% lignite recovery of the mining blocks to be produced in the near future.

**SGeMS input data for ordinary kriging and indicator kriging**

To use the production data correctly in the geostatistical software program SGeMS, the current dataset must be modified. The midpoints of the grid cells are used to create new data points for SGeMS, containing the following information:

- Easting / Northing / Property

This property is depending on the estimation method applied; for ordinary kriging the production recovery is used. For indicator kriging the input property is a 1 - representing a grid cell with 100% recovery - or a 0 - representing a grid cell with no 100% recovery. Figure 25 gives an overview of these grid cells containing information about the estimated production recovery (left) and the indicator approach (right). The two results of the variogram modeling of the production data can be found in Figure 26 and Figure 27.

![Diagram](Image)

**Figure 25: All grid cells indicate a production recovery (left) or the indicative value 0/1 (right)**
Figure 26: Variogram of the production data - recovery

Figure 27: Variogram of the production data - indicator
Results of ordinary kriging and indicator kriging

In Figure 28 and Figure 29 the visual results of the ordinary kriging and indicator kriging are depicted, respectively. The applied grid in which the kriging is estimated is of the same size as the grid used in Excel - 50 x 50 m cells. The results of this estimation method are bipartite; the left graph represents the estimated recoveries and the right graph represents the corresponding variances of these estimates. The estimated recoveries in Figure 24 are not solely located in the produced area, but predictions are made as well for multiple cells - up to 200 m - out of this area. The color coding on the right side of the figure is already giving an indication of these westward predictions. However, SGeMS is able to export these predicted figures - these results and the corresponding variances can be found in appendix B. The predicted recoveries in the northern part of the focus area are very positive; there is mainly low or no sand expected. The southern part - where lower recoveries have been found during the past productions - is expected to host lower recoveries as well. The variance - and therefore the standard deviation - is increasing for a larger distance away of the current mining borders.

Figure 28: The results for ordinary kriging; left the estimated recoveries, right the corresponding variances

Figure 29: The results for indicator kriging; left the indicative results, right the corresponding variances
Surveying data - "time-consuming data acquisition"

Introduction to the surveying dataset
Since the surveying team can only safely map a face when the bucket wheel excavator is not operating in that area, the interval between two subsequent cross-sections varies between 50 m and 150 m on average. For the focus area - easting 32000-34000 northing 59000-61000 - of this research the spatial distribution of 3,300 surveying data points is shown in the following Figure 30.

The present sand intrusions are mapped using a theodolite which generates an accurate 3D description; easting, northing and the depth below sea level (NAP) - i.e. the X-,Y- and Z-coordinates - are determined for several key points. The applied accuracy of these measurements is up to mm-scale. However, only sand intrusions with a thickness of >30cm are measured.
In comparison to the production data - for which the produced material has been estimated - the surveying data are giving an indication of the location of the border between stratigraphic layers. The 6C Frimmersdorf lignite seam is vertically enclosed by the underlying 6B Frimmersdorf Sand and the overlying 6D Neurath Sand. The discovered sand intrusions are furthermore all grouped in up to four sand intrusion groups, named "Bergmittel" B, C, D and G. Finally, sand channels called "Rinne" can be present in the upper part and the lower part of the 6C lignite seam. All these stratigraphic layers and the twelve borders differentiating those layers are depicted in the following schematic overview - see Figure 31.

Figure 31: A schematic overview of the stratigraphic layers which can be present in the 6C Frimmersdorf lignite seam
Creation of extra surveying points to strengthen the presence of sand partings

The sand partings found in the newly exposed faces are of great importance for the modeling of these impurities - and therefore for the whole path of this research. Hence, the surveying data indicating these partings are very valuable. The use of these surveying points in the current data processing procedures in place at RWE differ from the intensive use during the geostatistical modeling in this project. Therefore it is desirable to increase the amount of surveying data - and the corresponding resolution - provided that the data remains realistic.

The surveying data points were divided into two parts; one part containing loose points which are not related to any other point in the database, the other part containing points which are actually located on the same border of a stratigraphic layer. This second part is useful in the context of this data processing part, since on these borders between original surveying points, new surveying points are created. In the 3D-middle of two existing surveying points a new point is created to strengthen the sand lens. The number of data points increases therefore from the original number of 3,300 to finally 5,000 data points. See Figure 32 for a visual explanation.

![Figure 32: A schematic view of the creation of new surveying points - on the 3D-average location of two original surveying points](image)

Introduction of the grid cells and estimation of the coal percentage

The 50 x 50 m grid is applied to this dataset as well. For every grid cell the average depth of all the modeled layers is estimated;

- The 6C000 and the 6D000 layers are indicating the floor and the roof of the 6C Frimmersdorf seam, respectively. Combining the depth of these two layers results in the average thickness of the whole coal seam in that particular grid cell.
- The remaining ten layers are symbolizing the floor or the roof of the intermediate sand lenses. If the floor and the roof data of all the six sand lenses is in their turn combined as well, the average thickness of all these sand lenses can be estimated. The sum of these thicknesses is the total sand present in every grid cell.
- Finally, coal percentage is the desired output of this modeling. This percentage can be estimated by:

\[
\text{Coal percentage} = \frac{\text{total thickness} - \text{sand thickness}}{\text{total thickness}}
\]

\[
\text{Sand percentage} = 1 - \text{coal percentage}
\]

The surveying data are not available for all the grid cells, since there is often a significant interval distance between two subsequent surveyed faces.
Creation of dummy surveying points - "no sand present = no sand modeled"
The following Figure 33 - left graph - indicates the presence of the surveyed faces - small grey dots representing the surveying points for the top and bottom borders of the whole 6C seam - and the presence of the floor data for sand lens B - large colored dots indicating the depth of that 6C180 layer. It can be expected that in every situation where the face is surveyed, the presence of this stratigraphic border - in this case the 6C180 layer - is analyzed as well. This implies that it can be assumed that on the other locations where the survey team has been measuring, this sand parting is not present. This situation may seem obvious, however, any estimation tool applying these input data will generate unrealistic results - see the right graph of Figure 33.

![Figure 33: Wrong input data will lead to wrong ordinary kriging results](image)

The solution for the described problem is the creation of dummy surveying points at locations where the face is surveyed, but no stratigraphic layer - in this example, no lens B border - is present. This immense task is even in Excel 2007 undoable and therefore the numerical computing program Matlab is introduced. In Matlab several algorithms are written to make the following calculations;

The 1,600 data points of the 6C000 layer are representing the floor of the Frimmersdorf lignite seam and are measured during every surveillance. Based on these 1,600 X- and Y coordinates, all other stratigraphic borders are analyzed and for all these border a new database of 1,600 points is created. For every 6C000 coordinate it is calculated whether there is a border of a particular lens present within a horizontal range of 30 m - so in that whole vertical column.

- If a lens is present - apply the coordinates of these floor and / or roof surveying points to fill the new database.
- If a lens is not present within this 30 m range - apply the X- and Y coordinates of the 6C000 data and complete this with a new estimated depth for both the floor and the roof of that lens. It is of importance that the floor and the roof of the analyzed lens are assigned the same depth-coordinate in case there is
no lens present and new coordinates are generated. This to assure that the thickness of that lens is zero, which implies that there is actually no sand lens present.

- This new generated depth is a moving average; it calculates the average depth of that border within a range of 50 m. This moving average is introduced to assure that the volume of the lens will not unrealistically increase due to "spikes" - e.g. see in Figure 34 what might happen in modeling the lens when a depth of zero is assigned to the newly created lens coordinates. If there are no input values for this moving average - i.e. if that particular lens is not within 50 m - than the average of the overall lens depth is applied. Since the lens is already "closed" - i.e. has a thickness of zero - this vertical shift in depth will not create any spikes in the volume measurements.

- See Figure 35 for a schematic overview of the creation of the dummy surveying points.

![Diagram](image1)

**Figure 34:** A significant shift in vertical distance for the modeled surveying points might lead to the unwanted creation of unrealistic volume

![Diagram](image2)

**Figure 35:** Schematic overview of the modeling of the floor and roof depths of a sand lens
In Figure 36 the kriging results of the depth of the 6C180 layer are shown. Combining the kriging results of the floor and the roof datasets of all sand partings is resulting to valuable information. Since the created dummy points of the floor and the roof of a particular sand lens are set to the same depth, subtracting these two results will eliminate these points and only leave behind the actual sand lens input. Subtracting the roof dataset from the floor dataset will give the thickness of that sand lens. If this approach is repeated for all the six sand lenses, the total thickness of all the sand can be estimated. The thickness of this sand compared to the total thickness of the seam - which is known based on the 6C000 and 6D000 datasets - is resulting in the sand percentage of all the grid cells in the focus area.

**Geostatistical modeling of the surveying data**

**Available geostatistical modeling methods**

After foregoing data processing steps the surveying dataset is available for geostatistical modeling. Here, comparable with the production data, both **ordinary kriging** and **indicator kriging** are applied. For both methods, the surveying dataset will be used to predict a couple of mining cuts forward. Ordinary kriging will estimated the expected coal percentage of the mining blocks to be produced in the near future. Indicator kriging will supply the probabilities for a 100% lignite recovery of the mining blocks to be produced in the near future.
Figure 37: The variogram of the surveying data region 2 - applied to all ordinary kriging processes

Figure 38: The variogram of the indicative surveying data region 3
SGeMS input data for ordinary kriging and indicator kriging

In Figure 39 the spatial distribution of the input data for the ordinary kriging (left) and the indicator kriging (right) are depicted. Comparable with the approach applied for the production data, here the variogram model is of importance for the future estimation methods. The variogram is currently not based on the coal percentage of every grid cell, but based on the depth of the different stratigraphic layers. Since during the creation of the dummy surveying points new points are added to the database - which will lead to unrealistic variogram modeling - the variogram must be based on the original dataset. The variogram of the surveying data in region 2 is shown in Figure 37.

