Advances in Data Spacing and Uncertainty

by

Felipe Alfredo Cabral Pinto

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

Master of Science

in

Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Collecting information is of vital importance for the development of a mineral project. The capital costs of mining projects are high and there is significant risk due to the available data. At all stages of a project, from exploration to mine closure, decisions need to be made that are based off of the data available. Some of the important sources of data include field work, outcrops, geophysical and geochemical measurements, and drilling. The information gathered can be either quantitative or qualitative, and the specific data available depends on each individual project. Of the data typically available, the information gathered from drilling is the most direct approach to understanding the subsurface mineral deposit.

This thesis addresses two important decisions companies face regarding drilling, (1) What is the data spacing needed to achieve an acceptable level of uncertainty for a relevant scale? and (2) What is the rate that uncertainty changes when more data is collected?

There is a level of uncertainty for drillhole spacing in any scale. The level of uncertainty can be shown as a measurement such as the probability of the grade to be within a percentage of the mean, which can be calculated for a relevant production scale. This may suggest a drillhole spacing, that is on average, associated to a desired level of precision. For example, the estimated grade of monthly production volumes falling within ±15% of the mean 80% of the times. With a specific level of precision, the drillhole spacing associated with it increases as the production scale increases.

In certain circumstances, the uncertainty versus data spacing can be established analytically. A "Learning Curve" is established to gain a deeper understanding of how uncertainty relates to data spacing for different spatial structure. The Learning Curve summarizes the rate at which the scale of variability is resolved for additional drilling.

This work also addresses the influence of important explanatory factors on the total variability. Explanatory factors are mostly economic, geologic and geometric factors that explain the variability in a variable. Uncertainty does not depend solely on data spacing (geometric factor), local uncertainty is also influenced by conditional mean, conditional variance (economic factors) and entropy (geologic factor). The influence of these factors explaining uncertainty is modelled by statistical regression techniques.

A comprehensive case study is presented that includes geological and grade modelling. This full data spacing study is practically important to present the concepts reviewed in this thesis and to demonstrate new concepts regarding uncertainty versus drillhole spacing.

"...Lutar, lutar, lutar

Pelos gramados do mundo pra vencer..."

Hino do Clube Atlético Mineiro,

o Galo forte e vingador,

o Galo doido,

o Galão da Massa,

o Galo de ouro, a Seleção do Povo.

Acknowledgements

I would like to thank my supervisor, Dr. Clayton V. Deutsch for his support, patience and love for sharing his knowledge in geostatistics. Your guidance and encouragement made this thesis possible. In addition, I thank the Centre for Computation Geostatistics and sponsors from industry for the financial support of my research.

I would like to thank Dr. Jeff Boisvert for your lectures in geostatistics, Dr. Johnathan G. Manchuk for answering my questions and Alice da Silva for all support given to students. In addition, I am grateful to my colleagues at the Centre for Computational Geostatistics for their advice and assistance during my time at the University of Alberta.

Special thanks to my friends Eric Daniels, Cole Mooney, Patrick Donovan and Diogo Souza, who have been a source of real friendship throughout my time in Canada.

I would like to thank my family for their support and encouragement. Being so far away is only possible with your understanding and love.

Finally, I would like to thank my wife Cíntia for your support, patience and love during the time of writing this thesis.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

ACE	alternating conditional expectations
CDF	cumulative distribution function
CIM	Canadian Institute of Mining, Metallurgy and Petroleum
DHD	drill hole density
DHS	drill hole spacing
DND	distance to the nearest drill hole
EG	equivalent grade
EUC	expected uncertainty curve
EV	economic value
GSLIB	Geostatistical Software Library
HA	hectare
KV	kriging variance
LC	Learning Curve
MAF	Minimum/Maximum Autocorrelation Factors
NPV	Net Present Value
PPMT	Projection Pursuit Multivariate Trasnform
RT	rock type
SGS	sequential Gaussian simulation
SIS	sequential indicator simulation
SMU	selective mining unit
SSE	sum of squares of the errors
SSR	sum of squares of the residuals
SST	total sum of squares
$A(\mathbf{u})$	search area
c	categorical variable

D	domain
$DHD(\mathbf{u})$	drill hole density at a location $\mathbf{u}$
$DHS(\mathbf{u})$	drill hole spacing at a location $\mathbf{u}$
$Exp(\mathbf{h})$	exponential variogram model
ft	feet
$Gaus(\mathbf{h})$	Gaussian variogram model
$H_V(\mathbf{u})$	entropy at the volume $V(\mathbf{u})$
k	realization
K	total number of realizations
m	metres
$\overline{\mu}_V(\mathbf{u})$	conditional mean at the volume $V(\mathbf{u})$
N	number of SMUs found within the volume $V(\mathbf{u})$ over all realizations
n_{dh}	nominal number of drill holes
$n_s(\mathbf{u})$	number of samples found within the area $A(\mathbf{u})$
$n_v(\mathbf{u})$	number of samples found within the volume $V(\mathbf{u})$
$n(\mathbf{u})$	number of closest samples to search at a location $\mathbf{u}$
p_c	proportion of a category c
q	sample spacing along the drill hole
R	rate of uncertainty changing
R^2	coefficient of determination
$R(\mathbf{u})$	approximated length of the edge of a square for DHS calculation
$\mathbf{u}$	location in space
σ_k^2	kriging variance
$\overline{\sigma}_V(\mathbf{u})$	conditional standard deviation at the volume $V(\mathbf{u})$
S	data spacing
$s(\mathbf{u})$	data spacing at location $\mathbf{u}$
$Sph(\mathbf{h})$	spherical variogram model
$U(\mathbf{s})$	expected uncertainty given S
$V(\mathbf{u})$	search volume
$z(\mathbf{u})$	outcome of a random variable at a location $\mathbf{u}$

Chapter 1: Introduction

An important issue in the development of a mineral project is to know the minimum amount of data necessary to support a decision. Decisions are taken considering uncertainty in the geological model, grade estimates, the production scale and other technical factors. The range of decisions in a mine extends from daily grade control to long term planning. All decisions are based on the available data, numerical modeling and the experience of the decision makers.

In the early stages of exploration there are few samples from delineation and reconnaissance drilling. If initial results show promise then more data is collected. CRIRSCO (2013) and JORC (2012) define classification standards for resources based on a quantified level of geological confidence (Figure 1.1). According to the Canadian Institute of Mining, Metallurgy and Petroleum (CIM), the classification of mineral resources is dependent on the "…nature, quality, quantity and distribution of data…" (Postle et al., 2000). Regarding mineral resources classification, changing from inferred to measured requires a better understanding of the geology and improved grade and tonnage estimates. This is achieved with more data.

Infill drilling provides additional information in specific zones of the deposit. At times, the drillholes are regularly spaced at a chosen distance. Consider the simplest case of infill drilling optimization versus regular spacing shown in Figure 1.2. A few initial drillholes are used to generate a first geological model. Lower grade zones have less variability and less uncertainty. High grade zones are economically more attractive, but with higher uncertainty in the grade. Placing additinoal drillholes in the high grade zone may be efficient to delineate the important part of the deposit. If the goal is to uniformly reduce global uncertainty and increase the overall knowledge of the deposit, then samples could be collected on a regular grid over the area.

Although only a minor amount of work has been carried out specifically in data spacing and

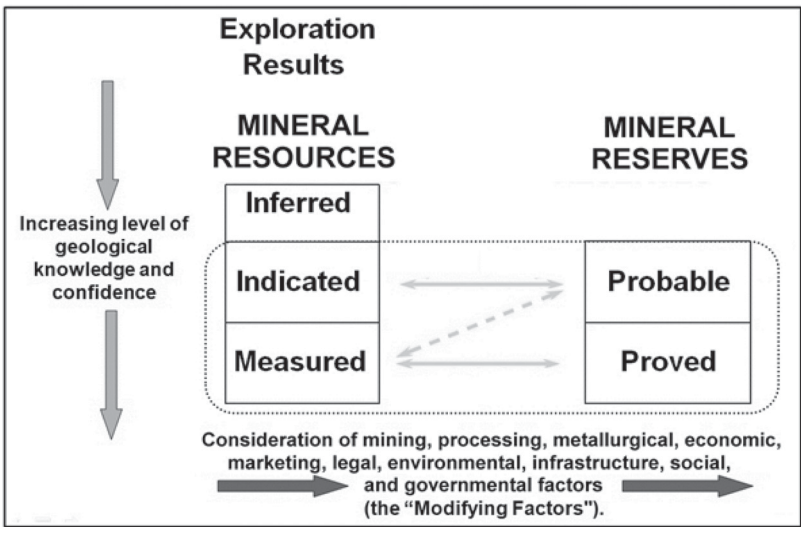

Figure 1.1: The JORC general mineral resources and reserves relationship. From (CRIRSCO, 2013).

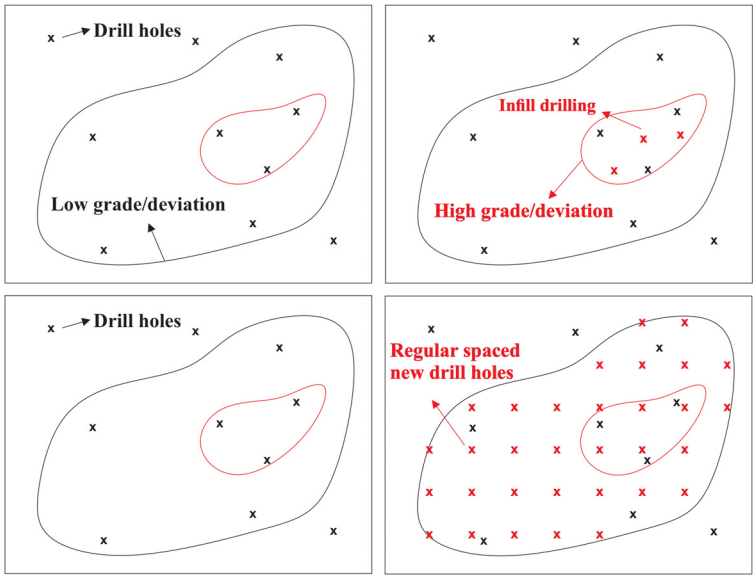

Figure 1.2: Illustrated schema of infill drillhole placement and regular spaced grid.

uncertainty, strong methodologies to relate these two factors have been developed. Firstly, it is important to know that in earth sciences, numerical models are created to reproduce data in its location and estimate properties away from it. Numerical models are the basis for decision making. The geological and grade models of a mine are commonly generated by geostatistical methods (Journel and Huijbregts, 1978). The uncertainty will be reduced with more drilling and other sources of information. This is referred to as the information effect (Rossi and Deutsch, 2014). Although data spacing and uncertainty studies can be done at any stage of mining, it is common to consider them relatively early in the lifecycle of a deposit.

This thesis explores an improved understanding of the relationship between data quantity and uncertainty. This will provide decision support information for drilling and data collection studies.

1.1 Literature Review

Most of the methodologies for infill drilling optimization consider minimizing the kriging variance (KV) due to the fact that KV can be calculated prior drilling, it considers anisotropies, accounts for spatial relationship between locations and is independent of the grade (Silva, 2015; Soltani and Hezarkhani, 2013). The first works on infill drilling optimization considered only 2D cases. The simplest scenario is the placement of a single drillhole at the location of highest KV, a second drillhole is placed after recalculating the KV (Gershon, 1987). Different methods were proposed such as the use of fixed point theory and iterative gradient based techniques (Sch, 1983) and integer programming using the branch and bound procedure (Gershon, 1987).

With the advancement of computing, new algorithms were developed to handle 3D models and optimizing more than one drillhole at a time. The use of a genetic algorithm to minimize the average KV in 2D and 3D cases was proposed by Soltani et al. (2011), whereas Mohammadi et al. (2012) considered the grade of blocks as the weight given to the average kriging variance in a simulated annealing based approach. In both methods

the new drillholes were considered vertical. Soltani and Hezarkhani (2013) proposed the optimization of directional drillholes by simulated annealing, in which the azimuth was fixed and the dip optimized. A more complete treatment to the problem, with more drillholes and no need for fixed azimuth, dip, and location was proposed by Silva (2015).

Geostatistical methodologies such as kriging and simulation are common to assess the spatial distribution of a property of interest including soil and groundwater properties (Webster and Oliver, 2007). McBratney et al. (1981a,b) proposed a method for designing optimal sampling schemes based on minimizing the standard error of a kriged estimate. The maximum error allowed must be set up by the modeller and the sampling density is calculated for a given variogram. The standard error is minimized if sampling is performed on an equilateral triangular grid for isotropic variogram cases. The sampling distances are modified to account for anisotropy, with smaller spacing in the direction of least continuity.

The problem of sampling schemes has been explored in groundwater monitoring and hydrogeology (Andricevic, 1990; Carrera et al., 1984; Criminisi et al., 1997; Loaiciga, 1989; Meyer and Brill, 1988; Rouhani and Hall, 1988; Storck et al., 1997; Zhang et al., 2005). Bueso et al. (1998) consider the spatial sampling design problem to find an optimal number of piezometric sites by sequentially adding to or deleting from the preexisting sampling network. In this work, the conditional entropy (the average amount of information of the mixing of variables) is calculated for the various number of deleted sites. The plot of number of sites against conditional entropy is used to understand the rate of information gain with respect to the number of deleted sites. They suggest an optimum solution is the one that minimizes the conditional entropy and coincides with the maximum rate of information. Bueso et al. (1999) extends the entropy-based approach to a multivariate framework. The objective function to be minimized is defined as a linear combination of the information on the variables an/or the location of data.

Drillhole spacing is commonly used as a criteria for mineral resource and reserve classification (Silva and Boisvert, 2014a). The magnitute of the estimate error are based

on the information from sample points, therefore all statements about the quality and quantity of mineral resources are dependent, among other factors, on the nature and quantity of sample information (Diehl and David, 1982). Although geometric criteria are considered neat and easy-to-understand parameter for classification, it is recommended to also consider a probabilistic analysis. The application of geometric thresholds depends on a number of factors such as local geology, comparison to similar deposits, uncertainty assessment, governmental guidelines/regulamentations and expert judgment (Bertoli et al., 2013; Leuangthong et al., 2006).

Important works on sampling schemes and data collection in mining have been done since the advances of geostatistics and computer sciences (Boucher et al., 2004; Deutsch and Beardow, 1999; Englund and Heravi, 1992; Froidevaux, 1982; Koppe et al., 2011; Pilger et al., 2001; Rojas and Cáceres, 2011; Wilde, 2010; Wilde and Deutsch, 2009). These publications are important for this thesis and the methodology herein presented, they are reviewed in more detail below.

Froidevaux (1982) shows how the precision of estimate (and, consequently, ore reserve classification) is affected by drillhole spacing. The precision measures how spread (distributed) the estimates are from the mean. The author uses geostatistics techniques to simulate a deposit with different number of drillholes and calculate the precision for each drilling pattern. The precision of the simulated estimates is plotted against the number of drillholes, see Figure 1.3. The actual precisions for three different drillhole spacing (shown as centers in the plot), show good agreement with the simulated estimates. Considerable changes in the precision are observed close to the inflexion point whereas marginal gains in precision are obtained away from it, even with large number of drillholes.

Englund and Heravi (1992) discuss the use of conditional simulation to optimize the sampling in contaminated soils. Sequential Gaussian simulation (SGS) is used to generate a representative true spatial distribution of the contaminant over the site. Given an objective function defined by the sum of sampling cost plus remediation costs plus cost of

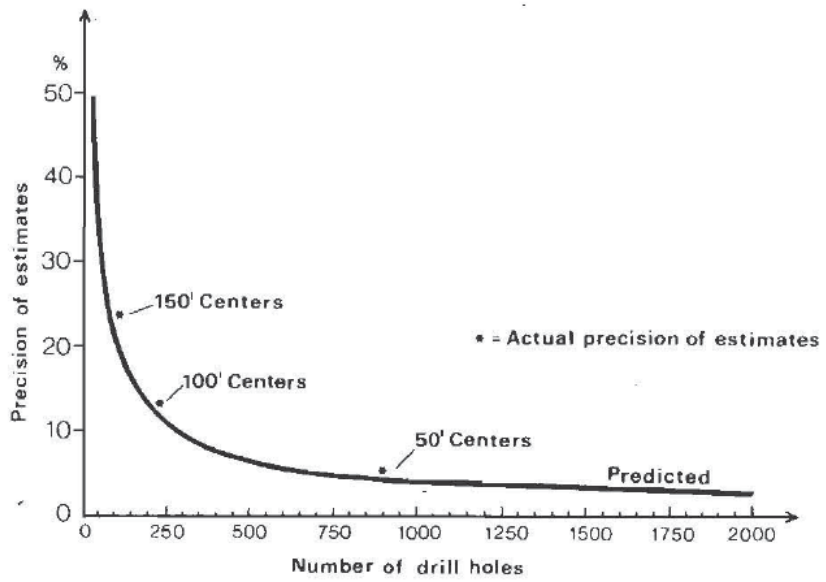

Figure 1.3: Precision of estimates versus number of drillholes. The actual precision for three different drillhole spacing are plotted as dots. From (Froidevaux, 1982).

residual contamination, the goal is to estimate the number of samples that would result in the lowest total cost. The simulated model is block averaged to a remediation unit (RU) and concentration and cost are calculated in the RU scale. For the same number of samples, the model is sampled in different random locations and the total cost is calculated. This process is repeated for a set of different number of samples. The optimum number of samples is the one related to the lowest total cost, as seen in the cross plot of Figure 1.4.

Deutsch and Beardow (1999) propose a methodology for optimal drillhole spacing in oil sands while assessing uncertainty in bitumen/fines predictions and reducing total cost. The total cost is a function of drilling cost and uncertainty cost. Block kriging or stochastic simulation can be used to assess uncertainty. A single measure of uncertainty (the average of the local variances over all locations) is retained for each drillhole spacing/density configuration. Moreover, the use of volume variance relations allows to establish an absolute measure of uncertainty at any scale. To simplify the application of the methodology, two assumptions are made (1) the cost of uncertainty in bitumen/fines is

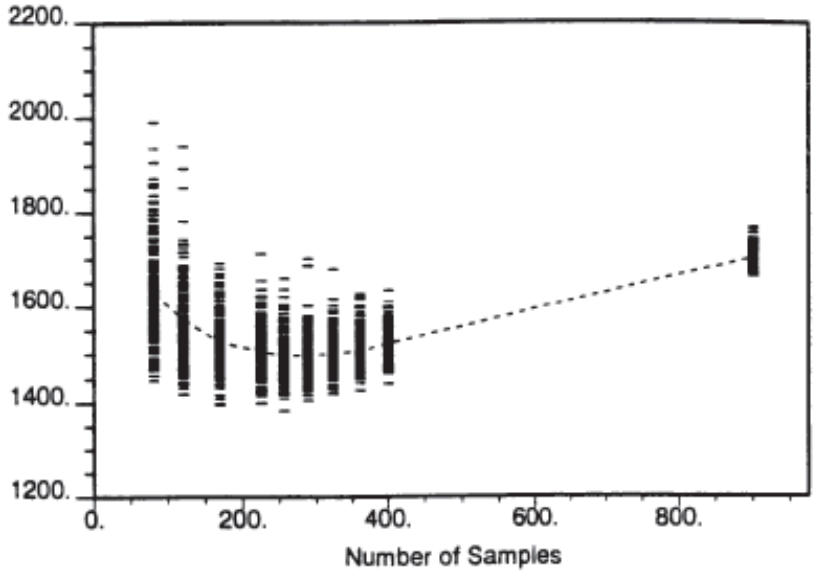

Figure 1.4: Total remediation cost (Y-axis) versus number of samples (X-axis). From (Englund and Heravi, 1992).

a linear function of uncertainty (indeed, it is a combination of many factors) and (2) drilling cost increases linearly with drillhole density. Moreover, among many factors that must be considered to satisfactorily arrive at the optimal spacing, the total cost is the single factor to choose the optimal density, see Figure 1.5.