Results of ordinary kriging and indicator kriging

The visual results of the ordinary kriging and indicator kriging are depicted in Figure 40. The color coding indicates a relative low coal percentage, compared with the prediction results of the production data. The actual predicted values are to be exported and validated in the next chapter.
Drill data - "pioneering data containing valuable ash content information"

Introduction to the drill dataset
The RWE Garzweiler open cast mine is an operation in which the drilled wells are used for multiple purposes. The easting and northing of all wells is available and combining the absolute drilling depth (m) and the depth below sea level (depth NAP (m)) of the surface where the drilling was conducted supplies finally the depth NAP (m) of the drill data. The drill data supply information about the coal percentage within a drill core comparable with the coal recovery for the production data and the coal percentage for the surveying data. However, the drill data have an added value since it also contains information about the ash content (%) in the 6C Frimmersdorf lignite seam. Different sources within the RWE business units are consulted to gather the drill data of all these wells to create a large database with close to 8,000 data points. These data are filtered on their easting and northing locations to focus on the 2 x 2 km area. Since in this research the focus is on the spatial distribution of the sand partings in the 6C Frimmersdorf lignite seam, the comprehensive drill data are narrowed down to only contain the following parameters; easting / northing / depth NAP / sample size / ash content.

Drillholes - sample sizes and ash contents
The 71 drillholes available in the focus area are all composed out of many different samples. The average length of the drillhole present in 6C Frimmersdorf seam is 13.5 m and the sample size is on average 0.75 m long. Every sample is indicating a different ash content - ranging between 0.8 and 85% - the actual distribution of the ash content can be found in Figure 41. Within the Garzweiler operation, samples with a higher ash content than 16% are regarded as sand partings - and inevitably, samples with a lower ash content are regarded as coal. This ash content threshold of 16% is applied during this project as well.

Figure 41: The positively skewed distribution of the ash content of the drill samples
**Coal percentages, weighted average ash content and horizontal slicing**

The determined ash content threshold of 16% is supplying a way to distinguish between coal and sand within the drill cores. If the ratio of the sample size containing coal and sand is estimated, the coal percentage for that drillhole is known. A drillhole with solely coal samples will result in a coal percentage of 1. The average coal percentage for the all the 71 drillholes combined is 0.96 - for comparison; the average coal percentage for the production figures is 0.85, the cause of this difference will be discussed in a later stage.

The weighted average ash content for every drillhole can be calculated by taking the sample size and the corresponding ash content value into account. However, as shown in Figure 42, this weighted average ash content is only slightly affected by the occurrence of a sand lens, since the height of this sand lens is often small compared to the total length of the drillhole in the coal seam.

![Figure 42: The weighted average ash content is currently only slightly affected by the presence of a - mostly - low sand lens](image)

To improve the resolution of this approach, the drill data are divided into multiple horizontal slices. For any estimation of the ash content during the further course of this project, the drill data are therefore divided into seven slices. For the upper and lower 3m of the drillhole within the 6C seam, composites of 1m are introduced. The remaining drill length in between these 6 slices is combined in one middle slice. However, note that any resolution can be achieved - this particular division is for example purposes only.

**Geostatistical modeling of the drill data**

Available geostatistical modeling methods

As with the two foregoing described data types, the ordinary kriging and indicator kriging methods are applied to the coal percentages. Furthermore, the spatial distribution of the drill data makes several other geostatistical methods interesting for usage as well. The sequential Gaussian simulation techniques are introduced for simulating both the coal percentages and the ash content.
SGeMS input data for ordinary kriging, indicator kriging and sequential Gaussian simulation

The spatial distribution of the 71 drillholes within the focus area is depicted in Figure 43. For both the ordinary kriging and indicator kriging the general input parameters are used, namely the coal percentage of the entire drillhole within the 6C seam for the ordinary kriging and the corresponding indicative information - 1 for 100% coal, 0 for not 100% coal i.e. sand within the drillhole - for the indicator kriging.

Figure 41 shows the positively distributed input data about the ash content. Since sequential Gaussian simulation requires a Gaussian distributed dataset, this original dataset has to be transformed. The software program SGeMS has a tool for this transformation, resulting in the Gaussian distribution shown in Figure 44. A Gaussian distribution is characterized by a mean of 0 and a variance of 1.
Results of ordinary kriging, indicator kriging and sequential Gaussian simulation

The left graph within Figure 45 shows the results of the ordinary kriging process. Here, the estimated coal percentage within a drillhole is used to determine the expected coal percentages in the grid cells around it. The right graph within Figure 45 represents the results of the indicator kriging process. The major sand intrusions in the southern area of the mined part of the focus area are visible in both graphs.

Sequential Gaussian simulation (SGSIM) is a simulation method; based on the input parameters - the Gaussian distributed dataset and the corresponding variogram - it creates multiple simulations of the reality. These simulations are created on a 50 x 50 m grid - comparable with estimation methods like kriging. In SGeMS this simulation technique is executed, resulting in 20 simulations of the possible distribution of the Gaussian dataset. However, these simulations do not supply any information about the actual ash contents in these grid cells. Therefore, the data in these simulations have to be transformed back to their original (positively skewed) distribution. This tool is available in SGeMS as well, requiring only the insertion of the minimum and maximum values and the distribution of the original weighted average ash content dataset. This transformation is shown in the following figure.
The 20 simulations made during the SGSIM step can be combined in order to create a final result of the try to model the actual distribution of the ash content. This combination gives the probability for every grid cell that it is below (or above) the threshold of 16% ash content; so the probability of each grid cell to be coal or sand. These results are depicted in the following Figure 47.

Figure 47: The probability of coal occurrence (left) or sand occurrence (right) in the middle slice

Figure 48: Sequential Gaussian simulated ash content of the whole drill length in the 6C lignite seam for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content.
Finally, Figure 48 is showing the results of the sequential Gaussian simulated ash contents (y-axis) for 40 mining grid cells (x-axis). These 40 mining grid cells represent roughly a monthly production of one bucket wheel excavator. The grey graphs represent the 20 simulated ash contents, the green graph represents the average of these simulated ash content and the red line indicates a possible threshold. This threshold can be set at any desired ash content. Note that no variance in the simulated ash content of all 20 simulations is present in two grid cells, indicated with a blue circle. These grid cells contain a drillhole which leads to certain ash content within these grid cells for all 20 simulations. The variance is increasing for grid cells further away from drillholes.

All evaluations and descriptions of potential benefits of these results are described in the next chapter.

**KOLA data**

The KOLA - an abbreviation for Kohle OnLine Analytics - data are another data type available for a more extensive modeling of the sand parting in the 6C Frimmersdorf lignite seam. It applies X-Ray diffraction in order to accurately assess the components of the produced lignite. The analyzed components are inter alia iron, sulfur, potassium, calcium and - of importance in the context of this research - the ash content of the produced lignite. For the period 2011/2012 this dataset contains > 300,000 samples. The Garzweiler opencast mine operates multiple KOLA measuring station, of which two are analyzing the coal from the Frimmersdorf lignite seam.

The KOLA data can supply a significant added value to the geostatistical modeling of the sand partings. Currently, the production data containing only the coal recovery are dually used; 1) they are used as input data for the geostatistical predictions of the coal recovery of the closely located mining blocks, and 2) they are used for the validation of all geostatistical predictions of the expected coal recovery, since it actually represents the real production figures. The scenario that the production data contain - besides the estimated coal recovery - also an estimation for the ash content will actually result in two improvements;

1. **The production data can be used for geostatistical estimation of the ash content**
   Currently, only the drill data contain information about the ash content. The new scenario will greatly increase the data density of this parameter and therefore the accuracy of the geostatistical estimations of the ash content.

2. **The production data can be used for a validation of the current geostatistical estimations of the ash content based on the drill data only**
   The geostatistical estimations of the ash content based on the drill data currently lack any possibility to be validated with actual produced figures. A proper application of the KOLA data will supply a solution for both the estimations based on the drill data and the estimations based on the production data itself - provided that improvement 1 is realized.
Although the potential advantages of the KOLA data are significant, a proper application of this data type is facing - unfortunately - too many obstacles. Firstly, it is very hard - if not impossible - to locate the exact origin of a sample, since the analyzed material is coming from different lignite layers, different operating bucket wheel excavators located at different locations from the measuring station. The current dataset - containing only the moment of measuring - is not able to resolve this. In addition, the analysis is only executed on material which is produced as lignite, not for material produced as waste. For an accurate modeling of the sand partings the ash content of the waste material within the lignite seam - i.e. the actual sand partings! - needs to be determined as well. Finally, the practical interpretation of the measuring process is causing a flaw as well. The samples are taken from the conveyor belt using a 30mm sieve. The irregular size distribution of coal / sand will cause a skewed representation of the reality - i.e. the sand is expected to mostly be contained within fragments larger than 30 mm.

**RECOMMENDATION:**
Analyze and - if possible - overcome the practical obstacles withholding the implementation of the valuable KOLA data
Chapter 7

Validation of geostatistical predictions based on the available data - Comparison of the geostatistical predictions with the actual production figures in the mined area

Introduction

The results of the geostatistical methods applied to the available data types are validated in this chapter. The validation process is crucial in the context of the overall project, since it gives an indication of the quality and practical application of the used methods. The quality is analyzed by taking four different figures into account:

1. **Geostatistical predictions** introduced in chapter 5
   = The predicted producible coal recoveries, coal percentage and coal volumes based on the different applied geostatistical methods.

2. **Actual production figures** introduced in chapter 6
   = The actual coal recoveries and coal volumes produced in the 6C Frimmersdorf lignite seam.

3. **Empirical error of prediction**
   = The difference between the geostatistical predictions and the actual production figures.

4. **Theoretical error of prediction** introduced in chapter 5
   = The theoretical error of the geostatistical predictions, related to the dimensions of the predicted zone and an estimate of the estimation error.
The predicted recovery is compared with the actual recovery, which results in the empirical error of prediction. Subsequently, this empirical error is compared with the theoretical error of prediction. The extent to which these figures correspond is an indication of the quality of the geostatistical prediction.

Overview data types and applied geostatistical methods
This chapter will give an analysis of the validation process of the three data types. Since most of the steps in this validation process will be the same for all three methods, only one extensive description will be given for the ordinary kriging and indicator kriging estimations. This description is given on the basis of the production data; the repeated step for the other data types can be found in appendix C. However, the analysis of the results is given for all three data types. At the end of this chapter the sequential Gaussian simulation is discussed, also on the basis of a full description of the applied approach.

- Production data - ordinary kriging Full description
- Production data - indicator kriging Full description
- Surveying data - ordinary kriging Analysis of validation results
- Surveying data - indicator kriging Analysis of validation results
- Drill data - ordinary kriging Analysis of validation results
- Drill data - indicator kriging Analysis of validation results
- Drill data - sequential Gaussian simulation Full description
Input regions for different moments in time based on the mining sequence

For the production data and surveying data, the overall dataset available in the focus area is based on figures acquired in 2011 and 2012. This area is divided into four regions (see Figure 49); regions 1 to 3 are used as input data for the validating predictions. Region 4 cannot be used for these predictions, since there will not be any production figures to compare the predictions with. These different regions can be seen as different moments in time during the progression of the open pit and all available data at that time will be employed to model the deposit. Therefore, it is evident that a region is also containing the data of the foregoing region, since in this way more input data are available.

The general parameter modeled during the geostatistical process is the coal recovery - for the production data - or the coal percentage - for the surveying data and the drill data. However, to generate a more useful outcome, this recovery is combined with the thickness data of the 6C seam for every grid cell to result in a prediction for the producible coal volumes. In this way the comparisons between the predictions and the actual production figures can be made for (weighted average) recoveries and (total) coal volumes.

*Figure 49: The mined part of the focus area is divided into four regions based on the mining sequence - representing different moments in time*
Validations made for different support regions - all with a practical relevance

The validation comparisons are made for three different comparison zones. It is expected that the absolute value of the error between prediction and actual figures will decrease with an increasing size of the support zone. Three different cases are differentiated, which have as well a practical relevance. Ordered from large to small these zones are:

1. The total predicted coal volume based on the data in one region
   These predictions are roughly representing a quarter-yearly production

2. The predictions of the producible coal volumes within a mining row
   A mining row is representing one bucket wheel excavator pass and approximately a monthly production.
   A row contains 40 50 x 50 m grid cells covering the whole length of the 2 km focus area from north to south. Here, row 1 contains the first 40 grid cells adjacent to the input region, row 2 contains the next 40 grid cells, etc. - see Figure 50 for a visual explanation.

3. And finally, the comparisons are made for all the 50 x 50 m grid cells separately. This comparison zone is representing a daily production of one bucket wheel excavator. The results are visualized in a scattergram; the x-axis depicts the predicted coal volume and the y-axis depicts the actual coal volume produced.

Note that a validation is only possible if a prediction exist for a particular grid cell and the grid cell is actually produced as well. Therefore can the amount of validated predictions based on different regions change as well.
Analysis of the validation process and results

Introduction
The following part of this chapter describes the necessary steps in order to validate the results of the ordinary kriging and indicator kriging of the production data. For the surveying data and drill data these steps are similar and therefore only the results of these validation process of these two datasets are described.

Production data - Ordinary kriging

Input data for the geostatistical modeling software program SGeMS
The datasets used for the different regions are containing three parameters; X / Y / coal recovery figures of the production data. For the surveying data and the drill data the third parameter is the coal percentage instead of the coal recovery - as discussed in chapter 2. The amount of data points is increasing for the regions later in time. Region 1 contains 164 data points, region 2 contains 310 data points and region 3 contains 446 data points. The spatial distribution and the recoveries per grid cell are shown in the following Figure 51: Input production data (left) for the three regions and the corresponding ordinary kriging results (right).Figure 51 Fout! Verwijzingsbron niet gevonden..

Variogram modeling of the production data
The creation of accurate variograms is required for a good estimation method. For all the different regions a variogram is generated. For a lower distance the variogram can quite accurately be connected to the actual
data points in the variogram model. For larger distances these points are more dispersed, caused by a trend component of the data, as discussed in chapter 5. The three variograms are listed in Figure 52.

**Ordinary kriging of the production data**
The visual results of the subsequent ordinary kriging step can be seen in the three overviews depicted at the right side of Figure 51. These kriging results are based on the variogram range resulting in a prediction range around 150 m. The prediction of a lower recovery in the southern part (blue) and a recovery of 100% in the northern part (red) of the focus area is clearly seen. The actual predicted recoveries are exported from SGeMS and processed in the next validation steps.

**Production data – Ordinary kriging**

*Figure 51: Input production data (left) for the three regions and the corresponding ordinary kriging results (right)*
Figure 52: The variogram results of the production data for the three regions
Validation results of the production data - comparisons of estimation results, theoretical figures and actual production figures

Validation overview of the production data - all predictions combined

In the table depicted in Figure 53 the overall results of the ordinary kriging validation are listed. A comparison of the predictions with the actual production figures is only possible if the predicted grid cell is mined as well. Therefore, the amount of the predicted cells depicted in the second column is changing for the different regions. In total there are 264 grid cells predicted and produced and are therefore useable for this validation process.

The difference between the predicted and actual figures in percentage terms depicted in the seventh column is the empirical error of prediction (%). Evidently, these differences are the same for the coal recovery as for the produced coal volumes, since the linear transformation from recoveries to volumes will not change this mutual ratio. The last column shows the theoretical error of prediction (%); implying a boundary in which 68.2% of the predictions are expected to be located. The quality of the prediction can be assessed by checking the percentage of predictions for which the empirical error is within the limits set by the theoretical error. The table gives the predicted and actual coal productions in million m³ and the predicted and actual coal recovery in a dimensionless way.

For both predictions based on region 1 and 3 the empirical error is within the limits of the theoretical error. However, the results for the predictions based on region 2 are showing a large difference between these two errors - an empirical error of -12.68% is estimated where the theoretical error sets the limits at 3.44% difference. The predictions based on region 2 rely on the low recoveries within this region, caused by the presence of major sand partings. When the occurrence of these sand partings decreases significantly within the prediction range of region 2, these predictions will show a large overestimation of the sand - i.e. a large underestimation of the coal volumes. This observation shows an important limitation of the ordinary kriging method - see the textbox on the next page for additional information. A more accurate view on these predictions is given in the following parts of this chapter; focusing on the predictions for multiple mining rows and for all the grid cells separately.

![Figure 53: The validation results of the ordinary kriging of the production data - divided into the predictions based on the different regions](image-url)
Limitations of the ordinary kriging method

The observation of the inaccurate predictions based production data in region 2 is showing a limitation of the applied ordinary kriging method. The available production data are sorted, combined and used to predict the future mining blocks. In the context of the stationarity assumption introduced in chapter 5, the ordinary kriging is only ‘limited’ able to follow sudden ‘regional’ or ‘structural’ changes. In this situation the (dis)appearing sand lens is affecting the stationarity of the dataset.

The next part about the row validation shows that the empirical error for the further located rows is increasing quickly. A continuous re-entry of the available data will improve the predictions and in this way decrease the empirical error in changing geological environments.

RECOMMENDATION:

Realize an automation step which applies a continuous re-entry of the available data. This automation step will generate a model where continuous input and usage of newly available data will result in the most up-to-date geostatistical predictions and a decrease of the empirical error.
Validation overview of the production data - predictions divided into multiple rows

Figure 54 shows the validation results of the geostatistical predictions grouped in rows containing 40 grid cells in south-north direction. Here, row 1 is the first adjacent row to the input data and the consecutive rows are the 40 grid cells located further away from the input data. The table gives for these different rows the predicted and actual coal productions in million m$^3$. The empirical error of prediction (%) of these two figures is described in the sixth column. This actual difference can be compared with the theoretical error of prediction (%) listed in the last column.

8 predictions are made for the producible coal volumes within a mining row, of which the empirical error of 4 of those predictions lie within the limits set by the theoretical error of prediction - i.e. 50% of the predictions. This is significantly less than the 68% of the predictions expected to lie within these theoretical limits by looking to a normal distributed set of predictions and its standard deviations. The predictions based on region 2 are all off; the cause of this is already described in the foregoing part of this chapter. Furthermore, it is seen that the predictions for the first row are matching the actual productions better than the consecutive rows. The differences shown in the last column are increasing for the further rows in all the regions. A continuous re-input of the production data will therefore be a valuable tool in the geostatistical modeling of the sand lenses.