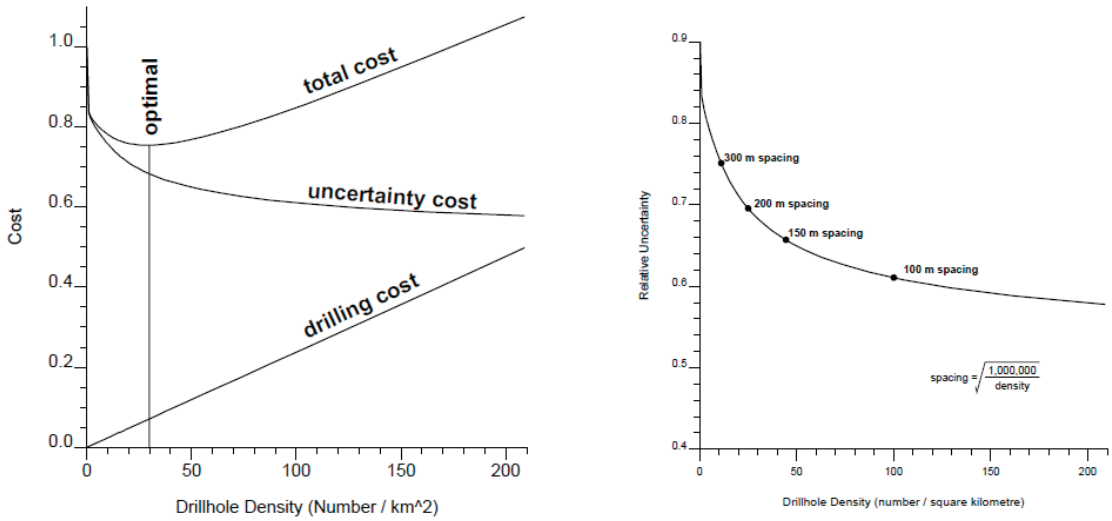

Figure 1.5: Optimal drillhole density against cost (left) and relative uncertainty (right). From (Deutsch and Beardow, 1999).

Boucher et al. (2004) propose a method for infill drilling assessment and optimization that is summarized in five steps. One realization of the attribute under study is generated by conditional simulation and treated as the "actual" deposit. This realization is sampled with different drilling schemes. For each scheme a set of realizations is run to simulate the attribute conditional to the data from the drilling schemes. The realizations are block averaged to a selective mining unit (SMU) scale and grade control and block classification are performed for each sampling scheme. The classification is then compared to the actual deposit (the first realization) using economic indicators such as the profit per tonne mined and profit per tonne milled. The optimal drilling scheme is the one that maximizes the total profit.

Wilde (2010); Wilde and Deutsch (2009) review the influence of factors such as local grade level, spatial variability of the grades, economic thresholds, scale, parameter uncertainty and data spacing/density on uncertainty and propose a methodology for determining uncertainty for different sampling schemes. The methodology is closely related to those proposed by Deutsch and Beardow (1999) and Boucher et al. (2004), although no effort is made to define an acceptable level of uncertainty, and no cost or objective function is minimized. This robust methodology evaluates the relationship between uncertainty and data spacing and can be adapted/modified to consider technical and economic factors defined by any modeller. The proposed methodology is illustrated in Figure 1.6. Realizations of the true distribution of a variable of interested is generated by SGS. These realizations are then sampled in any desired spacing. A new set of realizations is created now conditional to the new configuration of the samples. The fine scale grid of simulation is averaged up to a larger scale and uncertainty is summarized for each data spacing. The retained information is plotted in order to better support decisions.

Rojas and Cáceres (2011) use conditional simulation to evaluate the risk related to drillhole spacing and support an infill drilling campaign to improve resource classification. Conditional simulation is used to generate scenarios of the unknown reality of the deposit (geology and grade). The "reality" is sampled with four different drillhole spacing and the resource is estimated for each drilling scheme. The difference or errors in

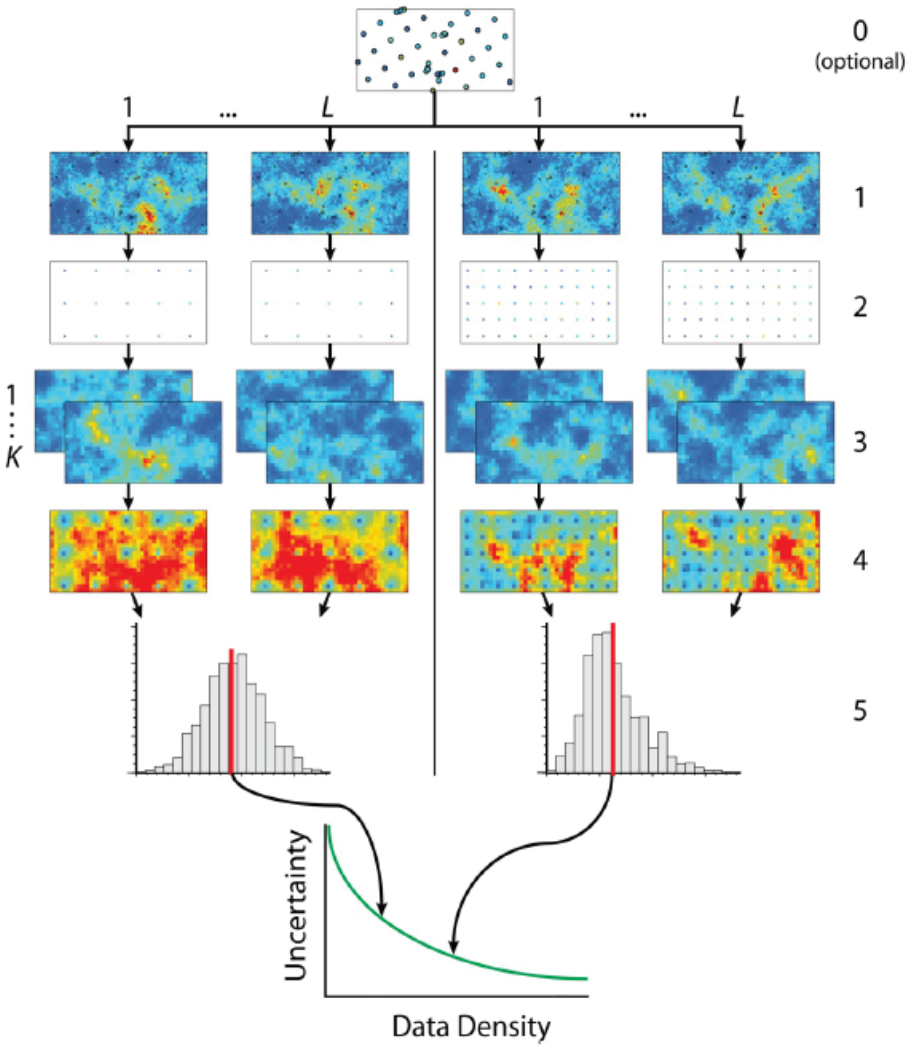

Figure 1.6: Illustrated schema of the proposed methodology to assess uncertainty and data spacing proposed by (Wilde, 2010). L realizations of the truth are generated by simulation and sample at different drilling spacing. K simulation realizations are generated conditional to the drilling schema. The average block uncertainty is plotted against uncertainty.

the estimated mean, metal content and tonnage are used to assist decision making. An important point in this study was the fact that a production volume was considered when calculating the estimation errors. The simulated models were "mining reconciliated" with the resource model (the reality) to consider an arbitrary production volume, allowing to define the optimal drillhole spacing to the desired mineral resource categories.

Koppe et al. (2011) compare the impact on uncertainty for two infill drilling schemes (1) regular (or quasi regular) grid and (2) placement of the new drilling in high variability location. The Net Present Value (NPV) of the deposit is used as transfer function of uncertainty. To calculate the NPV, the authors considered only the block grades and used mining and processing costs of copper, as well as the metal price as reference. The goal is to know which scheme would maximize the NPV, given a fixed number of additional samples to be added. The methodology used can be summarized in few steps. Several scenarios (realizations) of the attribute of interest are generated by geostatistical simulation with the initial drilling configuration by Turning Bands (Matheron, 1973). The set of simulated scenarios is used to map locations of high uncertainty. A simulated scenario is chosen to be sampled with the two drilling schemes. A new set of realizations is run conditional to the new data and later averaged up to a block scale for NPV calculation. As the real scenario is known, a comparison between real and simulated scenarios can be made. Post-processing of the information with analysis of the cost of the new drillholes and reduction of uncertainty support the choice of the optimal scheme. Other factors such as number of initial data, histogram, type of mineralization, redundancy of data (dependent of the variogram), sampling distance, number of realizations and cost of drilling affect uncertainty assessment, hence the NPV. The authors advise that the choice for the best drilling scheme should be made on a case-by-case basis.

The methodology proposed here adapts concepts from these approaches. The aim is not to define optimal infill drillhole location but to provide a practical and comprehensive workflow to understand the relationship between uncertainty and data spacing. This work investigates some factors affecting local uncertainty. The methodology considers the scale in uncertainty assessment. Proper interpretation and use of this methodology would help

companies choose the sampling spacing to achieve a desired level of uncertainty.

1.2 Concepts in Data Spacing and Uncertainty

For completeness, a discussion of the basic concepts of data spacing and uncertainty is provided.

1.2.1 Modeling Uncertainty

Uncertainty is due to heterogeneity at all scales and relatively widely spaced drilling (Caers, 2011; Pyrcz and Deutsch, 2014). The geological processes that created a deposit are too complex to be fully explained with relatively sparse data.

Geostatistics provides tools to build numerical models to assess uncertainty by kriging or simulation (Deutsch and Beardow, 1999). Assessing uncertainty with the kriging variance is limited due to the fact that its calculation accounts only for the sample locations and not the value of the samples (Goovaerts, 1997). Conditional simulation is used to generate multiple realizations of the spatial distribution of an attribute considering the variogram and conditioning data (Journel, 1974). Each realization is equally-probable and reflects a possible scenario for the spatial distribution of that attribute. The set of multiple realizations results in a distribution of predicted system response values, reflecting the uncertainty (Gotway and Rutherford, 1994).

Different measures of uncertainty are reviewed in Chapter 3. In a simulation context, uncertainty is assessed in a very high resolution (data scale) then averaged up to a larger scale that is relevant for decision making, mine planning or operation. Consider two measures of uncertainty calculated for grade in an arbitrary relevant volume, e.g. a monthly production volume. The standard deviation of a production volume is calculated from the simulated grades. The probability of the grade to be within a percentage of the mean, for example, within ±15% of the predicted grade for at least 80% of the monthly production volumes. If a set of 100 realizations are generated, a monthly production

volume would be considered acceptable if the number of realizations that the simulated grade fell within ±15% of the mean grade at least 80 times.

1.2.2 Geometric Measurements of Data Availability

Drillhole spacing (DHS), drillhole density (DHD) and distance to the nearest drillhole (DND) are the most common spatial measures of data availability.

1.2.2.1 Data Spacing

Data spacing is the average distance between adjacent data within an area. A relatively large area with few data is reported as large spacing. When drillholes are not uniformly distributed over an area, then data spacing must be calculated locally.

Mory and Deutsch (2006) proposed the calculation of DHS for vertical and non-vertical drills accounting for the anisotropy in grade continuity in 2D and 3D models. In two-dimensional cases, data spacing is defined as the average spacing of the drilling grid in each primary direction, as shown in Figure 1.7. The DHD at a location (**u**) is given by the number of samples found inside the rectangular or elliptical area ($n_s(\mathbf{u})$) divided by the area (A) times the reference area (10.000 squared meters for one hectare), as shown in Equation 1.1. The area (A) depends on the search shape and anisotropy and is calculated for elliptical (Equations 1.2) and rectangular (Equation 1.3) shapes. The average local spacing is then calculated by Equation 1.4. In three-dimensional cases where drillholes are not vertical nor aligned in the same direction, the calculations involve obtaining the nominal number of vertical parallel drillholes that would contain the same number of samples as the real configuration in the same volume. The volume height (Vh), see Equation 1.5, is used to calculate the nominal number of drillholes (n_{dh}) in Equation 1.6. Drillhole density and spacing are then calculated replacing $n_s(\mathbf{u})$ by n_{dh} in Equation 1.1.

$$DHD(\mathbf{u}) = \frac{n_s(\mathbf{u})}{A} \times 10000 \tag{1.1}$$

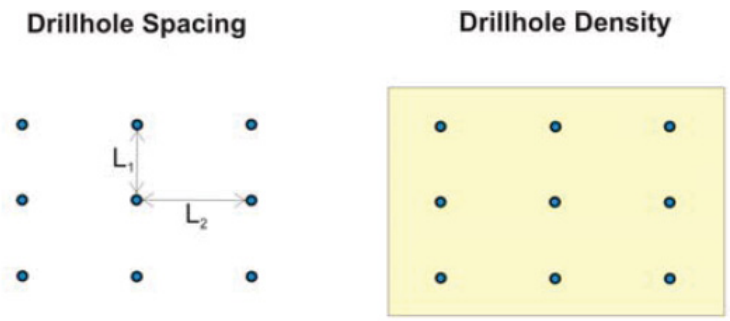

Figure 1.7: DHS and DHD parameters for calculation from Mory and Deutsch (2006). L_1 and L_2 are the drillhole spacing in horizontal direction. The yellow square is the area or volume for calculations.

$$A = \pi \times (sh_{max} \times a_h)^2 \tag{1.2}$$

$$A = (2 \times sh_{max} \times a_h)^2 \tag{1.3}$$

where:

sh_{max} - maximum horizontal search radius;
a_h - horizontal anisotropy;

$$DHS(\mathbf{u}) = \frac{L_1 + L_2}{2} = \sqrt{\frac{10000}{DHD(\mathbf{u})}} \tag{1.4}$$

$$Vh = \frac{2 \times sh_{max}}{a_v} \tag{1.5}$$

$$n_{dh} = \frac{n_s(\mathbf{u}) \times c}{Vh} \tag{1.6}$$

where:

c - composite length;
a_v - vertical anisotropy;

Naus (2008) proposed the use of Delaunay triangulation for calculation of point spacing in regular and irregular schemes, see Figure 1.8. The distance of every edge connecting one point to a neighbor is calculated, the spacing in a point is then given by the average of the edges lengths connected to the point.

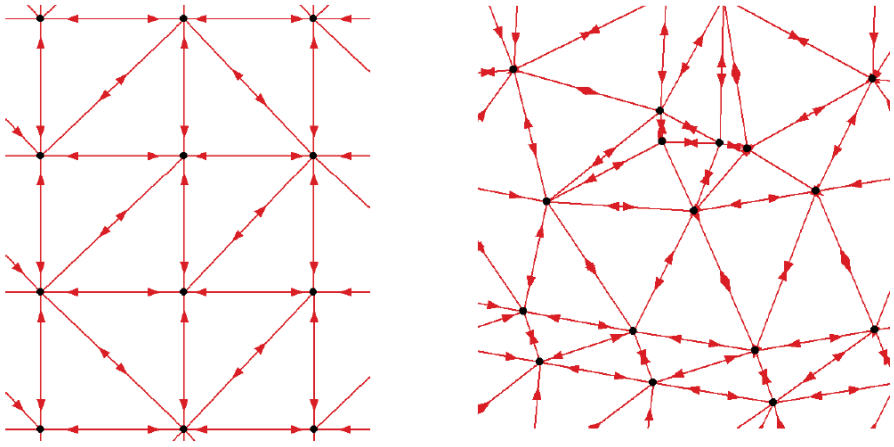

Figure 1.8: DHS calculation using Delaunay triangulation, from Naus (2008). The black dots are the data location, the red lines define the triangles.

Wilde (2010) proposed the calculation of data spacing for vertical and non-vertical drills (Figure 1.9) by Equations 1.7 and 1.8 respectively. The calculation of data spacing in three dimensions is reduced to two dimensions if drillholes are vertical, otherwise Equation 1.8 must be used. The calculation of data spacing requires either a constant volume $V(\mathbf{u})$ or number of samples $n_v(\mathbf{u})$ to search for. Data spacing will be noisy for very small volumes or too few samples and it will be over-smooth and locally imprecise for large volumes or too many samples (Leuangthong et al., 2006; Wilde, 2010).

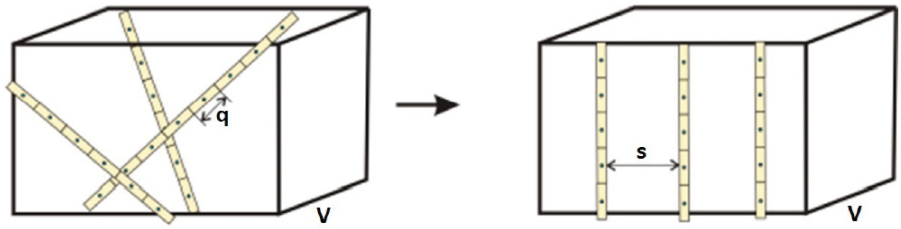

Figure 1.9: Ilustration of data spacing calculation in three dimensions for irregular and regular spacing, from Wilde (2010). V is the volume for calculations, q is the composite length and s the drillhole spacing.

$$s(\mathbf{u}) = \left(\frac{V(\mathbf{u})}{n_v(\mathbf{u})} \right)^{\frac{1}{2}} \tag{1.7}$$

where:

$s(\mathbf{u})$ - data spacing at location **u**;
$V(\mathbf{u})$ - search volume;
$n_v(\mathbf{u})$ - number of samples found within the volume $V(\mathbf{u})$;

$$s(\mathbf{u}) = \left(\frac{V(\mathbf{u})}{q \cdot n_v(\mathbf{u})} \right)^{\frac{1}{2}} \tag{1.8}$$

where:

q - sample spacing along the drillhole;

Silva and Boisvert (2014b) proposed the calculation of DHS using a single datum from each drillhole, reducing the dimension of the problem to a 2D calculation. The vertical tolerance is equal to the block vertical dimension, see Figure 1.10. DHS can be calculated given a search geometry (square or circle) or by a fixed number of data to search closest to the location being considered ($n(\mathbf{u})$). Equation 1.9 is used for a square search whereas Equation 1.10 is used for circular or number of data to search. In order to improve accuracy

and to approximate DHS calculation to equally spaced cases, the maximum distance for a search strategy is considered equal to the edge length of a square $R(\mathbf{u})$. The calculation of $R(\mathbf{u})$ is performed after finding $n(\mathbf{u})$. Equation 1.11 is used when a search geometry or the number of data is the parameter for DHS calculation, whereas Equation 1.12 is used for squared search. A smooth DHS is achieved using multiple parameters and averaging all calculations.

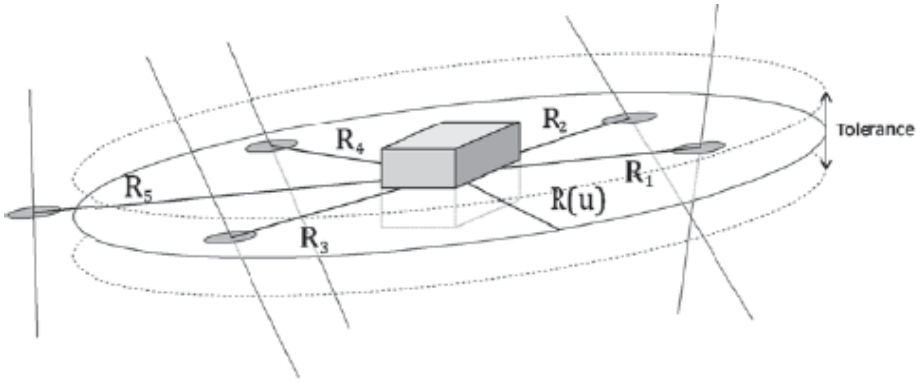

Figure 1.10: Illustration of DHS calculation and search scheme, Silva and Boisvert (2014b).

$$DHS(\mathbf{u}) = \left(\frac{R^2(\mathbf{u})}{n(\mathbf{u})}\right)^{\frac{1}{2}} \tag{1.9}$$

$$DHS(\mathbf{u}) = R(\mathbf{u})\left(\frac{2}{n(\mathbf{u})}\right)^{\frac{1}{2}} \tag{1.10}$$

$$R(\mathbf{u}) = \frac{R_n(\mathbf{u}) + R_{n+1}(\mathbf{u})}{2} \tag{1.11}$$

$$R_n(\mathbf{u}) = \mid X_n - X(\mathbf{u}) \mid + \mid Y_n - Y(\mathbf{u}) \mid \tag{1.12}$$

Data spacing measures the average distance between data and not directly the quantity of

data in a volume, therefore the estimation of cost of drilling is not as clear as it is for data density.

1.2.2.2 Data Density

Data density measures the number of data inside a reference volume. Data density is related to the number of drillholes and provides a direct manner to calculate the cost of drilling. Data density is commonly reported as the amount of data per section or hectare. Data density is high if many samples fall within a small volume and low if a large volume contains few samples.

Naus (2008) proposed the use of Voronoi diagram (Figure 1.11) for calculation of data density. Data density is calculated at any location assign to it the inverse of the polygon area or in terms of data per unit squared.

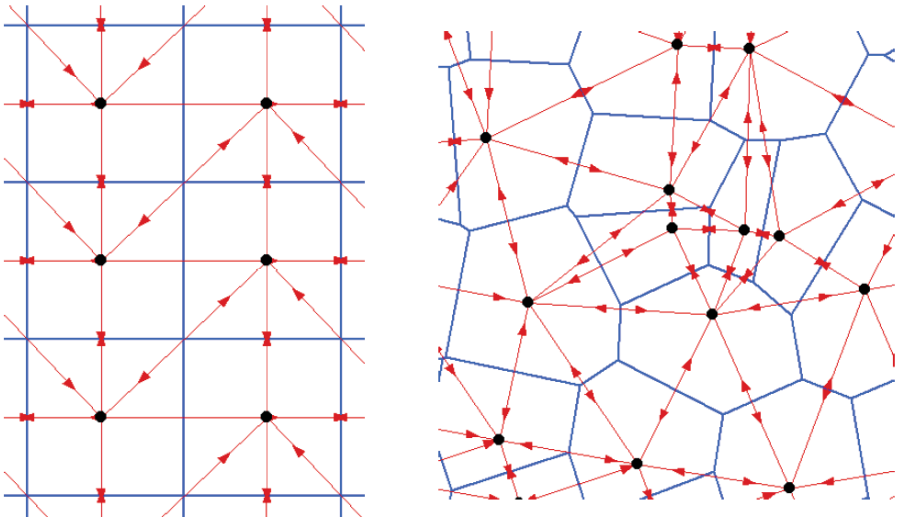

Figure 1.11: DHD calculation using Voronoi diagram, from Naus (2008). Red lines define the triangles (Delaunay), blue lines define the Voronoi polygons.

Wilde (2010) proposed the calculation of data density by Equation 1.13. Its calculation involves the same parameters $V(\mathbf{u})$ and $n_v(\mathbf{u})$ used in data spacing calculation.

$$DHD(\mathbf{u}) = \frac{n_v(\mathbf{u})}{V(\mathbf{u})} \tag{1.13}$$

Data spacing and density are measures dependent on the volume or number of data used in calculations. Distance to the nearest data is not affected by these parameters.

1.2.2.3 Distance to the Nearest Data

Distance to the nearest data (DND) is not as good measure of data availability as data spacing or density since it is highly locally variable and could be small in areas of widely spaced drillholes. Nevertheless, DND provides valuable complementary information to data spacing.

1.2.3 Data and Scale

Data quantity and quality affect the uncertainty and the ultimate goodness of a model. There is uncertainty in data measurements, data location and the measured data values. Different drilling methods have different sampling quality and precision.

The capacity of geostatistical models to assess uncertainty improves with more data, see Figure 1.12. The unknown truth is sampled with different number of drillholes (black dots); the simulated deposit with many drillholes better represents the truth.

Uncertainty is scale dependent and the choice of scale depends on the goals of the modeling (Caers, 2011; Pyrcz and Deutsch, 2014). Uncertainty decreases for larger volumes due the averaging of extreme values (high and low). The dispersion variance provides a measure of how variability changes with scale (Isaaks and Srivastava, 1989; Journel and Huijbregts, 1978; Pyrcz and Deutsch, 2014).

Uncertainty at the data scale is of little relevance for mine planning and operation. High resolution geostatistical models needs to be averaged-up to a relevant scale. The SMU is the smallest volume that relevant decisions are made in mine decision making. The use of

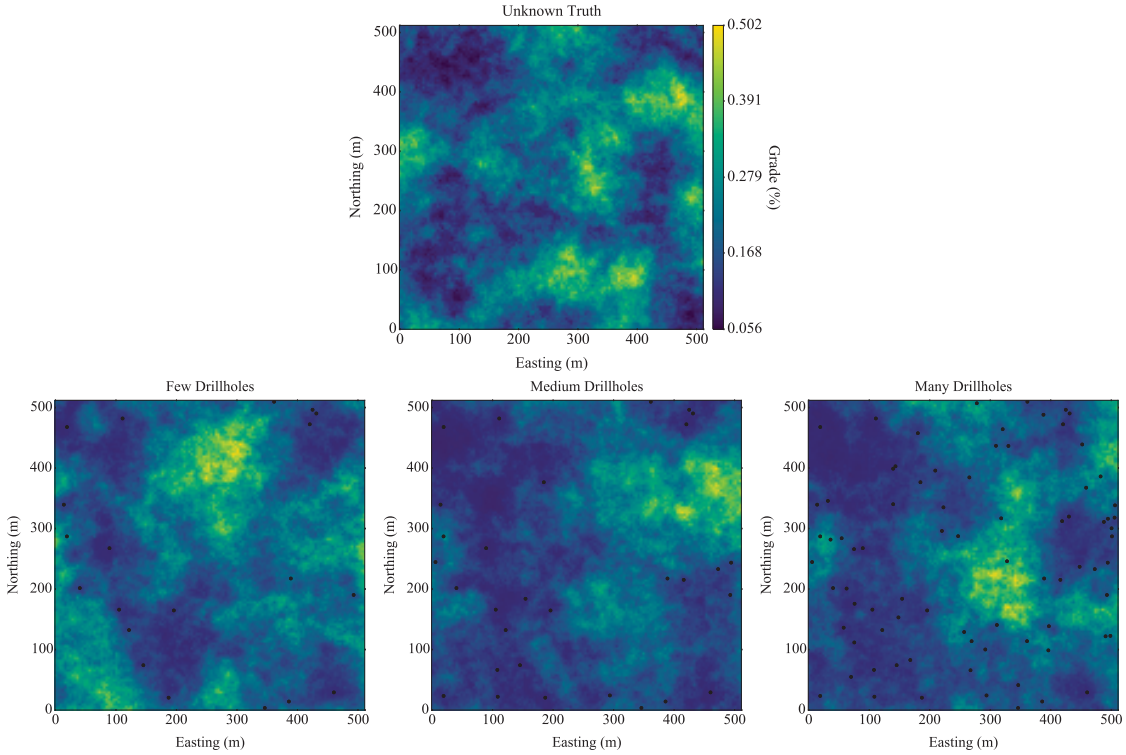

Figure 1.12: The capacity of a model to represent the unknown truth and data quantity shown at plan views. The simulated unknown truth is shown at the top. Three simulated models generated with different number of samples (black dots) from the truth are shown at the bottom.

a SMU scale model is due to the fact that ore tons and grade estimates in that scale will be similar to those found at the time of mining (Daniels, 2015). Although uncertainty in the SMU scale may be important in daily operations, drilling campaign and long term planning take into account large production volumes (a collection of SMUs). It is desirable to assess uncertainty in monthly, quarterly and yearly production volumes, for instance.

In this aspect, the production scale should be considered when assessing uncertainty. Three different manners of incorporating the scale into uncertainty assessment are discussed below.

1.2.3.1 Non-zero Cut-off and Fixed Volume

Given a model with ore and waste SMUs, the resources and uncertainty are assessed at the production volume scale. This is a block-averaging with no overlapping blocks. Figure 1.13 shows 9 production volumes averaging 36 SMUs each in a total of 324 SMUs over the grid. When calculating the resources within a production volume, the calculations are simple and straightforward, the expected values are calculated considering all SMUs within the production volume over all realizations of the grade.

The uncertainty in economic variables such as tonnes of ore, tonnes of waste, expected grade and quantity of metal inside a production volume is associated to that large scale and not to a specific SMU. All SMUs are reported with the same uncertainty. Consider Figure 1.13. It is expected that SMU 1 will have higher uncertainty as it is located in a transition zone (between ore and waste), than SMU 2. All realizations of SMU 2 are likely waste.

This approach may result in an overestimation of the uncertainty in some variables. Even inside a production volume that is all ore or all waste, SMUs may have different uncertainty levels. SMUs 3 and 4 are inside the same volume but SMU 4 is closer to SMU 5. A more reliable uncertainty measure in any SMU is determined considering its local neighborhood.

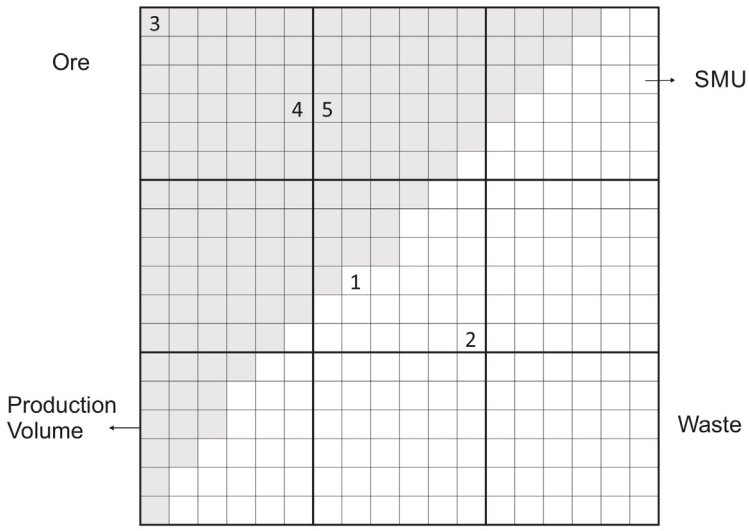

Figure 1.13: This is a schematic plan view of a bench. The small squares represent SMUs of 10-15 m on a side. The large squares represent the collection of SMUs (the production volume) that will be mined in a representative period of time. The gray SMUs are ore, the white SMUs are waste.

1.2.3.2 Zero Cut-off and Fixed Volume

Another way to assess uncertainty in a SMU is applying the concept of zero cut-off and fixed volume centered at the SMU. Calculations are performed centering the production volume at each SMU and assessing uncertainty based on a "moving window" with size equal to the production volume. If a cut-off is applied to the grade, and uncertainty is based on grade or tonnes of ore, then low grade areas are penalized. If uncertainty is based on quantity of metal then information on grade/tonnes is lost.

A zero cut-off makes the analysis of uncertainty less based on economics. The uncertainty is assessed considering the same time period in which each volume is the same. Figure 1.14 represents a model in which SMUs are either high or low grade. Consider the assessment of uncertainty in the SMUs based on a production volume moving window. Uncertainty in each SMU is now based on the adjacent blocks. The uncertainty in SMU 4 depends more on SMU 5, whereas in the previous case (Figure 1.13), uncertainty in SMU 4 was dependent

on SMU 3, a further block. Uncertainty in SMU 2 is no longer dependent on the high grade zone for that volume, the influence area around SMU 2 is all low grade.

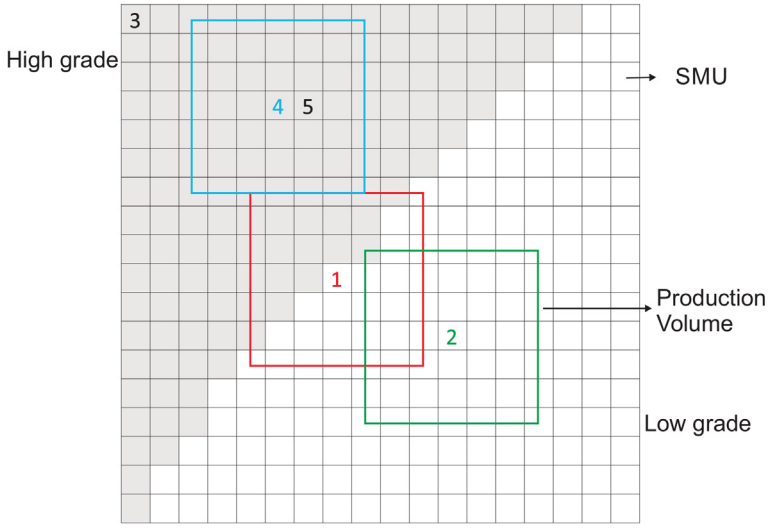

Figure 1.14: This schema is similar to the Figure 1.13. The gray SMUs are high grade, the white SMUs are low grade. The production volume is now centered at the SMU. Three different production volumes are shown in red, blue and green. They all have the same volume.

1.2.3.3 Non-zero Cut-off and Varying volume

A varying volume to match an economic criterion only makes sense if a cut-off is applied and the production volume window can expand or contract. Consider a monthly production formed by 25 ore SMUs, the blue square shown in Figure 1.15. Uncertainty in SMU 4 would be assessed by all 25 adjacent ore SMUs, matching the fixed monthly tonnes of ore. The production scale is now variant and may have to expand to match this production criterion. The varying production volume centered in SMU 1 has to expand to get the same 25 ore SMUs in order to have uncertainty assessed with the same economic reference scale as in SMU 4.

It is not clear how to expand this volume to get the target production. Expand in one

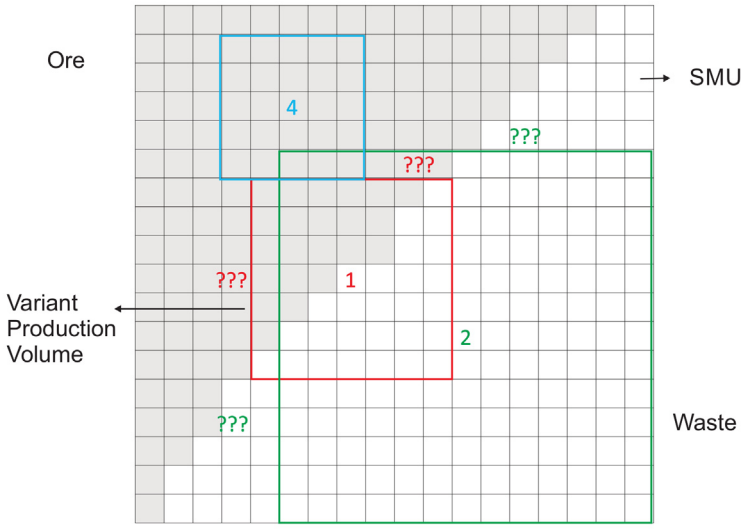

Figure 1.15: This schema is similar to the Figure 1.13. The production volume is now made variant to match an economic criterion, it can expand. For this reason, the production volumes are shown in three different sizes, the blue, red and green squares.

direction may get the same tonnes of ore than in another direction, but not the same quantity of metal, dependent on the grade. Moreover, when expanding the window on a regular grid, the tonnes of ore could match exactly the monthly reference production scale only if fraction of blocks are considered. An entire "row" or "column" of SMUs could result in a surplus of ore. Furthermore, this approach is awkward, since different realizations will have different bounds and different volumes, making results hard to understand and be interpreted. Uncertainty in some SMUs as SMU 2 would not be reliable, as the volume considered is too large compared to the SMU volume.

1.3 Methodology

The methodology herein proposed to determine uncertainty and data spacing is discussed and applied to an example.

Multiple scenarios of the unknown geology and grade distribution are generated by simulation. These models of uncertainty reproduce the spatial variability (variogram), and

the conditioning data, and histogram through a set of realizations. These realizations, built at a very high resolution (point scale) are block averaged to a SMU scale. Uncertainty is assessed in a production volume scale using one of the approaches of Section 1.2.3. Measures of uncertainty are calculated processing all realizations.

Uncertainty versus data spacing curves can be plotted to understand the influence of data spacing on local variability. A deeper study would consider additional explanatory factors. These factors could include local mean, local variance and entropy that explain more of the local variability. Entropy is the explanatory factor associated to the geology and mixing of rock types. All these factors are calculated relative to a production volume.

A set of plots are created to better explore and understand uncertainty and the explanatory factors. If multiple geological domains are considered, then uncertainty versus DHS curves are plotted per domain and analyzed separately. It seems reasonable to zone the deposit according to economic and technical factors, then apply all analysis to target drillhole spacing.

The methodology is summarized as follows:

1. Model uncertainty in categorical and continuous attributes by simulation;
2. Average the high resolution model from simulation to a SMU scale;
3. Assess uncertainty at the production scale;
4. Model the explanatory factors and understand their influence on local variability;
5. Plot uncertainty versus DHS curves;
6. If necessary, consider re-sampling the model with artificial drilling for more information;
7. Zoning the deposit and target DHS;
8. Decide what is DHD/DHS that leads to a desired level of uncertainty;

The following example demonstrates the capacity of the methodology to support decisions regarding DHS and uncertainty.

1.3.1 Example

The goal is to define data spacing necessary to classify SMUs as measured and indicated mineral resources based on different production scales. A complete case study is detailed in Chapters 4 and 5.

1.3.1.1 Modeling the Truth

The data used in this case study was generated by sequential indicator simulation (SIS) with a synthetic trend for four rock types (Goovaerts, 1997; Journel, 1983). SGS (Deutsch and Journel, 1998; Goovaerts, 1997; Isaaks, 1990) is used to simulate the grade within the four domains with a different variogram for each. These two models are then sampled to generate a set of drillholes, they are used as the data file in this example. The 140 drillholes and their spatial location are shown in Figure 1.16.

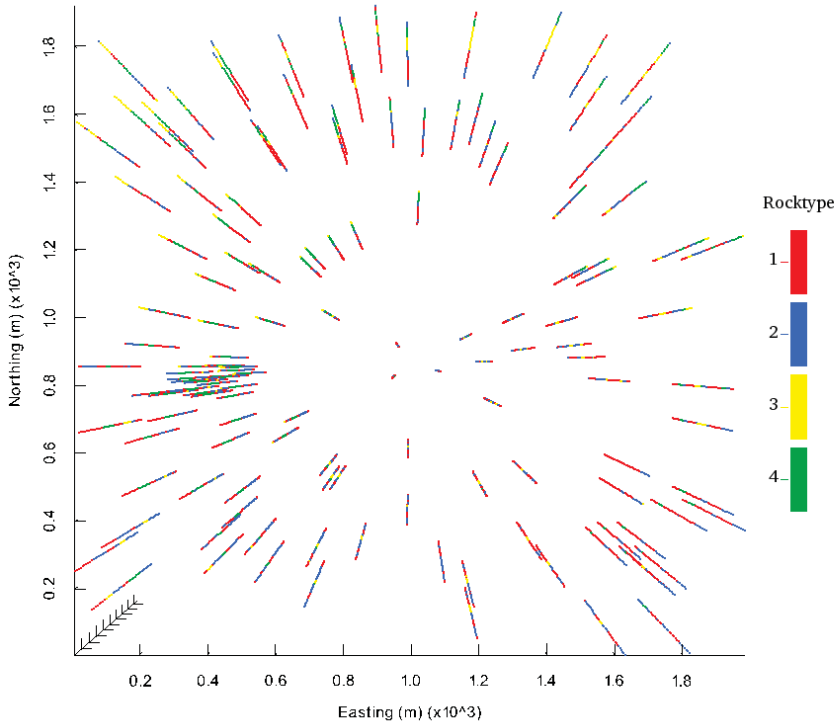

Figure 1.16: Location of drillholes projected onto a plan view and colored by rock type.