<table>
<thead>
<tr>
<th>Predictions based on region</th>
<th>Row 1</th>
<th>Row 2</th>
<th>Row 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictions based on region 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 1</td>
<td>39</td>
<td>0.12</td>
<td>NA</td>
</tr>
<tr>
<td>Row 2</td>
<td>4</td>
<td>1.12</td>
<td>NA</td>
</tr>
<tr>
<td>Row 3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Predictions based on region 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.12</td>
<td>1.00</td>
</tr>
<tr>
<td>Row 2</td>
<td>40</td>
<td>1.14</td>
<td>1.18</td>
</tr>
<tr>
<td>Row 3</td>
<td>36</td>
<td>1.00</td>
<td>1.18</td>
</tr>
<tr>
<td>Predictions based on region 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.31</td>
<td>1.33</td>
</tr>
<tr>
<td>Row 2</td>
<td>38</td>
<td>1.23</td>
<td>1.22</td>
</tr>
<tr>
<td>Row 3</td>
<td>27</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>All predictions combined</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row 1</td>
<td>119</td>
<td>3.55</td>
<td>3.69</td>
</tr>
<tr>
<td>Row 2</td>
<td>82</td>
<td>2.49</td>
<td>2.68</td>
</tr>
<tr>
<td>Row 3</td>
<td>63</td>
<td>1.90</td>
<td>2.06</td>
</tr>
</tbody>
</table>

Figure 54: The validation results of the ordinary kriging of the production data - divided into predictions for multiple rows
Validation overview of the production data - predictions for all grid cells separately

The three scattergrams shown in Figure 56 - region 1, Figure 57 - region 2, and Figure 58 - region 3, are showing the results of the ordinary kriging estimations for all the grid cells separately. Ideally, the predicted production volumes (m$^3$) are comparable with the actual production volumes (m$^3$) and are therefore located on the blue diagonal line $x = y$. The theoretical error of prediction for every grid cell separately is the regular standard deviation estimated from the variogram, so without any division by the square root of the amount of grid cells. The weighted average of these standard deviations is used as the theoretical error of prediction and represented by the orange and green lines. The orange lines represent the 1σ-limits and the green lines represent the 2σ-limits. The slope is estimated by using the theoretical error - for the 1σ-limits that is:

$$y = (1 \pm \sigma)x$$

And for the 2σ-limits that is:

$$y = (1 \pm 2\sigma)x$$

Referring to Figure 55, one can expect that 68.2% of the data points can be found within the borders of the orange 1σ-lines and 95.4% of the data points within the green 2σ-lines. The actual distribution of the data points representing the predictions vs the actual figures is summarized in Figure 59. Referring to Figure 56, it is seen that many data points within region 1 are not near the diagonal $x = y$ and that the limits of both σ-lines are exceeded. Since overestimating and underestimating predictions are both made based on the production data in region 1, the overall difference between the predicted coal production and the actual coal production is differing only ca. -2.5% (see Figure 53). For both the prediction based on regions 2 and 3, many of the data points are also located near the blue diagonal line. However, the prediction based on region 2 shows many data points left of the diagonal, implying an underestimation of the producible coal volumes.
**Region 1:**

Figure 56: The predicted coal production vs the actual coal production - shown for all the grid cells of region 1 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit.

**Region 2:**

Figure 57: The predicted coal production vs the actual coal production - shown for all the grid cells of region 2 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit.
Figure 58: The predicted coal production vs the actual coal production - shown for all the grid cells of region 3 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit.
Another added value of analyzing these grid cells separately is that now these outliers can be located. Firstly, the data points located outside the 2σ-borders are marked purple. Subsequently, the actual location of these outliers in our focus area can be determined. (see Figure 60).

It is seen that the majority of the outliers is located in the southern part of the focus area. When a closer look is applied to Figure 60, it is noticeable that the grid cells showing a large empirical error of prediction are located in areas where structural changes appear suddenly. This can be seen as well by comparing the predictions based on region 2 with the actual production volumes (see Figure 53). Reliable prediction within these changing environments are - as mentioned earlier in this chapter - a limitation of the ordinary kriging method.
Production data - Indicator kriging

Introduction
The indicator method requires a different approach. Now the input data do not contain any direct information about the coal recovery, but an indirect, binary system in which:

- 1 represents a grid cell containing only coal i.e. no sand present
- 0 represents a grid cell containing not only coal i.e. sand present.

The resulting indicator kriging estimates are indicating the probability that a certain grid cell contains only coal - i.e. the grid cell is an 1 in the indicator kriging approach. Since these probabilities are not comparable with the actual production figures - as opposed to the ordinary kriging results - another approach is applied. Five different probability ranges are composed; probability 0 - 0.2, 0.2 - 0.4, 0.4 - 0.6, 0.6 - 0.8 and 0.8 - 1. In this way, all the grid cells of the indicator kriging results are categorized into these 5 probability intervals. Remember that in reality the actual production figures can only be a 0 or an 1. For all these zones the amount of 1's is estimated; the percentage of these 1's compared to the total amount of grid cells in this zone should be close to the probability interval.

The datasets used for the different regions are containing three parameters; X / Y / indicator data (see Figure 62). The amount of data points for all the regions is the same as for the ordinary kriging approach.

Variogram modeling of the production data
The variogram of region 1 is based on too few data points and is therefore not applicable for the subsequent modeling steps. Therefore, only one variogram - the variogram based on the data from region 2 - is applied for the following estimations steps of all the three different regions. This variogram fits well to the data points and the geology of the three regions is comparable, justifying this decision.

![Figure 61: The general variogram model for the indicator approach of the production data - based on the data of region 2](image-url)
**Indicator kriging of the production data**

The indicator data and the corresponding variogram allow a larger range used in the estimation step. For all the three different regions a prediction range of 350 m is applied. The visual results of this kriging step is shown in the following three figures, in which the color coding symbolizes the probability for that grid cell to host solely coal - i.e. in the indicator approach that corresponds to contain the value 1.

**Production data – Indicator kriging**

**Figure 62: The input data (left) and the corresponding indicator kriging results of the three regions (right)**

- **Input data for region 1**
- **Indicator kriging results for region 1**

- **Input data for region 2**
- **Indicator kriging results for region 2**

- **Input data for region 3**
- **Indicator kriging results for region 3**

**Zones of interest:**

- Deep red: High probability for 100% coal recovery
- Uncertain predictions – additional exploration necessary
- Deep blue: High probability for reduced coal recovery caused by the occurrence of sand lenses
Validation results of the production data - comparisons of indicator kriging results, theoretical figures and actual production figures

The results for the validation process of the indicator kriging are comparable in a less extensive manner than those for the ordinary kriging. The numerical results of the kriging process are divided into five zones of different probabilities. The probability that only coal occurs in a grid cell in these zones - i.e. the probability of the occurrence of a 1 in the indicator approach - is therefore correlated to the probability range of these zones. In this way, a validation of the predictions can be made by comparing the percentage of 1’s in the overall predicted area. Referring to the table below, this validation approach can be summarized by comparing the actual probability (# grid cells with 100% coal recovery / # predicted grid cells) with the expected probability of that range (column 1).

<table>
<thead>
<tr>
<th>Probability range</th>
<th># predicted grid cells</th>
<th># grid cells with 100% coal recovery</th>
<th>Actual probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 0.2</td>
<td>290</td>
<td>34</td>
<td>0.12</td>
</tr>
<tr>
<td>0.2 - 0.4</td>
<td>100</td>
<td>17</td>
<td>0.17</td>
</tr>
<tr>
<td>0.4 - 0.6</td>
<td>76</td>
<td>31</td>
<td>0.41</td>
</tr>
<tr>
<td>0.6 - 0.8</td>
<td>59</td>
<td>34</td>
<td>0.58</td>
</tr>
<tr>
<td>0.8 - 1</td>
<td>30</td>
<td>27</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Figure 63: The validation results of the indicator approach of the production data

The validation approach for the indicator method is more devious than for the ordinary kriging method, but in general the validation step is showing good results. The actual probability depicted in the last column is fairly good comparable with the pre-set probability range. The outlying value of the second range - 0.2-0.4 - is mainly located near the major sand lens in the southern part of the focus area. However, this indicator approach will not deliver results for a one-to-one comparison of each production block, but rather give an expectation of the occurrence of sand lenses in a particular zone. By looking to the lower indicator kriging results shown in Figure 62 mainly three zones of interest can be distinguished:

- The pure blue zone predicted in the southern part of the focus area. Here a reduced recovery caused by the occurrence of sand lenses can be expected.
- The pure red zone predicted in the northern part of the focus area. Here a (nearly) 100% coal recovery can be expected.
- The occurrence of sand lenses in the zone in between those two areas is yet unsure. Extra attention can therefore be given to this area, to make the general expectation for the future mining blocks complete. Additional auger exploration drilling can be a good solution for these areas; a cheap and easy method which will deliver examples of the material to be produced quickly.
Surveying data - Ordinary kriging

For both the surveying data and the drill data not a comprehensive description of all modeling steps is given, solely an analysis of the results of the different geostatistical approaches. Chapter 2 of this report partly deals with the practical conversion of coal percentage to coal recovery. The relationship between those two figures is of importance for this data type, since the surveying data (and the drill data) contain information about the coal percentage, where a validation comparison is made to the actual coal recovery. For a visual explanation of this matter, see Figure 11 chapter 2. Here, the coal percentage is assumed to be the expected coal recovery.

Validation results of the surveying data - comparisons of estimation results, theoretical figures and actual production figures

Validation overview of the surveying data - all predictions combined

The validation results of the ordinary kriging of the surveying data are showing an overall underestimation of the coal volumes, as can be seen in Figure 64. This indicates that the geostatistical model estimates more sand in the lignite seam than actually is present.

For the predictions based on region 1 the total predicted coal production seems to fit the total actual coal production very well. However, when looking to the grid cells separately in a later stage of this chapter it is seen that it is actually a combination of underestimations and overestimations. The other two regions are showing a large underestimation of the coal volumes.

<table>
<thead>
<tr>
<th>Predictions based on</th>
<th>Predicted coal production (million m³)</th>
<th>Predicted weighted average coal recovery (%)</th>
<th>Actual coal production (million m³)</th>
<th>Actual weighted average coal recovery (%)</th>
<th>Error</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>region 1</td>
<td>4.56</td>
<td>0.78</td>
<td>4.56</td>
<td>0.78</td>
<td>0.01</td>
<td>+/- 5.61</td>
<td></td>
</tr>
<tr>
<td>region 2</td>
<td>4.56</td>
<td>0.77</td>
<td>5.24</td>
<td>0.88</td>
<td>-12.95</td>
<td>+/- 5.26</td>
<td></td>
</tr>
<tr>
<td>region 3</td>
<td>3.42</td>
<td>0.76</td>
<td>4.13</td>
<td>0.92</td>
<td>-17.18</td>
<td>+/- 5.46</td>
<td></td>
</tr>
<tr>
<td>All predictions</td>
<td>12.55</td>
<td>0.77</td>
<td>11.04</td>
<td>0.85</td>
<td>-9.07</td>
<td>+/- 5.28</td>
<td></td>
</tr>
</tbody>
</table>

Figure 64: The validation results of the ordinary kriging of the surveying data - divided into the predictions based on the different regions

This underestimation of the coal volumes can be caused by different factors.