One hundred realizations of the geological model were generated by SIS. The P50 realization is shown in Figure 1.17 for RT 1 and RT 3. RT 1 has the highest proportion. Finally, 100 realizations of the grade were generated for each rock type.

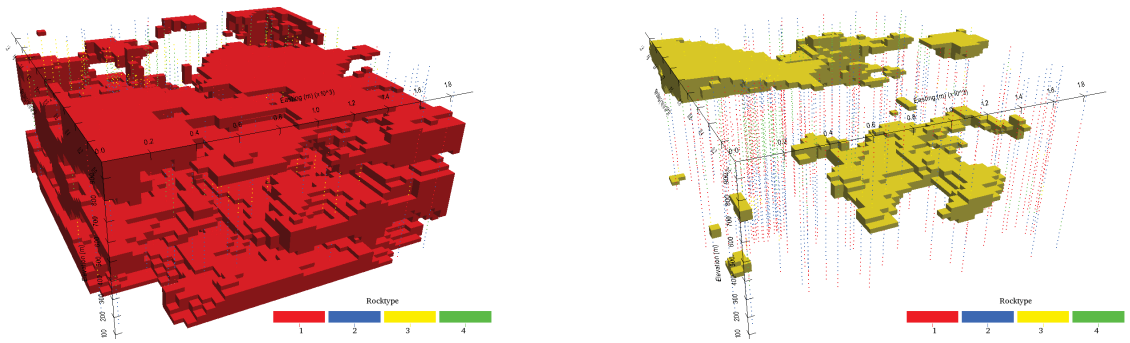

Figure 1.17: Oblique view of the P50 from categorical simulation modelling of RTs 1 and 3. The drillholes are also shown.

1.3.1.2 Uncertainty Assessment in a Production Scale

Four simulated nodes are average up to compose a SMU. Two different production volumes are considered: the weekly and monthly. Uncertainty assessment is done using a zero cut-off and fixed volume approach.

Processing the realizations from the simulation allows the calculation of measures of uncertainty. In this example, the probability of the grade to be within 15% of the estimated mean is the measure of uncertainty. The influence of scale in uncertainty assessment can be seen in Figure 1.18. The measure of uncertainty is plotted for the two production volumes for the same slice of the deposit. A lower uncertainty (higher precision) is associated to the larger volumes.

1.3.1.3 Explanatory Factors

Drillhole spacing is the first explanatory factor to be analyzed. The DHS and DHD over the deposit, as well as their distributions are shown in Figure 1.19.

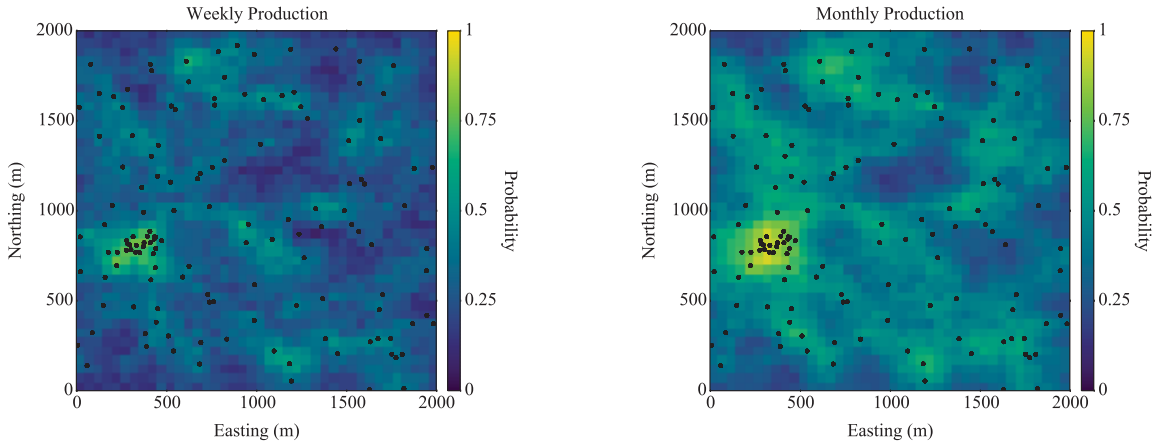

Figure 1.18: Plan view of the precision (15%) of the simulated grade for weekly and monthly production volumes in a slice of the deposit. The dots represent the drillholes.

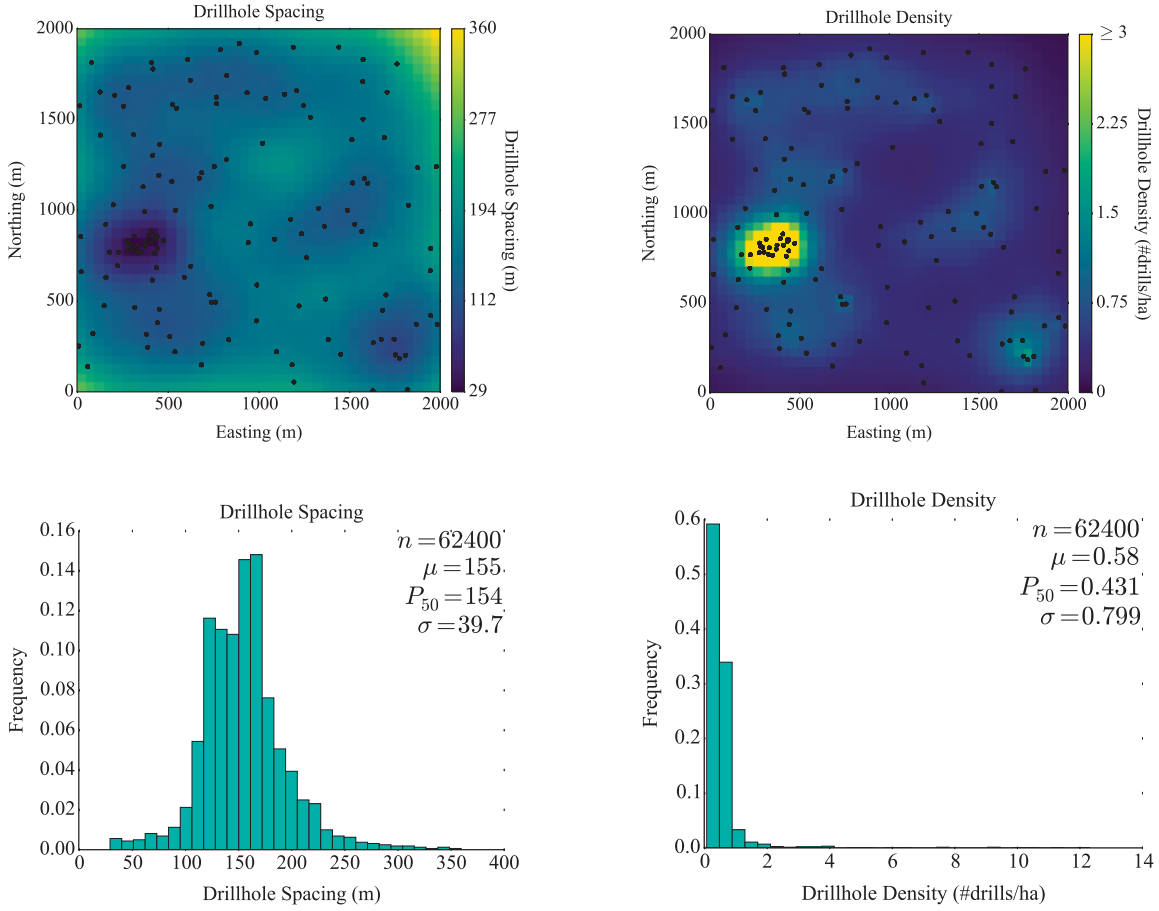

Figure 1.19: Plan view of the DHS and DHS over the deposit (at the top) and their distributions (at the bottom).

Conditional local mean and local standard deviation are calculated considering all SMUs inside a production volume and their simulated grade over all realizations. These two explanatory factors are important to explain the proportional effect, or the dependency of the variance on the mean grade. The conditional local mean is plotted against the standard deviation in Figure 1.20. This confounding factor is more significant for skewed distributions of a variable.

A slice of the deposit showing the conditional local factors is plotted in Figure 1.21 for a weekly production. The local standard deviation is also associated to transition zones. These zones are characterized by high grade deviation, usually located between two domains (different mineralization zones or lithology) or two locations with different grade level inside the same domain.

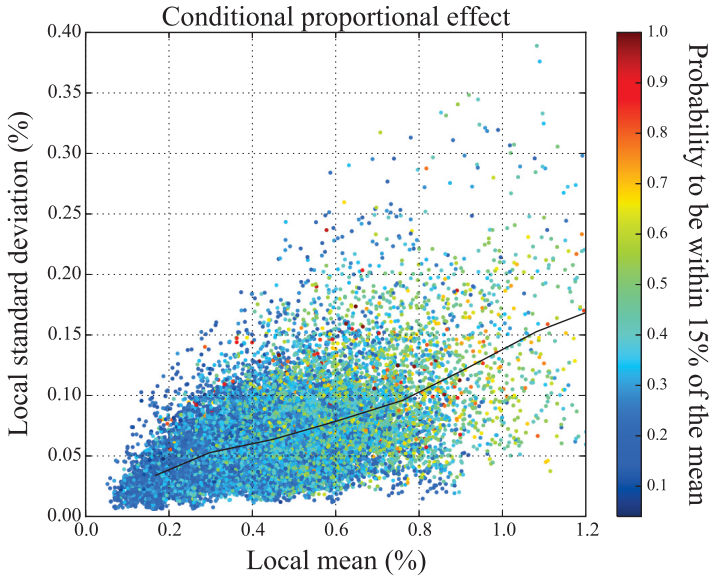

Figure 1.20: Conditional proportional effect with non parametric fitting regression. The dots represent the SMUs, colored by their precision (15%).

Entropy is high in the contact of different domains. If mineralization is associated to a specific rock type, zones with high entropy will affect the grade estimates, hence uncertainty will be high. A slice of the geological model is plotted with the entropy in Figure 1.22

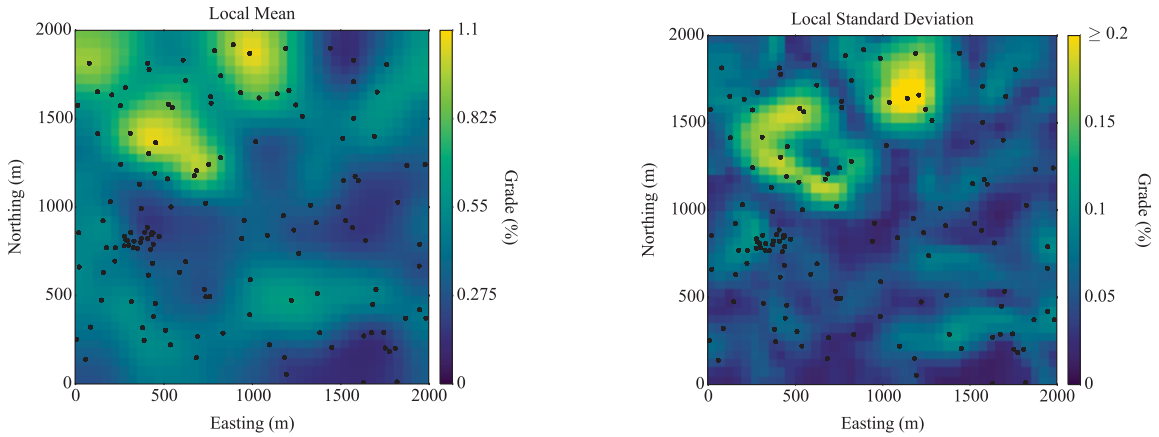

Figure 1.21: Plan view of the local mean and local standard deviation in a slice of the deposit - weekly production scale.

calculated for the weekly production scale.

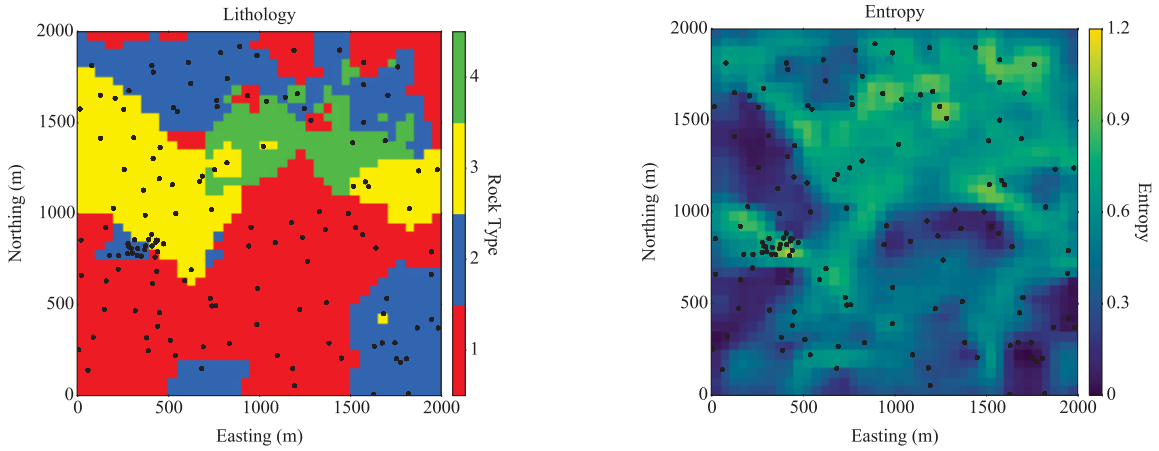

Figure 1.22: Plan view of the geological model and entropy (weekly production scale) in a slice of the deposit.

The explanatory factors and how to model them are better explained in Chapter 3.

1.3.1.4 Explanatory Models

When analyzing the relationship between predictor (factors) and response (uncertainty) variables, scatterplots reveal some information. Plotting the local mean against precision (uncertainty) can help understand the local variability due to that factor, for instance. The quantification of these relationships is given by statistical models that captures the portion of the variance due to each explanatory factor. A more detailed explanation of these models is given in Chapter 3. As example of an explanatory model, Alternating Conditional Expectations (ACE) is used to fit predictor and response variables in a non parametric manner, being able to capture underlying relations that simple statistical regression (linear or quadratic) cannot (Breiman and Friedman, 1985). The variance of the transformed predictor variable in this method can be used to quantify the impact of that factor on the total variability. In Figure 1.23 the drillhole spacing is plotted against its tranformation (on the left plot); and the regression plot of the transformed response versus the sum of the transformed ACE predictor values is shown on the right. Explanatory models must be done for each production volume, since scale changes the impact of each factor on uncertainty.

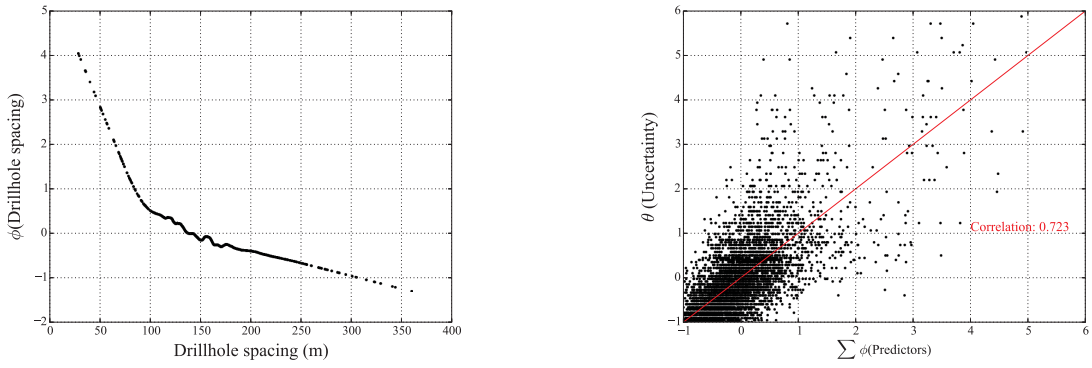

Figure 1.23: Left: Drillhole spacing predictor versus its ACE transform. Right: Regression accuracy for the transformed response (Y-axis) versus the ACE predictor values (X-axis).

1.3.1.5 Uncertainty Versus DHS Curves

When DHS is plotted against uncertainty, without any consideration to stationary domains (the different lithologies), a smooth expected uncertainty curve can be calculated. The expected uncertainty curve is calculated dividing the drillhole spacing into small intervals in which the measure of uncertainty is averaged. Since different domains may be sampled with different spacing and have unique geological features, it is not expected that uncertainty versus DHS curves be the same for all rock types in the model. The uncertainty versus DHS curves for the whole deposit and for RT 1 are shown in Figure 1.24.

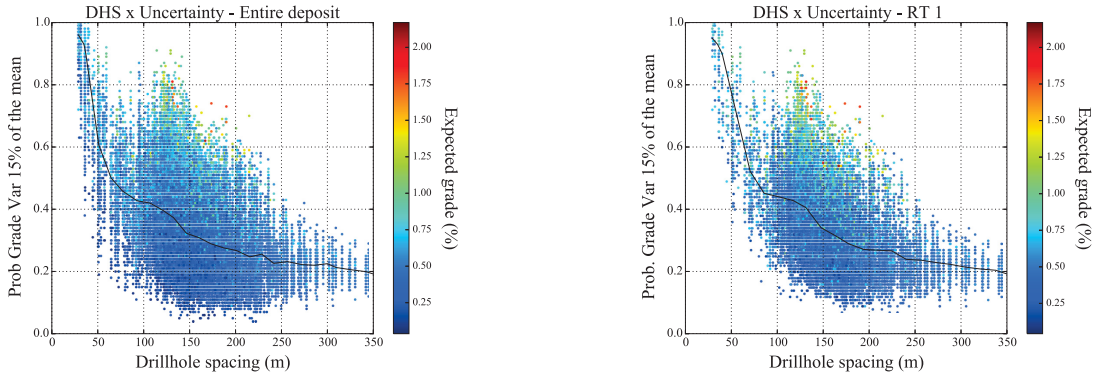

Figure 1.24: Uncertainty versus DHS curves for the entire deposit (left) and for RT 1 (right) based on a weekly production scale. Each dot represents a SMU, colored by its expected grade. The expected uncertainty curve as function of the DHS is shown as the continuous line.

The influence of the scale on uncertainty versus DHS curves is shown in Figure 1.25 for the entire deposit. The uncertainty curve goes up for larger scales (towards more precision). The DHS to achieve a same level of uncertainty is greater for larger production volumes.

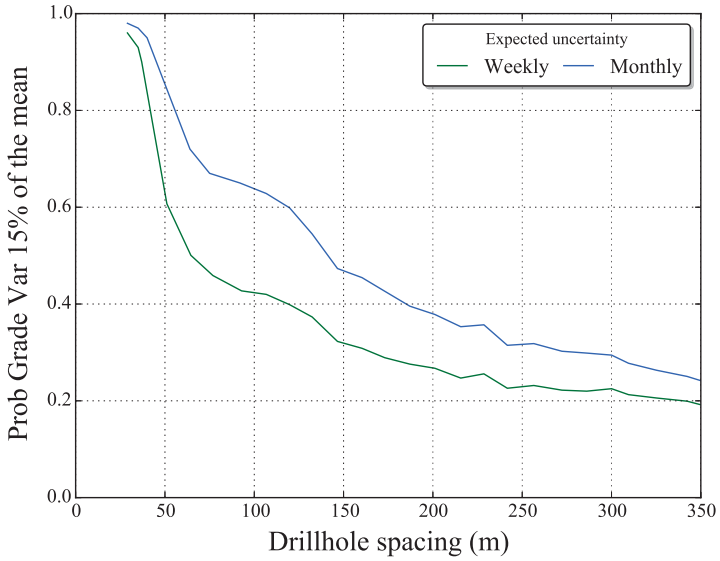

Figure 1.25: Uncertainty versus DHS curves for weekly and monthly production scales for the entire deposit.

1.3.1.6 Zoning

Zoning areas of interested is an alternative when drilling the entire domain or the entire deposit with regularly spaced drillholes is prohibitive (economic and technical factors). Expertise and knowledge of the deposit should be considered when choosing these zones. In this example, the drillholes are vertical. New drillholes will be likely vertical, therefore possible considerations of zoning this deposit may consider uncertainty and economic factors calculated over the vertical extent of the deposit.

The predominant rock type and the expected grade were calculated considering all SMUs vertically adjacent in the grid. Doing such calculations allows a reduction of the 3D grid to a 2D model and the information is summarized in plan view, see Figure 1.26.

Based on the expected grade, two high grade zones were defined, shown in Figure 1.27. Zone 1 has RT 1 as the predominant category. For any analysis, the uncertainty versus DHS curve for this RT is considered. Although RT 1 is the predominant RT in Zone 2, the presence of RT 2 will interfere in the DHS analysis. A higher entropy and local standard

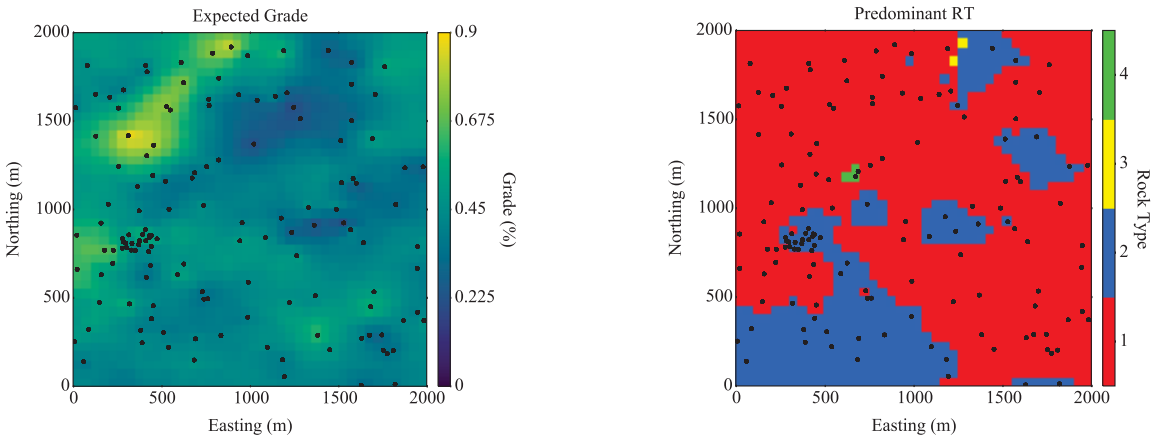

Figure 1.26: Expected grade and predominant category calculated vertically in the grid for a weekly production volume.

deviation in this zone due the contact between the two domains and different grade estimates are expected.

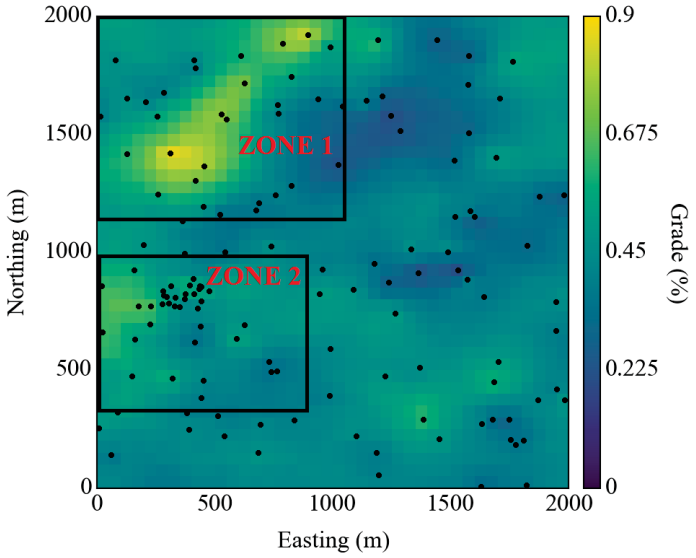

Figure 1.27: Zones of interest based on the expected grade - weekly production volume.

The uncertainty versus DHS spacing curves for weekly and monthly production scales for RT 1 are shown in Figure 1.28. Only confidence intervals greater than 50% and DHS smaller than 200 m were plotted to improve visualization. By the simplest analysis of these curves,

in Zone 1, the DHS required to classify SMUs as measured with a confidence interval of 80% is based on the uncertainty curve calculated within a weekly production scale, whereas the spacing to classify a SMU as indicated is given by the monthly curve. For measured and indicated, the DHS are approximately 50 meters and 65 meters respectively. These DHS can be used in Zone 2 as well, since the predominant RT is also RT 1. However, the influence of the explanatory factors in this zone should be analyzed closer in order to support a proper DHS.

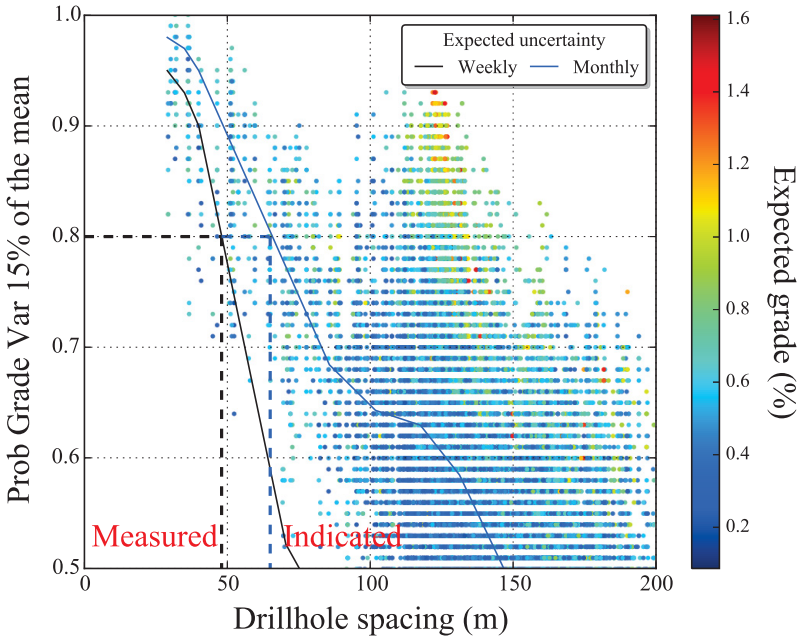

Figure 1.28: DHS required for measured and indicated classifications in zone 1 for a confidence interval of 80%. The SMUs are plotted based on a monthly volume uncertainty and colored by the expected grade.

1.4 Uncertainty Versus Data Spacing Curves

The shape of uncertainty curves depend on the measure of uncertainty plotted, Figure 1.29. When precision is plotted against DHS, the uncertainty curve goes down for large spacing. When the standard deviation of the simulated variable is plotted, the uncertainty curve goes

up, but the concept remains, there is more variability in regions of large DHS.

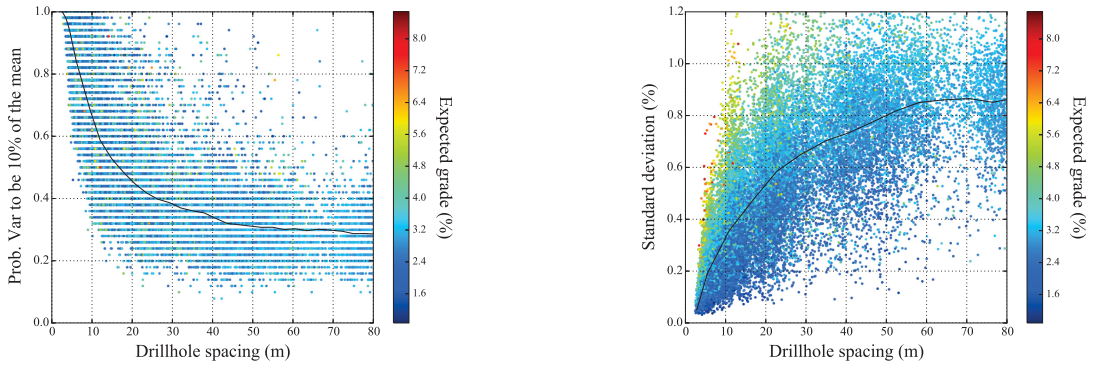

Figure 1.29: Uncertainty and DHS curves for two different measures of uncertainty, precision (on the left) and standard deviation (on the right).

Uncertainty and DHS plots are analyzed in terms of the expected uncertainty curve. Consider the same plot from the example given in the previous section, shown in Figure 1.30. For a given level of uncertainty the DHS is known and vise versa. The expected uncertainty curve is calculated averaging the uncertainty of all SMUs in a small interval of DHS. Therefore, for a given DHS, the uncertainty curve represents the expected value, there will be some SMUs and production volumes less and more certain.

All SMUs that are above the expected uncertainty curve are more certain (when considering precision), whereas all SMUs below it are uncertain, Figure 1.31. A closer look at these SMUs and their spatial location may reveal more about uncertainty in the model.

Increasing the scale of production will decrease uncertainty and may change the shape of the expected uncertainty curve, see Figure 1.32. When different production volumes are considered, for a given level of uncertainty, the DHS associated to it increases with the volume. When increasing scale, it is expected that the uncertainty curve goes up or down in the plots, that depends on the measure of uncertainty.

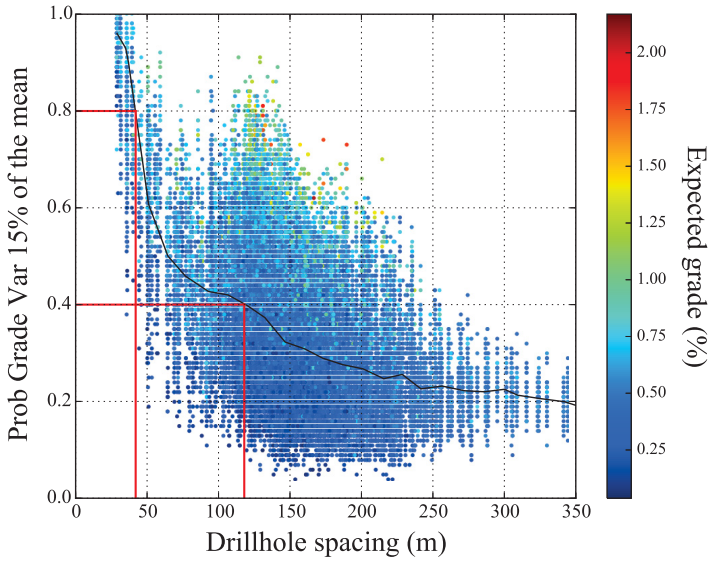

Figure 1.30: Reading information from the plot, for a given uncertainty level the DHS is known, and vice versa.

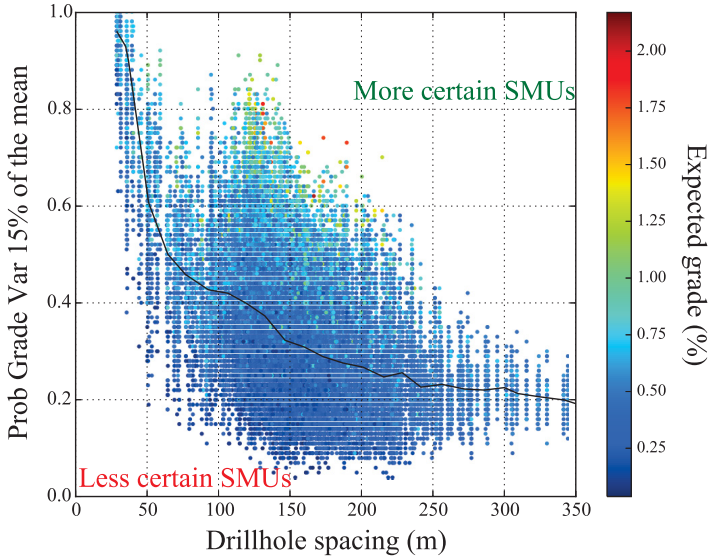

Figure 1.31: Reading information from the plot; more certain SMUs are above the curve, and less certain SMUs are below it.

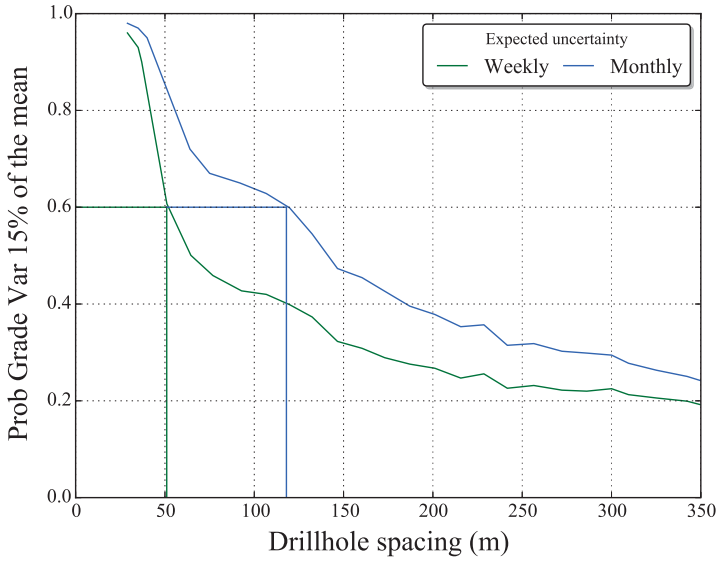

Figure 1.32: Production scale influence on uncertainty curve; for a given uncertainty level, a larger spacing is associated to a larger volume.

1.4.1 Re-Sampling Approach

In some cases of irregular drillhole spacing, there may be no closely drilled regions of the deposit. In such cases, re-sampling a simulated model and re-simulation can be used to add more information in uncertainty versus DHS curves.

Consider the drillhole configuration as shown in Figure 1.33. The actual data configuration is used to assess uncertainty in a simulation approach, and DHS is calculated using the original set of drillholes.

The DHS versus uncertainty for this drillhole configuration is shown in Figure 1.34. The standard deviation of simulated values is the measure of uncertainty. There is no information for standard deviation less than 0.2 units and drillhole spacing less than 40 units. One may try to extrapolate the expected uncertainty curve to small spacing, but this prediction would be very uncertain.

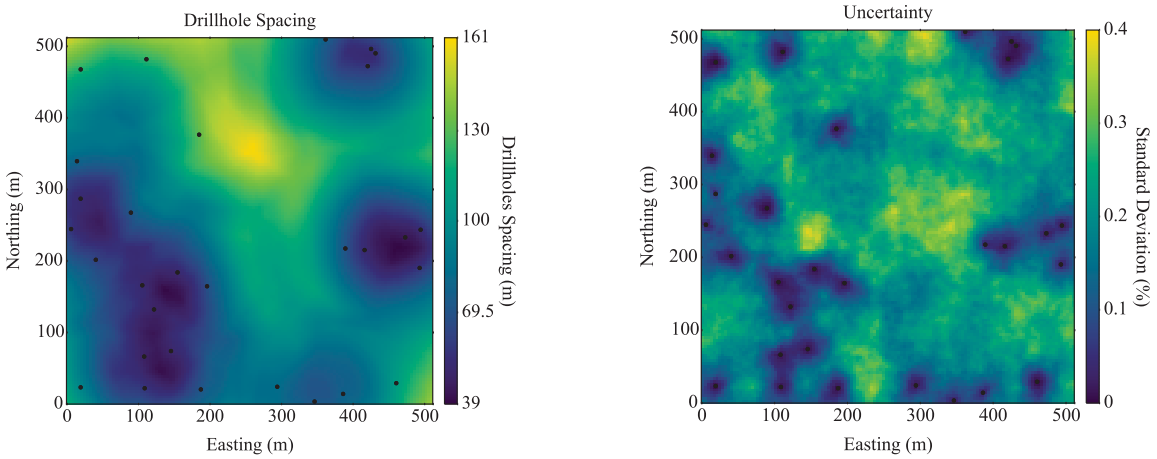

Figure 1.33: Drillhole spacing (on the right) and standard deviation (on the left) of the simulated values prior to resampling.

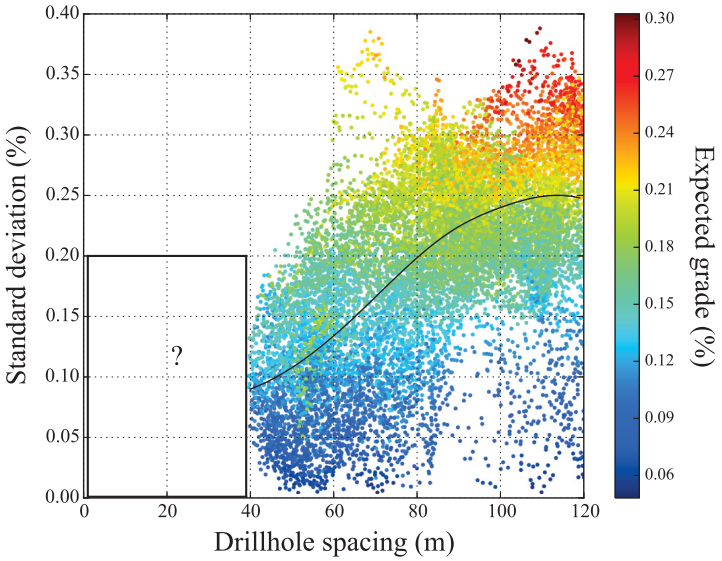

Figure 1.34: Uncertainty versus DHS curve prior to resampling. No information is available in small DHS.

Support with Simulated Data

An alternative to fill the missing information is to re-sample a realization as the unknown truth and re-simulate the attribute of interest. Consider sampling the model to create new artificial drillholes in any spacing required to predict a better expected uncertainty curve. From the original drillhole configuration, new drillholes are created and a new configuration is obtained. The updated DHS and uncertainty plots are shown in Figure 1.35.

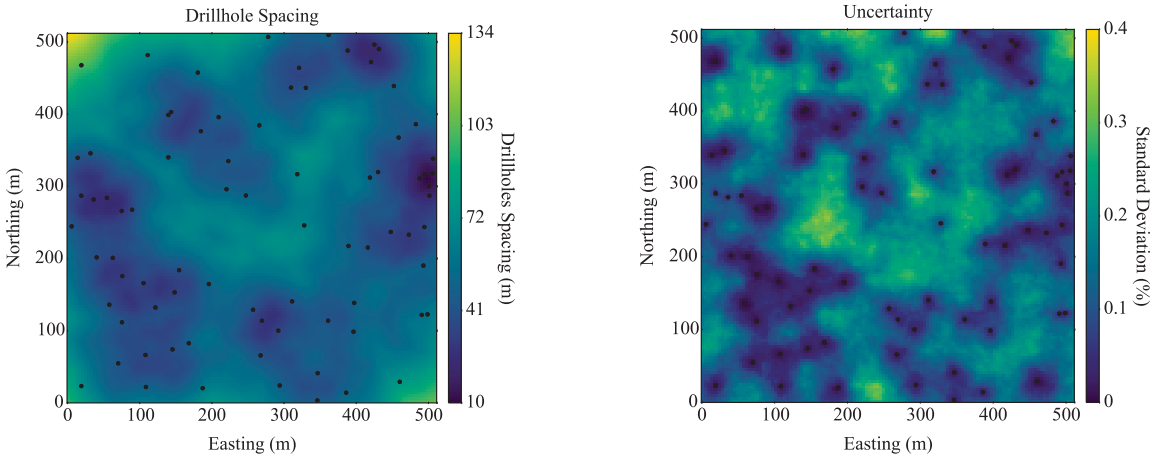

Figure 1.35: DHS and standard deviation of the simulated values after resampling (with new artificial drillholes).