- The most obvious reason for the underestimation of the coal can be found in the conversion of the conversion of the 2D thickness to the 3D volume applied during this project. This approach assumes that the 2D thickness of a sand lens visible and measured on the mining face is present in the entire 50 x 50 m grid cell where it is located in (see Figure 35 - chapter 6). This approach will overestimate the actual sand lens volume and therefore underestimate the coal volumes.
Referring to the coal percentage vs actual coal production section in chapter 2, the underestimation of the coal volumes can be visually explained by the following two figures (see Figure 65). 1) Minor sand lenses which are detected by the surveying crew are not recognized by the bucket wheel excavator operator or are too small to separate with the bucket wheel excavator. 2) The edges of sand lenses are produced as coal since it is hard to separate accurately between these two layers, as discussed in chapter 2. Both these situation will increase the production of coal compared to the initial coal percentage detected by the surveying team - with increasing ash contents in the coal as a consequence.

![Small sand lenses are too small to separate or not recognized and produced as coal](Image)

![Edges of sand partings are produced as coal](Image)

Figure 65: Visual explanation of the potential causes of the underestimation of the coal volumes

Furthermore, the original data from the surveying crew can be inaccurate. The experience that climate conditions can influence the conspicuousness of the sand partings and the raised chance of human errors in the interpretation- and measuring process are adverse points for the use of this data type in an accurate modeling process. The measuring is done manually and the amount of data points per sand lens are far from abundant, which could lead to incorrect modeling of the sand lenses. However, an overall underestimation of the coal volumes by 10% is unlikely to be caused by any measuring error.
Validation overview of the surveying data - predictions divided into multiple rows

Figure 66 shows the validation results for the different rows of all regions. For all the three regions only the variogram of the data of region 2 is applied (see Figure 67), since for the other two regions it is hard to fit a proper variogram and the geological setting is not expected to differ significantly between the different regions. This causes the fact that the theoretical error of prediction can be the same for the different regions, provided that the amount of predicted grid cells is similar as well.

The empirical error of prediction depicted in the sixth column of Figure 66 shows again a large difference with this theoretical error of prediction. In the following part of this chapter the data will be represented in a graphical way, which simplifies the understanding of this dissimilarity.

<table>
<thead>
<tr>
<th>Predictions based on region 1</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m$^3$)</th>
<th>Actual coal production (million m$^3$)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.17</td>
<td>1.15</td>
<td>1.67</td>
<td>+/- 4.00</td>
</tr>
<tr>
<td>Row 2</td>
<td>40</td>
<td>1.15</td>
<td>1.08</td>
<td>5.99</td>
<td>+/- 4.80</td>
</tr>
<tr>
<td>Row 3</td>
<td>40</td>
<td>1.17</td>
<td>1.11</td>
<td>4.63</td>
<td>+/- 5.43</td>
</tr>
<tr>
<td>Row 4</td>
<td>38</td>
<td>1.09</td>
<td>1.17</td>
<td>-7.42</td>
<td>+/- 5.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictions based on region 2</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m$^3$)</th>
<th>Actual coal production (million m$^3$)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.14</td>
<td>1.23</td>
<td>-7.88</td>
<td>+/- 4.00</td>
</tr>
<tr>
<td>Row 2</td>
<td>40</td>
<td>1.13</td>
<td>1.32</td>
<td>-13.96</td>
<td>+/- 4.80</td>
</tr>
<tr>
<td>Row 3</td>
<td>40</td>
<td>1.14</td>
<td>1.32</td>
<td>-13.88</td>
<td>+/- 5.43</td>
</tr>
<tr>
<td>Row 4</td>
<td>39</td>
<td>1.08</td>
<td>1.29</td>
<td>-16.44</td>
<td>+/- 5.77</td>
</tr>
<tr>
<td>Row 5</td>
<td>2</td>
<td>0.07</td>
<td>0.07</td>
<td>-4.51</td>
<td>+/- 26.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictions based on region 3</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m$^3$)</th>
<th>Actual coal production (million m$^3$)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.08</td>
<td>1.33</td>
<td>-19.03</td>
<td>+/- 4.00</td>
</tr>
<tr>
<td>Row 2</td>
<td>38</td>
<td>1.03</td>
<td>1.22</td>
<td>-16.05</td>
<td>+/- 4.92</td>
</tr>
<tr>
<td>Row 3</td>
<td>32</td>
<td>0.88</td>
<td>1.05</td>
<td>-16.30</td>
<td>+/- 6.07</td>
</tr>
<tr>
<td>Row 4</td>
<td>17</td>
<td>0.44</td>
<td>0.53</td>
<td>-16.84</td>
<td>+/- 8.74</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>All predictions combined</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m$^3$)</th>
<th>Actual coal production (million m$^3$)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>120</td>
<td>3.38</td>
<td>3.71</td>
<td>-8.93</td>
<td>+/- 4.00</td>
</tr>
<tr>
<td>Row 2</td>
<td>118</td>
<td>3.31</td>
<td>3.62</td>
<td>-8.70</td>
<td>+/- 4.84</td>
</tr>
<tr>
<td>Row 3</td>
<td>112</td>
<td>3.18</td>
<td>3.49</td>
<td>-8.69</td>
<td>+/- 5.61</td>
</tr>
<tr>
<td>Row 4</td>
<td>94</td>
<td>2.61</td>
<td>3.00</td>
<td>-12.98</td>
<td>+/- 6.34</td>
</tr>
<tr>
<td>Row 5</td>
<td>2</td>
<td>0.07</td>
<td>0.07</td>
<td>-4.51</td>
<td>+/- 26.46</td>
</tr>
</tbody>
</table>

Figure 66: The validation results of the ordinary kriging of the surveying data - divided into predictions for multiple rows
Validation overview of the surveying data - predictions for all grid cells separately

A visual overview of the validation comparisons of the separate grid cells can clarify the uncertain situation of the surveying data. The scattergrams in Figure 68, Figure 69 and Figure 70 are showing the graphical comparison between the predicted coal volume (m$^3$) and the actual production (m$^3$), in which a clear underestimation of the coal volumes can be noticed. However, if another regression than $x = y$ would be used - indicated with the vertical shift of the $x = y$ line to the dashed blue line in Figure 68 and Figure 69 - the uncertainty would be lower. It is therefore clear that a **bias** is present. Three possible causes of this bias are described in the beginning of this part of the chapter; however, based on the available data the exact causes cannot be determined.

\[
\text{Empirical error of prediction} = \text{uncertainty} + \text{bias}
\]

It must be noted that the empirical error shown in the validation tables in this chapter is not simply the same as the uncertainty of that prediction. In fact, a potential **bias** must be taken into account as well. Based on the available data, the cause of the bias present for the predictions based on surveying data cannot be explained. Possible reasons for this bias are described in the beginning of this part of the chapter.
Region 1:

Figure 68: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 1 separately. The orange lines symbolize the $1\sigma$-limit and the green lines symbolize the $2\sigma$-limit. The dashed blue line indicates a regression other than $x = y$ which might lower the uncertainty.

Region 2:

Figure 69: The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 2 separately. The orange lines symbolize the $1\sigma$-limit and the green lines symbolize the $2\sigma$-limit. The dashed blue line indicates a regression other than $x = y$ which might lower the uncertainty.
Region 3:

**Figure 70:** The predicted coal percentage vs the actual coal production - shown for all the grid cells of region 3 separately. The orange lines symbolize the 1σ-limit and the green lines symbolize the 2σ-limit.

**Figure 71:** The amount and percentage of grid cells within the sigma-borders for the three regions.

The standard deviation of the distribution of the separate grid cells of the surveying data is causing a much larger range within the σ-limit than for the production data. However, the amount of grid cells within these limits are for four of the six cases still more than theoretically expected, as can be seen in Figure 71.

Similar to the location of the outliers of the production data, the grid cells showing the largest difference between the prediction and the actual figures are located near the sand lens in the southern part of the focus area (see Figure 72).
Improvements for the data acquisition and data processing of the surveying data

To improve the approach to convert the available surveying data into useful data for the geostatistical modeling, the data acquisition and data processing of the surveying data needs to be changed. The three recommended changes are listed below:

1. Measure more surveying points on the floor surface of the sand lens, and particularly at the roof surface of the sand lenses. Currently, only the outer ends of the sand lenses are recorded. Advised is to record more data points in between the currently existing points, which will lead to a much more realistic modeling of the sand lenses.

2. Realize a zero thickness of the sand lenses on locations where the sand lens is not present, by setting the roof- and floor surface to the same depth. This will simplify the simulation of the sand lenses. Practically, this means that the most outer surveying points in horizontal direction should be included in both the floor- and roof surface dataset.

3. Currently, the depth of a surface is determined by estimating the average depth within a mining grid cell. Assuming the introduction of changes 1 and 2, the roof- and floor surfaces of the sand lenses can be simulated in order to create a more realistic model of the sand lenses.
**Surveying data - Indicator kriging**

The indicator method of the surveying data is based on the same disputable data modeling approach as for the ordinary kriging method, so the expectations for these results should not be too high. The kriging results are visually depicted in Figure 74 and the subsequent validation for the five probability ranges is shown in Figure 73. Almost all grid cells - 90% - are located in the probability range 0 - 0.2, indicating that in a large area the expectations for any sand lens occurrence are high. For the remaining four ranges - 0.2-0.4, 0.4-0.6, 0.6-0.8 and 0.8-1 - it is hard to draw any solid conclusions, since the low amount of grid cells in these ranges. However, a large discrepancy between the expected probability - first column - and actual probability - last column - is noticeable, indicating the presence of a bias, which cannot be detected based on the data available.

![Surveying data - indicator kriging validation](image)

*Figure 73: The validation results of the indicator approach of the surveying data*

![Surveying data – indicator kriging](image)

*Figure 74: The indicator results based on the surveying data in region 3*
Drill data - Ordinary kriging
The last data type to be validated is the drill data. Compared to the foregoing two data types, the drill data have the advantage that it is not bounded to the mining limits, but available for the whole focus area. First the ordinary kriging and indicator kriging results are validated; the last part describes shortly the simulation approach and subsequently the (validation) results of the simulation method.

For both the surveying data and the drill data not a comprehensive description of all modeling steps is given, solely an analysis of the results of the different geostatistical approaches. Chapter 2 of this report partly deals with the practical conversion of coal percentage to coal recovery. The relationship between those two figures is of importance for this data type, since the surveying data (and the drill data) contain information about the coal percentage, where a validation comparison is made to the actual coal recovery. For a visual explanation of this matter, see Figure 11. Here, the coal percentage is assumed to be the expected coal recovery.

Predictions based on a region vs. predictions within a region
As opposed to the production data and the surveying data - for which the predictions are only available in regions based on the data of a particular regions - the predictions of the drill data are available for the whole focus area.
Therefore, extra attention must be paid for the location of a prediction;
- All predictions combined = predictions within regions - for a complete overview of the data
- Validation of rows = predictions based on regions - for a better comparison with the other two data types

Validation results of the drill data - comparisons of estimation results, theoretical figures and actual production figures
Validation overview of the drill data - all predictions combined
The validation results of the ordinary kriging of the drill data are shown in Figure 75. Here, the predictions and the actual production figures within the four regions are compared. The theoretical error of prediction of the drill data cannot be estimated with the help of the variograms and the lag distances of each consecutive row. Therefore, the kriging variances produced in SGeMS are transferred to the standard deviation per grid cell and used for the calculation of this theoretical error. For regions 1,2 and 3 the empirical error is larger than the theoretical error.

An overall overestimation of the coal volumes is seen. This indicates that a certain percentage - for the four regions on average 13.5% - of the sand intrusions is not detected by the current wells. The cause of this overestimation of the coal volumes is described in a later stage of this chapter.
### Drill data - Ordinary kriging validation

<table>
<thead>
<tr>
<th># predicted grid cells</th>
<th>Predictions</th>
<th>Actual figures</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted coal production (million m³)</td>
<td>Predicted weighted average coal recovery (%)</td>
<td>Actual coal production (million m³)</td>
</tr>
<tr>
<td>Predictions in region 1</td>
<td>159</td>
<td>5.82</td>
<td>0.95</td>
</tr>
<tr>
<td>Predictions in region 2</td>
<td>146</td>
<td>5.08</td>
<td>0.95</td>
</tr>
<tr>
<td>Predictions in region 3</td>
<td>136</td>
<td>4.79</td>
<td>0.97</td>
</tr>
<tr>
<td>Predictions in region 4</td>
<td>126</td>
<td>4.42</td>
<td>0.98</td>
</tr>
<tr>
<td>All predictions combined</td>
<td>567</td>
<td>20.10</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Figure 75:** The validation results of the ordinary kriging of the drill data - divided into the predictions in the different regions

**Figure 76:** The ordinary kriging results based on the drilling data
Validation overview of the drill data - predictions divided into multiple rows

The validation results divided into different rows are shown in Figure 77. Here the predictions are based on regions - as opposed to Figure 75, where the predictions are made within the regions. This is for a better comparisons with the validation results of the other data types. As can be seen in the last two columns of the table, the predictions based on region 1 are showing the largest difference between the theoretical and empirical error of prediction. The predictions based on the other regions are better; the empirical error is mostly within the borders of theoretical error of prediction. Overall, of the 12 predictions made for the different rows, the empirical error of 7 of those predictions are within the limits set by the theoretical error - a percentage of 7/12 or 58%. This is not supporting the theoretical expectations of 68% of the predictions within the 1σ-limit.

<table>
<thead>
<tr>
<th>Predictions based on region 1</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m³)</th>
<th>Actual coal production (million m³)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.42</td>
<td>1.16</td>
<td>23.00</td>
<td>+/- 12.83</td>
</tr>
<tr>
<td>Row 2</td>
<td>40</td>
<td>1.39</td>
<td>1.11</td>
<td>25.55</td>
<td>+/- 12.99</td>
</tr>
<tr>
<td>Row 3</td>
<td>40</td>
<td>1.41</td>
<td>1.13</td>
<td>25.06</td>
<td>+/- 12.99</td>
</tr>
<tr>
<td>Row 4</td>
<td>40</td>
<td>1.42</td>
<td>1.24</td>
<td>13.82</td>
<td>+/- 12.90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictions based on region 2</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m³)</th>
<th>Actual coal production (million m³)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.43</td>
<td>1.23</td>
<td>15.91</td>
<td>+/- 12.94</td>
</tr>
<tr>
<td>Row 2</td>
<td>40</td>
<td>1.42</td>
<td>1.32</td>
<td>7.99</td>
<td>+/- 13.23</td>
</tr>
<tr>
<td>Row 3</td>
<td>40</td>
<td>1.44</td>
<td>1.32</td>
<td>8.75</td>
<td>+/- 13.60</td>
</tr>
<tr>
<td>Row 4</td>
<td>40</td>
<td>1.42</td>
<td>1.32</td>
<td>7.05</td>
<td>+/- 13.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictions based on region 3</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m³)</th>
<th>Actual coal production (million m³)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>40</td>
<td>1.41</td>
<td>1.33</td>
<td>6.26</td>
<td>+/- 14.33</td>
</tr>
<tr>
<td>Row 2</td>
<td>38</td>
<td>1.33</td>
<td>1.22</td>
<td>8.50</td>
<td>+/- 14.99</td>
</tr>
<tr>
<td>Row 3</td>
<td>32</td>
<td>1.12</td>
<td>1.05</td>
<td>6.56</td>
<td>+/- 16.73</td>
</tr>
<tr>
<td>Row 4</td>
<td>16</td>
<td>0.56</td>
<td>0.53</td>
<td>5.60</td>
<td>+/- 24.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>All predictions combined</th>
<th># predicted grid cells</th>
<th>Predicted coal production (million m³)</th>
<th>Actual coal production (million m³)</th>
<th>Empirical error (%)</th>
<th>Theoretical error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>120</td>
<td>4.27</td>
<td>3.72</td>
<td>14.66</td>
<td>+/- 13.37</td>
</tr>
<tr>
<td>Row 2</td>
<td>118</td>
<td>4.14</td>
<td>3.65</td>
<td>13.49</td>
<td>+/- 13.71</td>
</tr>
<tr>
<td>Row 3</td>
<td>112</td>
<td>3.97</td>
<td>3.50</td>
<td>13.36</td>
<td>+/- 14.28</td>
</tr>
<tr>
<td>Row 4</td>
<td>96</td>
<td>3.39</td>
<td>3.10</td>
<td>9.52</td>
<td>+/- 15.18</td>
</tr>
</tbody>
</table>

Figure 77: The validation results of the ordinary kriging of the drill data - divided into predictions for multiple rows
Validation overview of the drill data - predictions for all grid cells separately

Referring to Figure 78, the one-to-one comparison for all grid cells is showing a remarkable result. Most of the data points in this graph are closely located near the diagonal x=y line, as opposed to a situation in which a scattered distribution might lead to the same overall differences between the predictions and the actual figures.

![DRILL DATA - ORDINARY KRIGING VALIDATION](image)

Figure 78: The predicted coal percentage vs the actual coal production - shown for all the grid cells of all the regions separately

The remarkable aspect of these results lies in the fact that all outliers in Figure 78 are representing a situation with overestimated coal volumes. The sand parting distribution in the Frimmersdorf lignite seam is subject to a "high nugget effect" which reveals a limitation of the drill data. This can be explained with the help of Figure 79. Generally, two scenarios can be distinguished;

1. The drillholes hit the existing sand lens - represented by a big red dot - resulting in a proper prediction of the sand.

2. The drillholes do not hit the existing sand lens - represented by a big black dot - resulting in predictions which will not take the sand into account. This inevitably leads to an underestimation of the sand, i.e. an overestimation of the coal volumes.

To reduce the chance of missing a sand lens, an increased drilling density must be applied. It is needless to mention that an increased drilling density is extremely expensive and therefore unrealistic.
Figure 79: Schematic visualization of the two drill scenarios; 1) the drillhole hitting the sand lens, and 2) the drillhole not hitting the sand lens.
Figure 80: Map of outliers of the drill data - most outliers are located near the sand lens in the southern part of the focus area, however, compared to the other two data types the outliers are more scattered.
Drill data - Indicator kriging

The indicator approach of the drill data shows for all the grid cells containing a drillhole whether sand is present in the 6C seam or not. Figure 81 depicted below shows the subsequent kriging results of these indicator input values. For the mined area in the eastern part of the focus area two distinctive areas can be observed. 1) The major sand lens in the southern part - corresponding with the pure blue area indicating a low probability for a 100% coal recovery, or in other words - a high probability of sand occurrence. 2) The northern part in which the pure red zone is indicating an area with a high probability for 100% coal recovery. Both characteristics have been detected multiple times before and are clearly visible using this indicator approach.

Focusing on the unmined area several interesting zones can be analyzed. The presence of a well hitting a sand layer in the 6C seam is indicated with a blue dot in the figure below, and especially the north-western and south-western areas of the focus area are of importance since there multiple wells are hitting sand. These areas would require additional exploration in order to improve the certainty about the occurrence of this sand.