The overall uncertainty decreases with more drillholes, and now there is information to predict a reliable expected uncertainty curve for DHS less than 40 units. All SMUs inside the squared area in Figure 1.36 were obtained from the simulation with the new drillhole data. Of course, we do not really have more drillholes at close spacing; these values are just to understand uncertainty versus drillhole spacing.

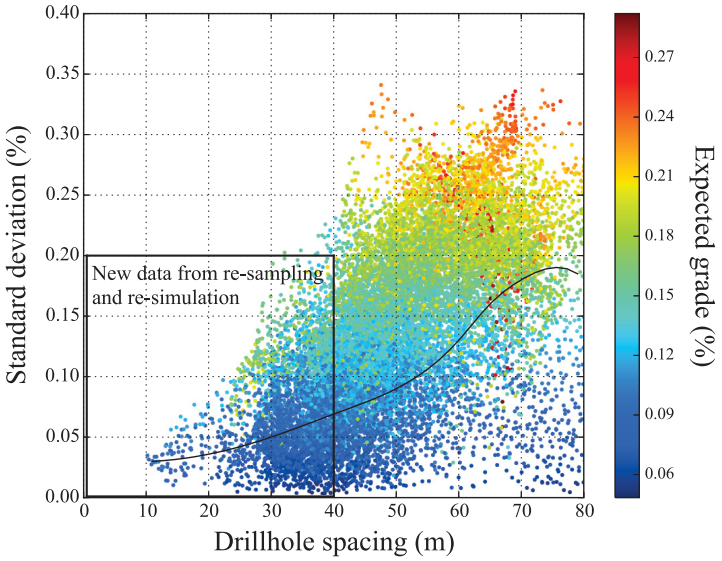

Figure 1.36: Uncertainty versus DHS curve after resampling. SMUs inside the squared area were obtained from the simulation with the new drillhole data.

1.5 Thesis Outline

Chapter 2 introduces and discusses the concepts of the Learning Curve and the value of collecting more data in terms of the rate of information gained. The regimes of uncertainty are important when targeting the right interval of DHS for a level of uncertainty. Chapter 3 reviews the various factors affecting uncertainty (mitigating factors) and discusses the use and application of different statistical models to quantify the influence of the explanatory factors on local variability. The case study is divided into Chapter 4 and Chapter 5. Chapter 4 presents the full workflow for uncertainty modelling. Categorical and continuous variables are modelled with modern multivariate techniques. Chapter 5 contains the data spacing and uncertainty study. Conclusion, remarks and advice are given in Chapter 6. The new programs developed during the time of this work are present in Appendix A.

Chapter 2: Learning Curve

The understanding of a mineral deposit is determined by the available data. In general, the more data that is available, the more that is known about a deposit. Drilling is expensive and there are additional safety and environmental costs. The search for the minimum amount of drilling that leads to an acceptable level of uncertainty is important. The relationship between uncertainty and data spacing depends on the variogram and many other factors. Understanding this relationship would help support decisions regarding further drilling.

A methodology based on kriging is developed to calculate the uncertainty as a function of data spacing and the variogram. The rate at which the variability is resolved with additional data is given by the first derivative of this function. This rate of information gained depends on the variogram model such as the spherical, Gaussian and exponential, and the scale considered for uncertainty reporting. There is less uncertainty at a block scale than a point scale.

The Learning Curve (LC) will show different values for different data spacing intervals. The rate of information gained is explained by the Learning Curve. An example based on the example of Section 1.3.1 is given to provide a better understanding of the concept.

2.1 Introduction

Data spacing is a geometric factor that influences uncertainty. Although data spacing units are easily understandable, commonly expressed in feet or meters; dimensionless units are a more general way to express data spacing. A practical dimensionless measure is the data spacing divided by the variogram range.

The variogram represents the spatial variability of a random variable and controls how uncertainty decreases with increasing data density (Deutsch and Beardow, 1999). The

variogram is used to calculate the uncertainty. Different methodologies such as kriging and simulation can be used to assess the uncertainty versus data spacing for a given variogram.

Consider the uncertainty versus data spacing plot from Figure 1.29, also shown below in Figure 2.1. For the example, one unconditional realization over a domain of 512 m x 512 m is generated and used as the true distribution of a random variable. This realization is sampled with grids of different average data spacing. Each of these samples is used as conditioning data for 100 realizations generated by SGS. The variogram is given by two nested isotropic spherical structures, with no nugget, 30 and 70 percent contributions and ranges of 16 and 64 m. The realizations are up-scaled to blocks of 4 m x 4 m and post-processed for uncertainty assessment. The probability of the grade to be within 10% of the mean is the measure of uncertainty.

The continuous line represents the conditional mean of uncertainty given the data spacing; referred to the expected uncertainty curve (EUC). The relationship between data spacing and uncertainty is clearly non-linear, but it is evident that uncertainty is higher for greater data spacing. Uncertainty is low for small data spacing and increases until a stable value at a spacing close to the maximum variogram range.

The rate that the uncertainty increases is higher at small data spacing. Consider an interval of values for the data spacing (S), from a spacing S_0 to another S_1 units. This interval is given by ΔS. Consider the change in the uncertainty (U) for this interval, say ΔU = $U_1(S_1)$ - $U_0(S_0)$ units. The rate (R) of uncertainty change for this interval is calculated by Equation 2.1.

$$R(S) = \frac{U_1 - U_0}{S_1 - S_0} = \frac{\Delta U}{\Delta S} \tag{2.1}$$

In general, the rate of information gain can be written as the first derivative of the expected uncertainty in relation to data spacing:

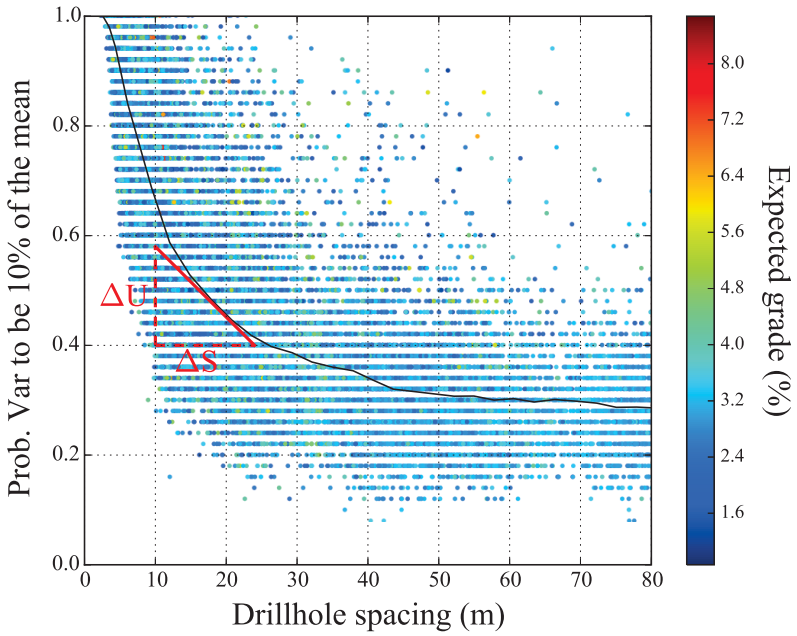

Figure 2.1: Rate of uncertainty changing in a uncertainty versus DHS curve.

$$R(S) = \frac{d(E\{U \mid dataspacing\})}{d(dataspacing)} = \frac{dU}{dS} \tag{2.2}$$

Consider two regions: (1) data spacing decreasing from S_1 = 30 m to S_2 = 10 m and (2) data spacing decreasing from S_3 = 50 m to S_4 = 30 m. The expected conditional uncertainty given these data spacing are U_1 = 38%, U_2 = 68%, U_3 = 30% and U_4 = 38%. The variation of uncertainty is then ΔU_{2-1} = 68 - 38 = 30% and ΔU_{4-3} = 38 - 30 = 8%. The rate of uncertainty changing for an interval of data spacing is then calculated by Equation 2.1, thus R_{12} = 1.5 %/m and R_{34} = 0.4 %/m.

The rate that the uncertainty changes for data spacing varying from 30 to 10 m is about four times higher than that for data spacing from 50 to 30 m. A greater understanding of this rate of learning would inform decisions related to data spacing.

2.2 Background

The commonly used variogram models are the spherical, Gaussian and exponential given in Equations 2.3, 2.4 and 2.5 (Rossi and Deutsch, 2014). As the variogram models have different continuity or shape, it is expected that uncertainty will be resolved at different rates for each model.

$$Sph(h) = \begin{cases} 1.5(h/a) - 0.5(h/a)^3, & \text{if } h \leq a \\ 1, & \text{otherwise} \end{cases} \tag{2.3}$$

$$Gaus(h) = 1 - exp(-3(h/a)^2) \tag{2.4}$$

$$Exp(h) = 1 - exp(-3h/a) \tag{2.5}$$

where:

a - range;
h - lag distance;

For a fixed variogram, there will be less uncertainty when the data spacing is small compared to the variogram range. For a continuous variogram, small changes in data spacing may not impact the uncertainty, since data is more redundant. Moreover, there may be data spacing values that do not affect uncertainty. The expected relationship between uncertainty and data spacing is given by Figure 2.2. Note the similarity with Figure 2.1.

The EUC of Figure 2.2 can be divided in three regions. Region 3 is related to the early stages of exploration, when data is sparse and uncertainty is high. Region 2 is the region of interest, for a small change in data spacing value, a large amount of the variability is

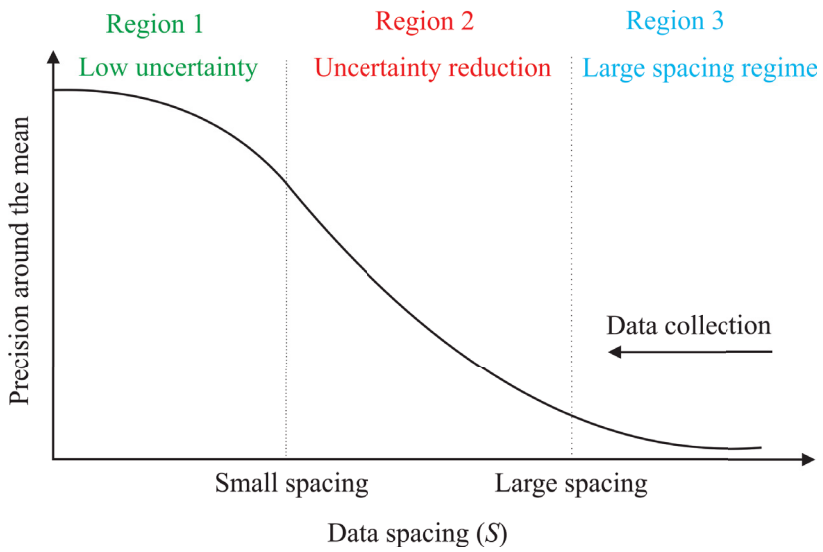

Figure 2.2: Expected relationship between uncertainty and data spacing for a fixed variogram. The regions of uncertainty changing for an interval of DHS are shown.

resolved. Region 1 is that region for small data spacing values, where uncertainty is very low since much information is already available.

The rate that the uncertainty changes for different data spacing is presented in Figure 2.3. This curve is the first derivative of the expected uncertainty curve of Figure 2.2. How variability reduces for an interval of data spacing is presented in three "regimes" based on the three data spacing regions. The arrows indicate the direction of data collection. Uncertainty is reduced when data is collected; the rate of reduction is what changes.

Data is sparse in Regime 3 and every extra drillhole will have a small contribution reducing the uncertainty. The highest rate that the uncertainty is resolved is in Regime 2. This is an important regime to help decide the data spacing that will lead to an acceptable level of uncertainty. The last regime (Regime 1) is where additional data does not influence the uncertainty significantly, and the decision of stopping data collection may be considered. Extra drillholes in this regime may not be worth the cost.

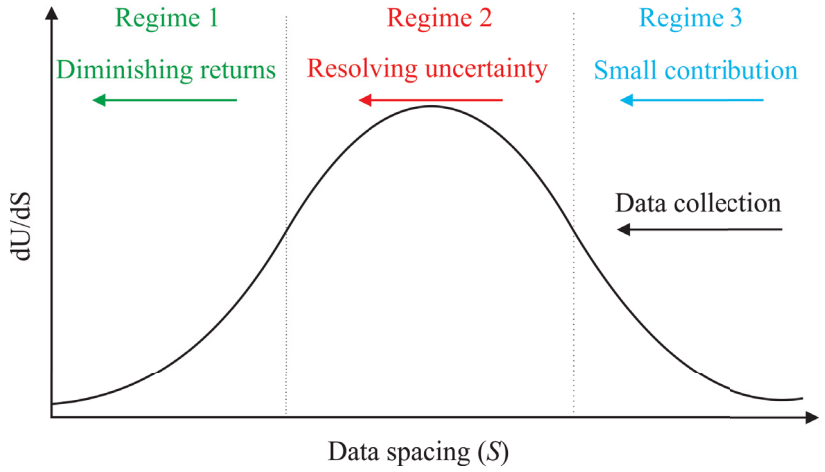

Figure 2.3: Expected derivative of the uncertainty as function of data spacing. The rate of information gained (the regimes) depends on the uncertainty regions defined in Figure 2.2.

2.3 Calculation of the Learning Curve

The Learning Curve calculated below is based on the kriging variance and relies on the standard stationarity assumption. The variogram is kept fixed and only the data spacing changes. For a given configuration of data in which drillholes are regulary spaced, the averaged kriging variance is calculated. The 1D example is the simplest case, see Figure 2.4. The kriging variance and the uncertainty as a function of data spacing are given by Equation 2.6 and Equation 2.7 respectively.

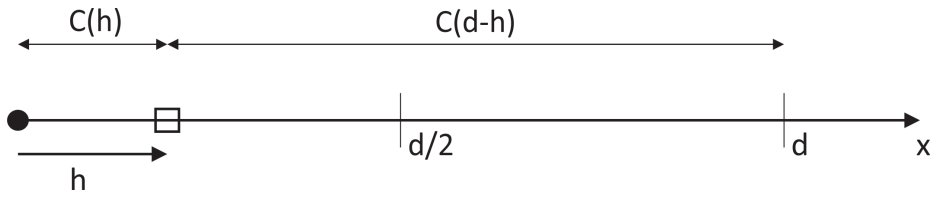

Figure 2.4: Scheme for data spacing (d) and estimation location for kriging-based methodology in 1D case.

$$\sigma_k^2 = 1 - \lambda_1[1 - C(h)] - \lambda_2[1 - C(d - h)] \quad \text{for} \quad h \in [0, d/2]; \quad d \leq a; \quad a = 1 \tag{2.6}$$

$$U(d) = \int_0^{d/2} \sigma_k^2 \, dh \tag{2.7}$$

where:

λ	- kriging weight;
a	- maximum standardized range;
$C(h)$	- covariance at lag h;
d	- distance (spacing);
h	- lag distance;

The model in 1D has no practical application. Moreover, the equations for a 2D case must respect geometric relationship between estimation and data locations. These equations are too complex and present some instability issues that avoid their direct use. It is necessary then, to work with krigring approach within a grid and discretized domain. In Figure 2.5, the area given by the large square is of interest, half the size of the data spacing (d), and discretized in $n \; x \; n$ estimation points or blocks, regularly spaced. In this case, a more general equation and its practical form are given in Equations 2.8 and 2.9.

$$\sigma_u^2(d) = \frac{1}{d^2} \int_{-\frac{d}{2}}^{\frac{d}{2}} \int_{-\frac{d}{2}}^{\frac{d}{2}} \sigma_k^2 \, dx \, dy \tag{2.8}$$

$$\sigma_u^2(d) = \frac{1}{n^2} \sum_i^n \sum_j^n \sigma_{k_{i,j}}^2 \tag{2.9}$$

These equations are used to calculate the EUC for a fixed variogram model with different data spacing configurations. For each model of data spacing, the kriging variance is calculated for all points or blocks and the average is calculated. Calculating the expected

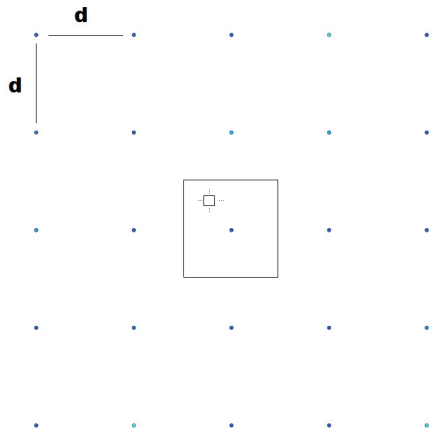

Figure 2.5: Scheme for data spacing (d) and estimation location for kriging-based methodology in 2D case. Data location is shown as dots.

uncertainty for different spacings using the same variogram model allows calculating and plotting the EUC. The first derivative of this curve is the Learning Curve.

2.4 Interpretation of the Learning Curve

The kriging-based methodology was used to plot the Learning Curve in different cases for point and block scales. The regions and regimes previously introduced are seen in these plots. The process to better know the deposit is continuous. Interpretation of the results should be in terms of resolving a scale of variability for different intervals of data spacing. For the example given in Section 2.1, the number of new drillholes required to change the data spacing from 60 to 40 m is much less than what is required to change the data spacing from 30 to 10 m, see Figure 2.6. Since an economic analysis is not given by the LC, a trade-off study is necessary to evaluate the cost of collecting more data.

2.4.1 Learning Curve Conceptual Plot

The uncertainty curve and the Learning Curve are better understood when plotted together and against dimensionless data spacing; such as the ratio of data spacing over the variogram range. In order to better understanding the building of the LC by the

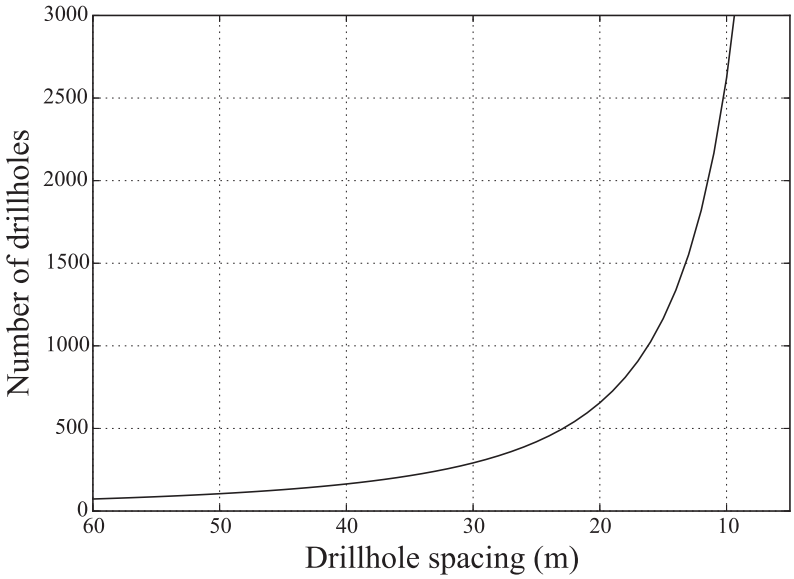

Figure 2.6: Relationship between number of drillholes and regular spacing for a 512 m x 512 m domain.

kriging-based methodology, consider the first of 3 plots shown in Figure 2.7. The blue line represents the EUC, whereas the red line is its first derivative, the LC. Moreover, consider the conceptual plots of Figures 2.2 and 2.3. The abscissa in all plots is switched; the dimensionless data spacing is plotted from large to small values, the direction of data collection.

Consider an initial large data spacing, say 2.5 times the variogram range, as the starting point. The measure of uncertainty is given by the average kriging variance for that spatial configuration of data widely spaced. Uncertainty at this point is high. Once more data is added into the model and the data spacing decreases, the average kriging variance (the blue line) starts to decrease, but at a low rate (the red line). The contribution of extra data resolving a scale of variability is small.

At a certain point (Figure 2.8), when a minimum quantity of data is already considered, the contribution of extra data increases considerably. In this regime, the rate of information gained when changing data spacing is high and variability is resolved faster.

There is a data spacing that is associated to a maximum rate of information, beyond this

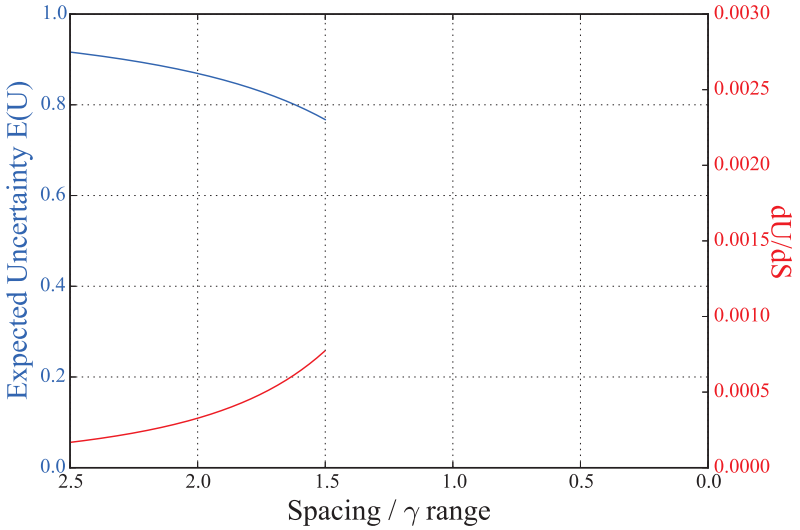

Figure 2.7: Step 1 in the Learning Curve; large data spacing region and small contribution regime.

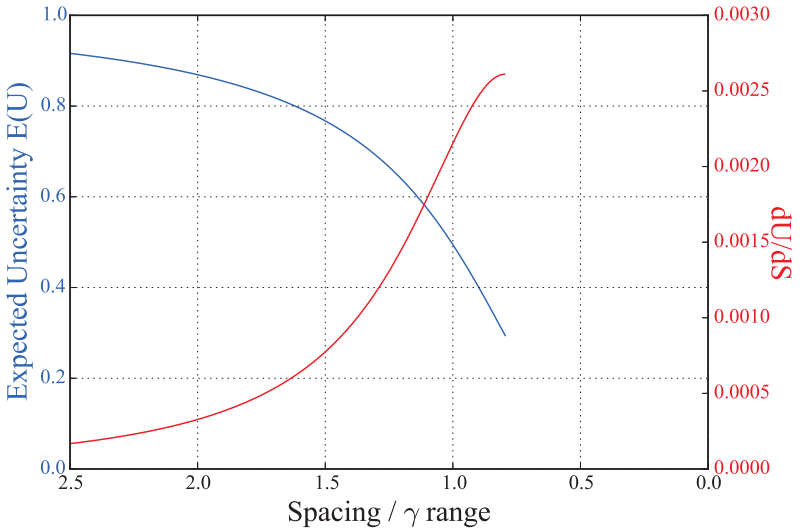

Figure 2.8: Step 2 in the Learning Curve; data spacing decreasing and resolving uncertainty regime.

data spacing (the inflection point) the contribution of extra data resolving a scale of variability decreases. It is important to note that uncertainty is reducing when data spacing changes throughout the EUC, however the rate that variability is resolved depends on certain intervals of data spacing. This rate of changing is captured by the LC. The EUC and the LC shown in Figure 2.9 was calculated using an isotropic Gaussian variogram with no nugget. The three regions of data spacing and regimes of uncertainty are shown as well.

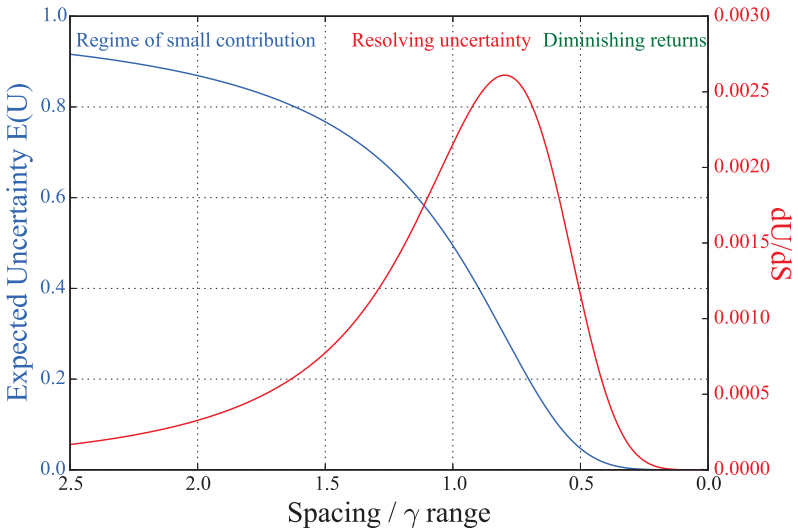

Figure 2.9: Step 3 in the Learning Curve; small spacing region and diminishing returns regime.

2.4.2 Point-Scale

The EUC and the LC for the spherical, exponential and Gaussian variogram models are shown in Figure 2.10. The greater continuity of the Gaussian model can be seen in the EUC for any data spacing when compared to the other models, since uncertainty is lower throughout the curve when this model is used. The values for the expected uncertainty and Learning curve for very small data spacing for the Gaussian model are close to but not equal to zero. The linearity near the origin of the spherical variogram can be seen in the LC for data spacing smaller than 0.5. The result from the exponential variogram is expected,

since the variance of this model increases fast for small ranges of correlation. The rate that this model resolves a scale of variability keeps increasing for smaller data spacing but does not have a maximum value as the other models. The maximum rate that uncertainty is resolved is reached with data spacing of 0.68 and 0.8 for the spherical and Gaussian model, respectively.

The Learning Curve shows the Gaussian as the model that faster resolves a scale of variability for data spacing greater than about 0.6. The exponential is the model that presents the higher rates below data spacing of 0.5. The regimes of uncertainty are a function of the variogram model. Each model has an interval of data spacing that resolves faster and a threshold or inflection point where the rate of information diminishes.

The influence of the nugget effect for the spherical variogram is shown in Figure 2.11. The nugget effect will increase the global variability, since for any data spacing value the expected uncertainty will be higher. Only the spherical model is shown; however the same analysis can be made for the Gaussian and exponential variograms.

2.4.3 Block-Scale

In practice, point-scale uncertainty is less relevant than for block-scale production volumes. A lower variability will be seen for block models. The expected estimation variance is lower for blocks since high and low values are averaged out when points are up-scaled to blocks (Figure 2.12). Despite the changes in the rate of the Learning Curve, the data spacing that provides a maximum value for the rate, in spherical and Gaussian models, do not change from points to blocks. The greater change observed is at the exponential model, which now has a data spacing of 0.17 resolving a maximum of the variability. For large data spacing regions, the expected uncertainty for the spherical variogram is the highest of all models, which is different than the point-scale model. However, when the LC is analysed, the shape of the curves are similar. The regimes defined from the point-scale LC are the same for the block-scale. The main difference between the two models is the greatest potential to resolve a scale of variability when a larger block scale is considered.

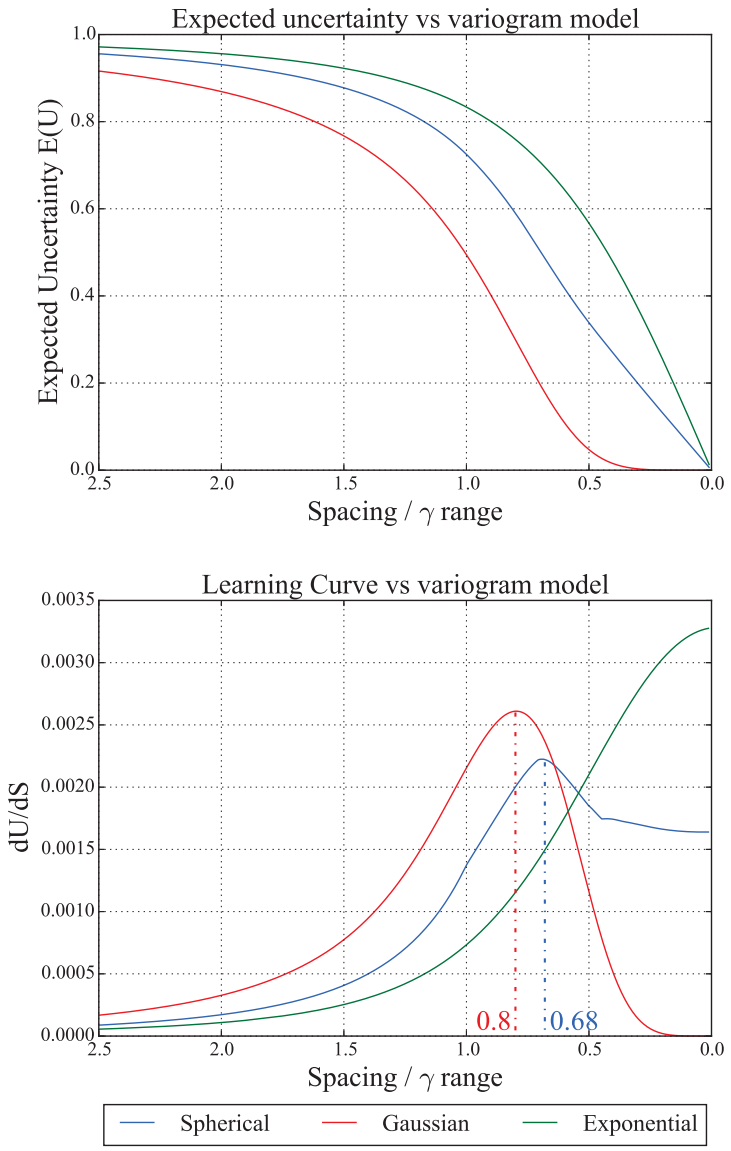

Figure 2.10: Expected uncertainty curve and the Learning Curve for a given variogram model at point scale.

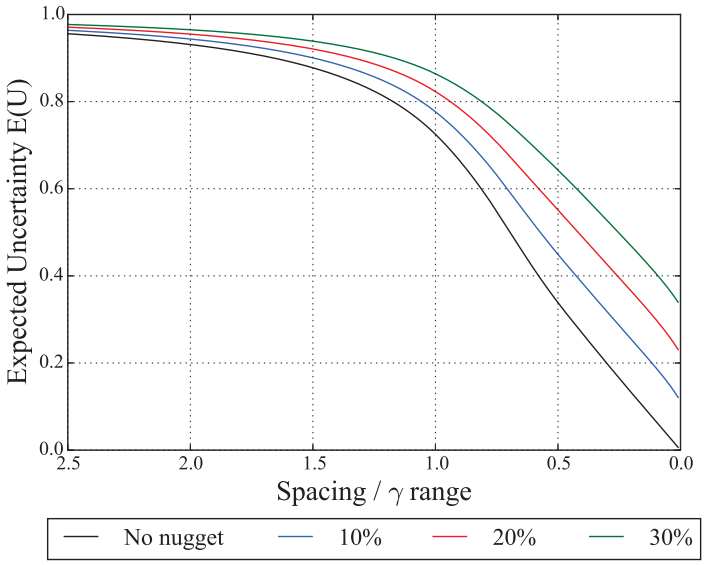

Figure 2.11: Expected point-scale uncertainty curve as a function of the nugget effect for the spherical variogram model.

For fixed variogram models, when the block size increases, the expected uncertainty decreases. Figure 2.13 shows the EUC and the LC for three different block sizes for the spherical and Gaussian variograms. Larger blocks will average out more low and high values, leading them to resolve faster a scale of variability.

The influence of the nugget effect in larger scale is shown in Figure 2.14 for a block size equals to 20% of the variogram range. For data spacing greater than about 0.6, the expected uncertainty decreases proportional to the nugget and block size. However, for small data spacing, a higher nugget will result in higher expected uncertainty. The contribution of the nugget effect on the variability becomes more important than the condition data for small spacing. Although the EUC changes for different intervals of data spacing, the LC will show little change, and the nugget effect will directly affect how the variability can be resolved. Increasing the nugget will lead to less capacity to resolve uncertainty, regardless of the block size or variogram model.

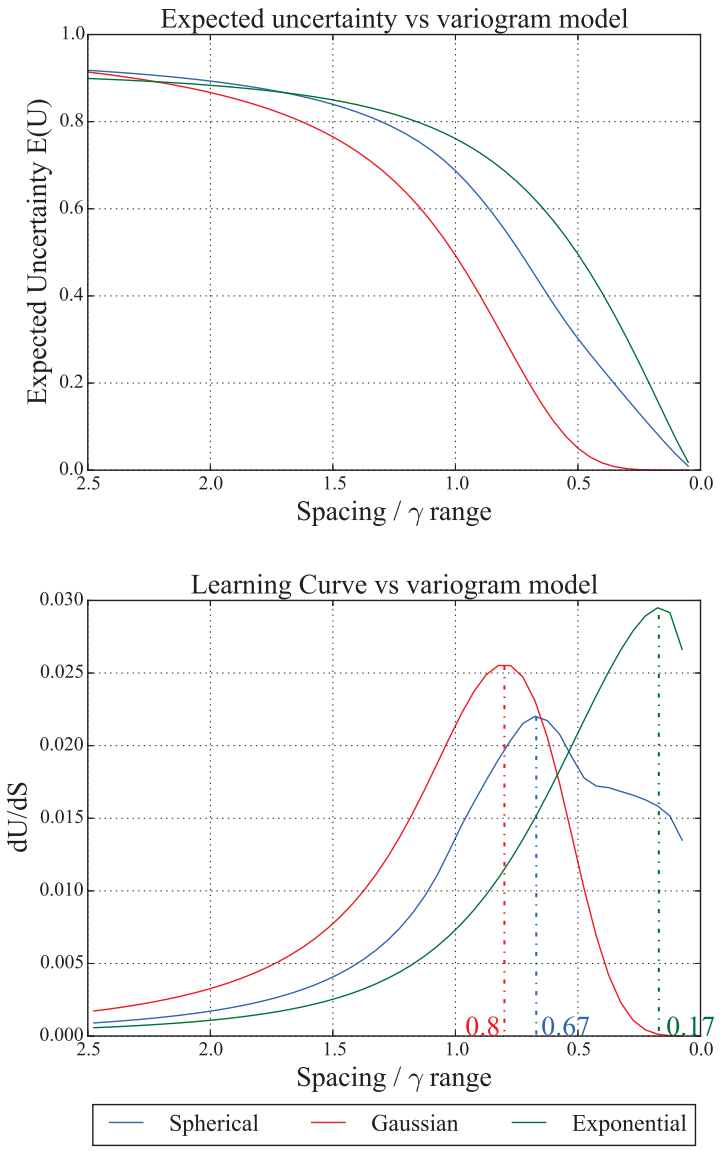

Figure 2.12: Expected uncertainty curve and the Learning Curve for a given variogram model at block scale.

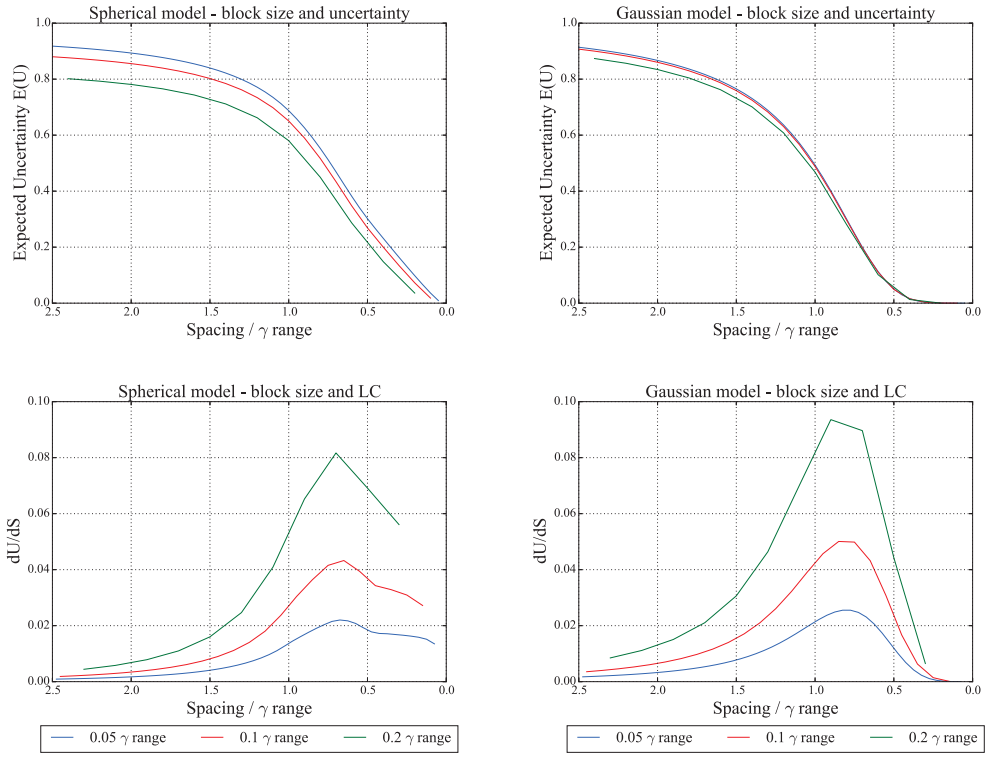

Figure 2.13: Block size influence on the expected uncertainty curve and the Learning Curve. Spherical variogram on the left and Gaussian variogram on the right.

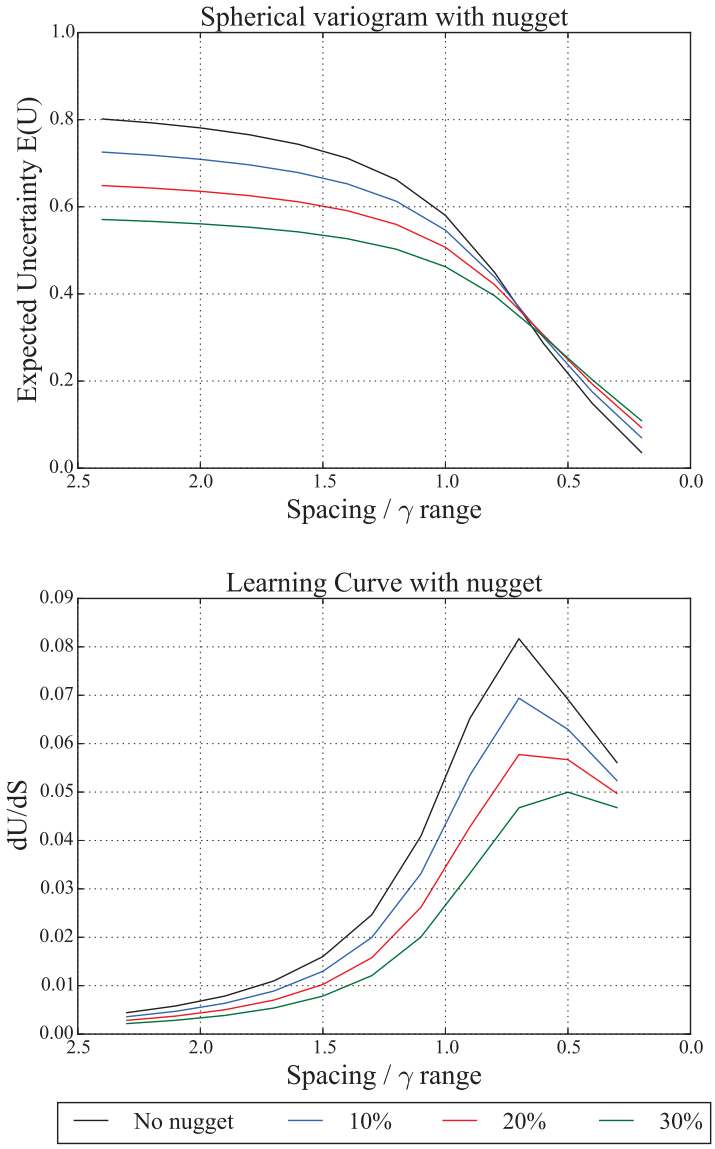

Figure 2.14: Expected block-scale uncertainty curve as a function of the nugget effect for the spherical variogram model.