Figure 81: The indicator kriging results of the drill data - in the unmined part of the focus area drillholes are containing sand in north-western and south-western area
As described above, the added value of the indicator approach is not so much the actual prediction of the probability of the occurrence of sand or not, but moreover the distinct zoning of interesting areas. Therefore, the validation results in Figure 82 should not be taken to close. Although the actual probabilities show a large overestimation of the coal volumes for the three middle probability ranges, the indicator method for the drill data is a valuable geostatistical tool.

**Distribution of the empirical error and the relevance of simulation**

When a closer look is applied to the empirical error of the predictions based on the drill data, a remarkable observation can be done. The actual distribution of the empirical error is not symmetric, but shows a positively skewed distribution. Since kriging assumes a symmetric distribution, the applicability of kriging in general can be questioned. The use of the simulation approach - described in the next part of this chapter - could supply a solution.
Drill data - Sequential Gaussian simulation

Sequential Gaussian simulation applied to two drill data parameters

The distribution of the empirical error described with the help of Figure 83 is questioning the application of kriging methods. Simulation, on the other hand, can offer a solution since it does not require a normal distributed dataset. Besides this benefit of the simulation approach, the drill data also contain an added value compared to the other two data types; the examination of the ash content of the drill cores. Therefore, as opposed to the production data and the surveying data, the drill data supply two available parameters:

1. The expected coal recovery

\[
\text{coal percentage} = \frac{\text{coal length of the drill core in seam } 6C}{\text{total length of drill core in seam } 6C}
\]

Combined with the practical conversion of the coal percentage to expected coal recovery, as discussed in chapter 2.

2. Ash content

The drill core is examined for the ash content. This ash content is an indication for the quality of the lignite; a higher ash content will decrease the coal quality. The threshold of an ash content of 16% indicates the border between coal (<16% ash) and sand (>16% ash).

Both these drill data types are applied in the sequential Gaussian simulation. Firstly, the expected coal recovery is simulated 20 times and these simulations are compared with the actual recoveries. Secondly, the ash content of different horizontal slices of the 6C seam is simulated. Modeling those simulations can locate potential mining grid cells with an ash content nearing the threshold of 16% ash, or other thresholds applied.

Simulation process - applicable to both data parameters

First, the drill data are uploaded into SGeMS. In order to properly use the simulation approach, the dataset is transformed into a Gaussian distribution (see Figure 84). Based on the Gaussian distributed drill data, 20 realizations of the expected coal recovery (or ash content) are created. The distribution of these 20 realizations is again transformed back to the original distribution. These results can subsequently be used in different ways, described later in this part of the chapter.
Validation of the sequential Gaussian simulated coal recovery

For the four regions in the focus area, the drill data are employed to realize 20 simulations of the coal recovery per grid cell. The minimum, maximum and average simulated coal recovery - including the corresponding standard deviation - and the actual coal recovery in the four regions, are listed in Figure 85. The last column of this table gives the empirical error (%); which compares the average simulated coal recovery with the actual average coal recovery. Clearly, a systematic error is present in the simulated drill data which can be explained by the fact that the current drill spacing is not able to detect all sand lenses. The fact that more sand partings are present in the east part of the focus area is resulting in the higher empirical error in these regions (region 1 and 2).
Figure 85: The simulated coal recovery figures vs the actual coal recovery within the four regions

Figure 86 gives the results of the simulations for the predictions based on the regions, divided into the simulations for multiple rows. The relative low actual recoveries in the area predicted based on the drill data in region 1 are causing a large empirical error of prediction. Increasing actual recoveries westward induce a better fit with the simulations for the predictions.

The following three scattergrams are showing the simulated coal volume vs the actual coal production per grid cell, for the simulation with the minimum, average and maximum simulated coal recovery, respectively.
Figure 87: The simulated coal volumes with the minimum recovery vs the actual coal production - shown for all the grid cells of all the regions separately.

Figure 88: The simulated coal volumes with the average recovery vs the actual coal production - shown for all the grid cells of all the regions separately.
Analysis of the results of the sequential Gaussian simulated ash content

Since there are not any actual production figures containing the ash content available, the results of the sequential Gaussian simulated ash content cannot be validated. The KOLA data can offer a possibility to validate these results, however - as described in chapter 6 - the current status of the KOLA dataset is far from applicable. Nonetheless, the availability of the ash content offers an added value to the drill data.

To improve the resolution of this simulation approach, the drill data containing the ash content are divided into multiple horizontal slices. The average length of the drillholes in the 6C Frimmersdorf seam is 13.5m and for the upper and lower 3m of the drillhole, composites of 1m are introduced. The remaining drill length in between these 6 slices is combined in one middle slice. However, note that any resolution can be achieved - this particular division is for example purposes only.

The same simulation process is followed as described in the “Simulation process - applicable to both data parameters” part of this chapter; 20 realizations of the ash content in every grid cells - divided into 7 horizontal slices - is the result. These simulations can be employed to model the ash content for a particular mining block. For the remainder of this part of the chapter, the simulation results of the ash content of a 40 cells mining row - representing a monthly production - are analyzed.

Figure 89: The simulated coal volumes with the maximum recovery vs the actual coal production - shown for all the grid cells of all the regions separately
Figure 90: Sequential Gaussian simulated ash content of the upper horizontal slice for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold.

Figure 91: Sequential Gaussian simulated ash content of the middle horizontal slice for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold.

Figure 92: Sequential Gaussian simulated ash content of the whole drill length in the 6C lignite seam for a monthly mining cut. The grey graphs represent the 20 simulated ash contents, the green graph represents the average simulated ash content and the red line indicates a possible threshold.
Figure 90, Figure 91 and Figure 92 are showing the results of the sequential Gaussian simulated ash contents (y-axis) for 40 mining grid cells (x-axis). These 40 mining grid cells represent roughly a monthly production of one bucket wheel excavator. Figure 90 shows the results for the upper horizontal slice, Figure 91 shows the results for the middle horizontal slice and Figure 92 shows the results for the whole drill length within the 6C Frimmersdorf lignite seam. The grey graphs represent the 20 simulated ash contents, the green graph represents the average of these simulated ash content and the red line indicates a possible threshold. This threshold can be set at any desired ash content. Note that no variance in the simulated ash content of all 20 simulations is present in two grid cells, indicated with a blue circle. These grid cells contain a drillhole which leads to a certain ash content within these grid cells for all 20 simulations. The variance is increasing for grid cells further away from drillholes.

The benefits of these simulations are to more accurately pinpoint "risky" mining grid cells. Here, "risky" implies the possibility of increased sand occurrence. Depending on the simulated ash contents and the levels set for the thresholds, different actions can be undertaken;

1. **Different risk levels of sand occurrence**
   The simulations results offers the possibility to divide the daily productions into different levels of probability for sand occurrence. This will increase the comfort of the operator, by knowing better what to expect for the coming daily production.

![Simulated ash content per monthly mining cut - whole drill length in 6C lignite seam](image)

*Figure 93: The simulated ash content can offer an indication of the daily expectation for the risk of sand occurrence*
2. **Operational blending**

By knowing that a particular bucket wheel excavator will be producing in an mining area with an increased chance of sand occurrence, the operational management can decide to locate another bucket wheel excavator in a mining area with lower risks of sand occurrence. This will level the ash contents of the produced material during that period of time and this operational blending will ensure a supply with steady ash contents to the power plants.

3. **Additional exploration**

Finally, the decision can be made to enhance the knowledge of the pinpointed risky area by executing additional exploration. Auger drilling is an example of a relative cheap and quick method to learn about the material to be produced in the near future.
Chapter 8

Conclusions, discussion and recommendations

Introduction
The conclusions set out in this chapter are twofold. With reference to the objectives stated in chapter 4, conclusions are drawn concerning 1) the available data types and 2) the applied geostatistical methods. Subsequently, the resulting insights and the corresponding paths to achieve them are discussed. Finally, several operational recommendations and a future outlook are formulated.

Thesis objectives and the corresponding accomplishments
The purpose of this thesis, as summarized in its title:

"Appraisal of geostatistical methods and geostatistical prediction of predominantly marine sand inclusions in the Frimmersdorf lignite seam in the Garzweiler open cast mine, Germany"

is to investigate the opportunities for RWE Power AG to apply geostatistical modeling at its Garzweiler lignite operation. An analysis of the available data and the possible geostatistical methods - including their strengths and shortcomings - is given. Combining available data and the most suitable methods will finally lead to recommendations for the best geostatistical modeling methods and predictions for the unmined area of the Garzweiler mine.

The five objectives and the main corresponding accomplishments are:

1. Analyze the available data and current data acquisition and data processing in place.
   \textit{All data types are thoroughly described and their strengths and limitations are clearly indicated.}

2. Analyze and compare the possible geostatistical methods.
   \textit{An introductory analysis of the field of geostatistics is given and the most suitable geostatistical methods are selected.}

3. Give geostatistical predictions of these sand inclusions.

4. Give certainties for the producible coal estimations.
Based on available data and using the most suitable geostatistical methods, predictions and corresponding certainties of the coal percentage and ash contents are given.

5. Give recommendations for deposit modeling, future exploration and data processing procedures. Several operational recommendations are given throughout the report and are summarized at the end of this chapter.

Conclusions

Conclusions of the results analyzed in the validation chapter

The conclusions which can be drawn based on the findings in this chapter are twofold. Firstly, the three data types are assessed on the basis of seven quality parameters. Secondly, the selected geostatistical methods are rated.

Assessment of the three data types

The assessment of the three data types is done on the basis of seven quality parameters, as indicated in the overview below. With the introduction of these quality parameters the whole process of geostatistical modeling is covered. A short description of every data type will follow.
Production data - "the sleeping beauty"

The production data have the valuable advantage of being always available within the mined area. This high data density and the fact that it is currently not used for any modeling purposes enhances the additional value of this data type. The production data - composed out of the GPS location of the bucket wheel excavator and the figures for the produced material, both digital information - have a high potential for automation. The spatial occurrence of the data and the combination with the ordinary kriging approach limits the prediction range, however; but with a continuous input of the available data this disadvantage can be reduced.

Two main improvements are to be made:

1. The creation of production blocks during the data processing must be more realistic. The current method of creating a production block based on the minimum and maximum GPS values must be replaced by a method taking into account the actual 3D movement of the bucket wheel.

2. To fully resource the potential of this data type, a proper application of the KOLA data must be realized. The required improvements for the KOLA data are described in chapter 6.

Surveying data - "a lot of work for a bias"

Much time goes into the acquisition and processing of the surveying data, while the benefits for the geostatistical modeling are limited. The experience that climate conditions can influence the conspicuousness of the sand partings and the raised chance of human errors in the interpretation- and measuring process are adverse points for the use of this data type in an accurate modeling process.

The geostatistical predictions based on this data type clearly show a bias, and even though the cause of this bias cannot be fully determined based on the available information, a high expectation of a deficient data processing approach exists. Reducing the dimensions of the grid cells will not completely solve the problem of inaccurate data, but surely lead to a suppression of the symptoms. To further improve the approach to convert the available surveying data into useful data for the geostatistical modeling, the data acquisition and data processing of the surveying data needs to be changed. The three recommended changes are listed below:

1. Measure more surveying points on the floor surface of the sand lens, and particularly at the roof surface of the sand lenses. Currently, only the outer ends of the sand lenses are recorded. Advised is to record more data points in between the currently existing points, which will lead to a much more realistic modeling of the sand lenses.

2. Realize a zero thickness of the sand lenses on locations where the sand lens is not present, by setting the roof- and floor surface to the same depth. This will simplify the simulation of the sand lenses. Practically, this means that the most outer surveying points in horizontal direction should be included in both the floor- and roof surface dataset.
3. Currently, the depth of a surface is determined by estimating the average depth within a mining grid cell. Assuming the introduction of changes 1 and 2, the roof- and floor surfaces of the sand lenses can be simulated in order to create a more realistic model of the sand lenses.

**Drill data - “pioneering data containing valuable ash content information”**

Drill data show a low data density due to the high costs per drillhole. However, these data are very precise and they are available outside the mined areas, which makes them useful for long term predictions, as opposed to the other two data types. Another added value of the drill data is that they contain information about the ash content. The simulated ash content for every grid cell and the introduction of a proper threshold for these ash limits can be a valuable improvement for the modeling process and the operational management. The introduction of the KOLA data is required in order to validate the results of the simulated ash contents.

**Assessment of the applied geostatistical methods**

**Ordinary kriging - “a solid method - calls for a dynamic approach”**

The ordinary kriging method is widely in use and in general it provides good predictions. These predictions are accompanied by an estimate of the estimation error - i.e. the kriging variance. In the context of the stationarity assumption - one of the fundaments of kriging - a significant limitation arises; ordinary kriging is only poorly able to follow sudden “structural” changes. Introducing a continuous, iterative entry of newly available data will generate a permanently updated model which will result in a decrease of the empirical error, also within these changing environments.

**Indicator kriging - “locate zones of interest”**

The indicator approach is not able to give accurate predictions of the parameter investigated, but moreover it can pinpoint zones of interest. The predictions with a very low - i.e. < 0.2 - or very high - i.e. > 0.8 - probability to host purely coal are mostly showing a good match with the actual occurring coal percentages. The intermediate probability ranges are a more uncertain area where additional exploration can be of value.

**Sequential Gaussian simulation - "blends well with ash contents"**

As opposed to kriging, simulation is not requiring a normally distributed dataset. The creation of multiple realizations is giving a better feeling for the potential scenarios of the future mining blocks. Since this geostatistical method is mainly applied to simulate the ash content, the validation step is not (yet) possible. However, combining this method with the ash content data is expected to create a great opportunity for operational management.

**Discussion**

The results of this thesis provide the basis for a significant improvement of the modeling process for the large lignite projects operated by RWE. This may lead to 1) more accurate predictions - including a certainty range - of the producible coal volumes and 2) potential improvements in the operational management with a view to reduce risks of producing low quality lignite and to ensure a steady production level.

The main hindrance experienced during this thesis project was the unavailability of suitable data. Huge amounts of data had to be collected and processed in order to make them suitable for geostatistical modeling.
Two main limitations of this study should be acknowledged. Both limitations are related to the assumption that the processed data in fact represent the actual reality. First, the production data - created by the combination of the production data of the excavators and the corresponding GPS locations - are assumed to represent the actual production figures; but in fact they are just a proxy. Second, the difference in depth between the 6C000 and 6D000 surfaces is assumed to represent the actual thickness of the 6C seam. Now, any inaccuracy in the manual processing of the data produces as a result scenarios which are not fully reflecting reality.

However, a general idea of the potential use of geostatistical modeling - including the encountered obstacles, limitations and strengths - is provided. The application of the labor intensive and therefore expensive application of the surveying data can be questioned, while a proper combination of the production data and the drill data can result in a significant improvement of the modeling approach.

Data collection should never be an aim in itself. Data collected without any clear purpose are solely bytes on a computer or drawers filled with files. Purposeful data collection presupposes a theory or a model which gives meaning and significance to raw data by processing them as information that is useful and needed for a specified goal. In the context of this project, this goal is the accurate prediction of the composition of a lignite seam.

Referring to chapter 3, where the second remark on deterministic vs stochastic modeling stated

- A model is not cast forever. It should be updated whenever the goal of the study changes, additional data become available, or the physics of the phenomenon becomes better known.

The fact that new (production) data become available on a daily basis provides the opportunity for a significant improvement. The production data are always available for the mined area and these data are digitally stored, possessing high potential for automation. This automation step can generate a model where continuous input and usage of the available data will result in permanently updated geostatistical predictions and a reduction of the empirical error. The drill data are not restricted to the mining limits and can therefore be used for the long term planning. Finally, the simulation of the ash content based on drill data can indicate risky areas and a proper use of these data in operational management can ensure reliable production levels.

**Recommendations and future outlook**

Throughout the report several recommendations are formulated. This section summarizes these recommendations and gives an outlook on the future operations when the geostatistical modeling techniques are applied.

1. Improve the approach of combining the production dataset and the GPS location of the excavator in order to more realistically simulate the 3D movement of the bucket wheel and the volumes produced.

2. Improve the approach to convert the available surveying data into useful data for the geostatistical modeling, by measuring more surveying points at both the floor and the roof of the sand partings,
"close" the sand parting by setting the depth of these surfaces to the same level and simulate - as opposed to use an average value - the depth of those surfaces.

3. Analyze and - if possible - overcome the practical obstacles preventing the effective use of potentially very valuable KOLA data.

4. Reduce the dimensions of the grid cells in order to increase the resolution of the geostatistical modeling results.

5. Implement an automation step which generates a continuous re-entry of the available data. This automation step can generate a model where continuous input and usage of the available data will result in permanently updated geostatistical predictions and a reduction of the empirical error.

Upon actual introduction and application of geostatistical modeling techniques, the geological modeling department will not only provide periodic geological forecasts, but it can introduce a permanently updated geological model as well. To fully exploit its potential, this model requires a close cooperation between the geological modeling department and the operational management of the mine.
References


RWTH Aachen, d. o. (2011, November). lecture slides EMC.


Schlumberger. Geoframe Introduction to CPS-3.


Appendices

Appendix A: The "SNESIM" algorithm

Appendix B: Export predictions and corresponding variances based on the production data

Appendix C: Validation results of the surveying data
Appendix A: The "SNESIM" algorithm

The SNESIM - Single Normal Equation Simulation - algorithm reproduces the patterns of the training image and at the same time honors any well- or seismic data (Strebelle, 2000). SNESIM is a sequential simulation algorithm, much in the style of well-known methods such as sequential Gaussian simulation and sequential indicator simulation (Isaaks, 1990); (Goméz & Journel, 1993).

Multiple-point geostatistics relies on the same principle of stationarity. The estimation or construction of a stochastic model relies on the availability of "repeated pattern information". Nevertheless, actual deposits may contain many non-stationary elements. Based on the principles of rotation and affinity transformation one can borrow patterns from a stationary training image and construct non-stationary reservoir models from it (Caers & Zhang, 2002). How this is done is shown in the follow figure - accompanied with a modification of the sequential simulation algorithm discussed in the previous part of this chapter.

1. Construct a fine 3D grid of well-data to closest grid cells
2. Define a random path
3. Until each non-datum cell u on the random path is visited
   1a. Search for closest nearby well data and previously simulated cells. Call this set a "data event"
   1b. Rotate and affinely transform the data event accordingly. The center of such transformation is the visited cell location u to be simulated
   2. From the training image, read a probability for the property to be simulated based on the rotated data from step 1b. and possibly any seismic data
   3. Draw an outcome from the probability model in step 2 and assign that value to the current grid cell

   End simulation
### Estimated coal recovery

Based on the production data

<table>
<thead>
<tr>
<th>Easting</th>
<th>Northing</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
</tbody>
</table>

---

Appendix B-1: Ordinary kriging results of the production data
## Appendix B-2: The ordinary kriging variances of the production data

### Corresponding ordinary kriging variance

#### Based on the production data

| Easting | Northing | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
|---------|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 0.0458  | 0.0452   | 0.0383 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.0441  | 0.0392   | 0.0384 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| 0.0429  | 0.0389   | 0.0386 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |

---

Master of Science Thesis

Frans van Beuningen
Appendix C: The input data (left) and the corresponding indicator kriging results of the three regions (right) of the surveying data.