2.5 Example

In pratice, the Learning Curve is calculated taking the first derivative of the EUC. The methodology presented above is based on kriging under a strong assumption of stationarity. For real data, the EUC is calculated from one of the methodologies discussed in the first chapter (Section 1). The theory developed to calculate the EUC and LC for a fixed variogram is academic. The practical calculation should be done on a case-by-case basis.

Consider the uncertainty versus DHS plot for a weekly production scale for RT1 in the example given in Section 1.3.1. Figure 2.15 shows the EUC in blue and the LC in red for this example. The LC is calculated taking the first derivative of the expected uncertainty curve. Since there is no precise or parametric mathematical function of uncertainty given the drillhole spacing for this case, the derivative is calculated for very small intervals of uncertainty and DHS along the EUC (see Equation 2.1). This is the reason of the noisy LC. Different numerical differentiation schemes could be considered.

The contribution of extra data reducing DHS from 350 m to 200 m is minimal. Uncertainty starts to decrease at a faster rate in the DHS interval of 200 m and 140 m. The interval of DHS between 140 m to 90 m shows little capacity to resolve a scale of variability, note the almost flat EUC in this interval. The highest rate of information gained is given by the DHS interval of 90 m and 50 m. Beyond this threshold the contribution of extra data diminishes, although uncertainty continues to decrease.

The Learning Curve is applied to a real data in the case study of Chapter 5.

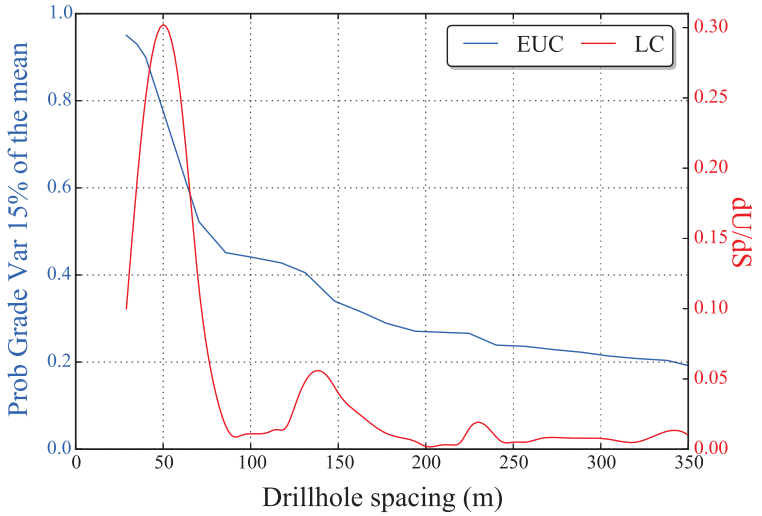

Figure 2.15: Expected uncertainty curve (blue) and the Learning Curve (red) calculated from synthetic example of Section 1.3.1.

2.6 Conclusion and Limitations

The Learning Curve is an interesting concept to help understand the influence of data spacing and the variogram on uncertainty for point and block scales. Expected and intuitive analyses are observed with synthetic variogram models.

The Gaussian variogram, because it provides the most continuous model, presents a higher capacity to resolve uncertainty. For the spherical and Gaussian models, the data spacing related to the maximum rate dU/dS, is the same for points and blocks. The exponential variogram presents a peak only when blocks are considered, and its capacity to resolve uncertainty is higher than the other models only in the regime of small data spacing. Uncertainty is also resolved faster for blocks rather than points.

The Gaussian variogram is a model generally used for continuous attributes, such as thickness and surfaces. The nugget effect is analysed only for the spherical model. Results show that the nugget decreases the expected block estimation variance for data spacing larger than 60% of the variogram range. For values smaller than that, a higher nugget will

increase the expected uncertainty, regardless of the block size.

For fixed variogram models increasing the block size results in lower uncertainty. This is more evident for the spherical and exponential variograms. Moreover, large blocks will resolve uncertainty faster than small blocks. The data spacing associated with a maximum rate remains the same, regardless of the block size.

The Learning Curve relies on a strong assumption of stationarity. The data configuration is perfectly spaced in a regular grid and the variogram is considered frozen over the domain. The application of the LC in real cases should be done over the EUC calculated for that particular deposit and drillhole spacing configuration. The calculation of the EUC for real deposits can be done by the methodology discussed in Section 1.3. The drillhole spacing does not need to be regular, the variogram can be made local and other parameter uncertainty can be considered when assessing uncertainty. The contribution of the LC in understanding drillhole spacing versus uncertainty is given by the analysis of the intervals of DHS that resolves faster a scale of variability (the regimes of uncertainty). This is done by taking the first derivative of the EUC.

Chapter 3: Uncertainty and Explanatory Factors

Explanatory factors are known factors that explain uncertainty. Besides data spacing, the most common explanatory factors are the local conditional mean, local conditional variance and entropy calculated across rock type proportions. The impact of these factors on uncertainty can be measured by different statistical regression models. Three explanatory models are presented: (1) simple and quadratic regression, (2) alternating conditional expectations and (3) stepwise removal of the factors. The capacity of these models to explain the total variability varies according to the regression model and the data. The order of importance of the explanatory factors depends on the geological domains.

Different measures of uncertainty, the explanatory models and the confounding factors are discussed in this chapter.

3.1 Review of Geostatistical Workflow for Uncertainty

Two methodologies to determine uncertainty versus drillhole spacing are presented. The methodology discussed in Section 1.3 can be applied to any deposit in which drillholes are not regularly spaced. When the deposit is drilled with different spacing, uncertainty can be related to a great range of drillhole spacing values and the expected uncertainty curve can be calculated with confidence. The plot of uncertainty versus DHS in this case is characterized by a cloud of points. There is a level of uncertainty for any DHS.

This methodology follows the steps below (as presented in Section 1.3):

1. Simulate multiple realizations to summarize uncertainty;

2. Average up the high resolution model to a large scale;
3. Assess uncertainty at the production volume scale;
4. Model and understand the influence of factors on local variability;
5. Plot uncertainty versus DHS/DHS curves;

A regular or quasi regular spacing would provide uncertainty for a single DHS value. Determining uncertainty at different spacing is achieved by sampling the model at different regular or random spacing to assess uncertainty for other drilling configurations. In relation to the previous case, two steps are added into the workflow after simulating the truth: (1) sampling the simulated truth at different regular/non-regular spacing and (2) generating a new true model conditional to the new sampled data. This methodology is discussed with an example in Section 3.3. The common steps of the uncertainty assessment workflow are discussed below.

3.1.1 Simulate the Truth

Realizations of the spatial distribution of a random variable $z(\mathbf{u}), \mathbf{u} \in D$ are generated by SGS for a number K of realizations, $\left\{z^k(\mathbf{u}), \mathbf{u} \in D, k = 1, ..., K\right\}$. A random variable is a function that associates real number (possible outcomes of an experiment) to each element in the sample space, assigning probabilities to the possible values (Ross, 2010; Walpole et al., 2012). These realizations are considered equi-probable, honouring the data, the data distribution and the variogram (Goovaerts, 1997). Measures of uncertainty are calculated from the different possible values of the random variable from a set of realizations. The histogram and variogram are reproduced within statistical fluctuations (Journel and Huijbregts, 1978).

When necessary, a simulated model can be sampled at a specified spacing. The data are then used to generate a new set of realizations.

3.1.2 Block Average

Journel and Huijbregts (1978) suggest simulating at fine resolution point scale, then averaging the simulated values up to a relevant scale of production such as the SMU. It is common practice to average 9 or more simulated points to compose a SMU.

The SMU value $\overline{z}_k$ is calculated by the arithmetic average of the simulated nodes n_v within the volume $v(\mathbf{u})$, if the random variable $z(\mathbf{u})$ scales arithmetically, see Equation 3.1 (Journel and Huijbregts, 1978). Different measures of uncertainty are calculated from the distribution of the K possible SMU values $\{\overline{z}_k(\mathbf{u}), \mathbf{u} \in D, k = 1, ..., K\}$.

$$\overline{z}_k(\mathbf{u}) = \frac{1}{|v|} \int_{v(\mathbf{u})} z_k(\mathbf{u}) d\mathbf{u} \simeq \frac{1}{n_v} \sum_{i=1}^{n_v} z_k(\mathbf{u}_i) \quad i = 1, ..., n_v \tag{3.1}$$

3.1.3 Common Measures of Uncertainty

The uncertainty in a location $\mathbf{u}$ depends on the data distribution, the variogram used for simulation and on the volume v. The probability distribution of all possible values (all realizations K) for a SMU characterizes the uncertainty. A single measure of uncertainty $\overline{U}_k$ for a realization is calculated as the average of the SMUs uncertainty at all location n_u (Equation 3.2). For a given data set, a single measure of uncertainty $\overline{U}$ is calculated averaging the uncertainty of all realizations K, Equation 3.3 (Wilde, 2010).

$$\overline{U_k} = \frac{1}{n_u} \sum_{i=1}^{n_u} U_k(\mathbf{u}_i) \tag{3.2}$$

$$\overline{U} = \frac{1}{K} \sum_{k=1}^{K} \overline{U_k} \tag{3.3}$$

Three measures of uncertainty are the most useful for geostatistics: the standard deviation, difference between percentiles and precision (Goovaerts, 1997). The data set previously

introduced in Section 1.4.1 is used as an example.

3.1.3.1 Standard Deviation

Given a random variable X with expected value of μ, the standard deviation σ is calculated by Equation 3.4.

$$\sigma = \left\{ E\left[(X - \mu)^2 \right] \right\}^{1/2} \tag{3.4}$$

The standard deviation is a measure of the spread of a distribution around its mean; it is the square root of the variance. The standard deviation has the same units as the variable. The uncertainty versus DHS curve when the standard deviation is used as measure of uncertainty is shown in Figure 3.1.

The coefficient of variation (CV) can also be used (see Figure 3.2). This coefficient is calculated dividing the standard deviation by the mean of the distribution, $CV = \sigma/\mu$. In general, the standard deviation increases when data spacing increases.

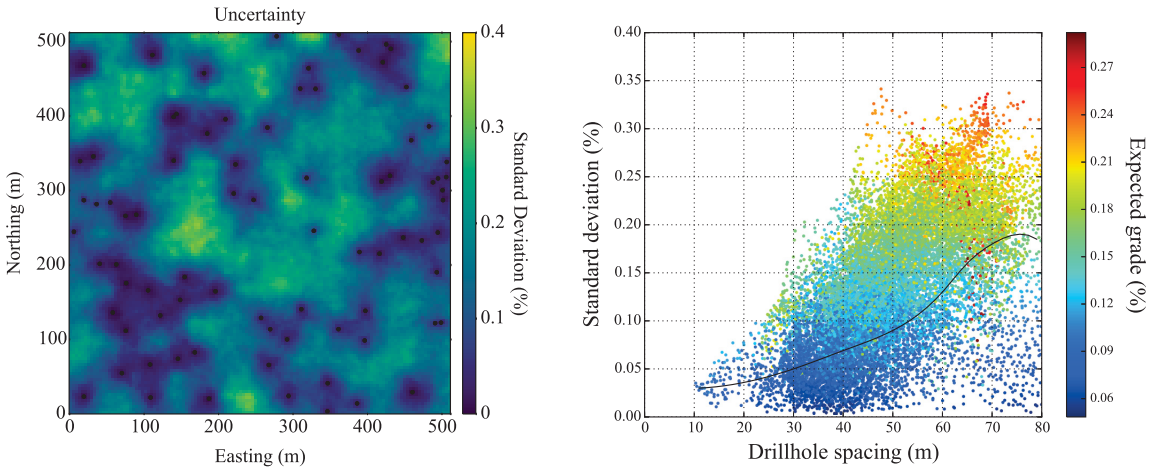

Figure 3.1: Uncertainty versus drillhole spacing curve when the standard deviation is the measure of uncertainty. Plan view of the uncertainty on the left and uncertainty versus DHS curve on the right.

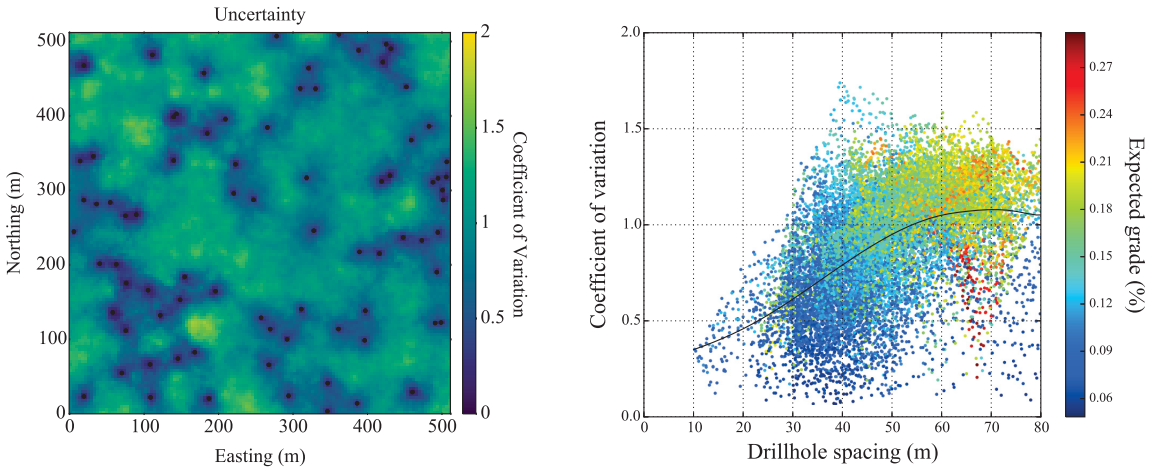

Figure 3.2: Uncertainty versus drillhole spacing curve when the coefficient of variation is the measure of uncertainty. Plan view of the uncertainty on the left and uncertainty versus DHS curve on the right.

3.1.3.2 Difference Between Percentiles

Montgomery and Runger (2002) define the percentile as the set of values that divide the sample into 100 equal parts; the pth percentile is the data value (X_p) such that p% of the observations are at or below this value. Difference between percentiles is also a measure of the spread of a distribution. The percentile values X_p are calculated based on the cumulative distribution function (CDF), as shown in Figure 3.3.

The difference between two symmetric percentiles such as P90 and P10, $\Delta_p = X_{0.9} - X_{0.1}$, is used as a measure of uncertainty. The difference between percentiles have the same unit of the variable. It is also common to standardize this difference by the P50, for a unitless measure of uncertainty, as shown in Equation 3.5.

$$\Delta_p = \frac{X_{0.9} - X_{0.1}}{X_{0.5}} \tag{3.5}$$

The DHS and uncertainty relationship when the standardized difference between percentiles

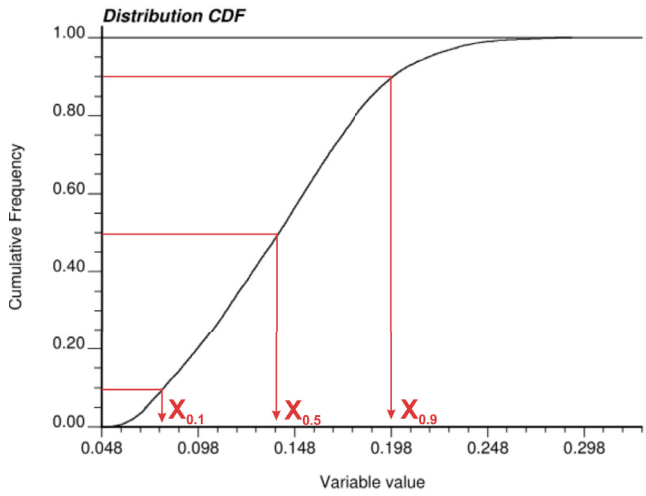

Figure 3.3: Three percentiles (P10, P50 and P90) from a CDF and their respectively values ($X_{0.1}$, $X_{0.5}$ and $X_{0.9}$).

is used as the measured of uncertainty is shown in Figure 3.4. This difference is larger when the data spacing increases.

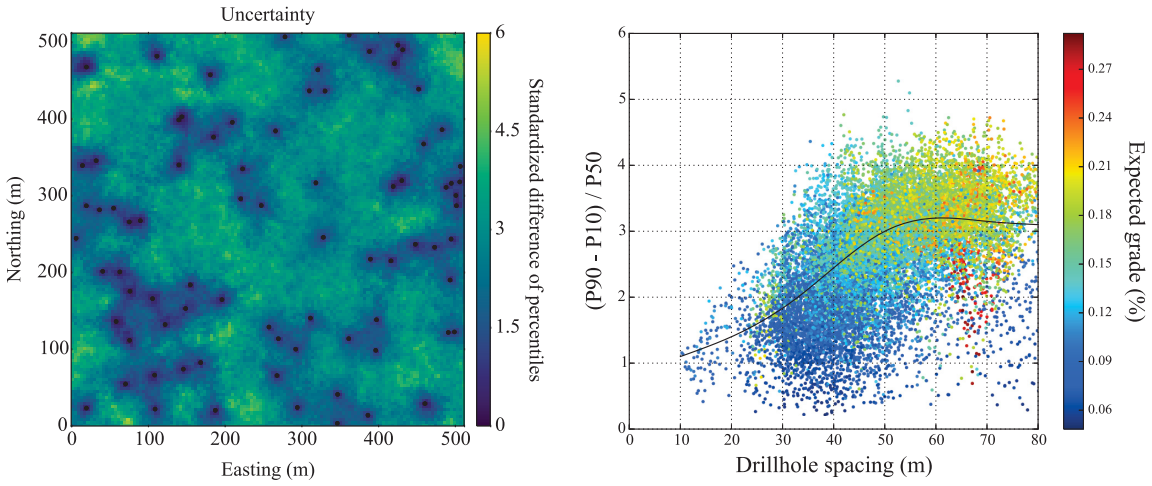

Figure 3.4: Uncertainty versus drillhole spacing curve when the standardized difference between percentiles is the measure of uncertainty. Plan view of the uncertainty on the left and uncertainty versus DHS curve on the right.

3.1.3.3 Precision

Precision is another measure of uncertainty. Precision calculates how many times the value of the random variable falls within an interval of the distribution mean. Given a constant value r (the probability), the distance h from the mean $\mu(\mathbf{u})$ is calculated by the product $h = r \times \mu(\mathbf{u})$. Thus, a precision of 20% measures the probability of the simulated SMU grade values $z(\mathbf{u})$ to be within ±20% of the mean. Precision is calculated counting the number of realizations that the simulated SMU value fell within that range of values.

Let $C_k(\mathbf{u}; \mathbf{h})$ be the indicator defined by Equation 3.6. The precision at a location $\mathbf{u}$ is then calculated averaging $C_k(\mathbf{u}; \mathbf{h})$ for all realizations k (Equation 3.7).

$$C_k(\mathbf{u}; \mathbf{h}) = \begin{cases} 1, & \text{if } \mu(\mathbf{u}) - h \leq z(\mathbf{u})_k \leq \mu(\mathbf{u}) + h \\ 0, & \text{otherwise} \end{cases} \tag{3.6}$$

$$p(\mathbf{u}; \mathbf{h}) = \frac{1}{K} \sum_{k=1}^{K} C_k(\mathbf{u}; \mathbf{h}) \tag{3.7}$$

In Figure 3.5, the probability of the SMU grade values to be within 15% of the mean is plotted against DHS. Precision and DHS are inversely related. Simulated SMU grade values are more precise when DHS decreases.

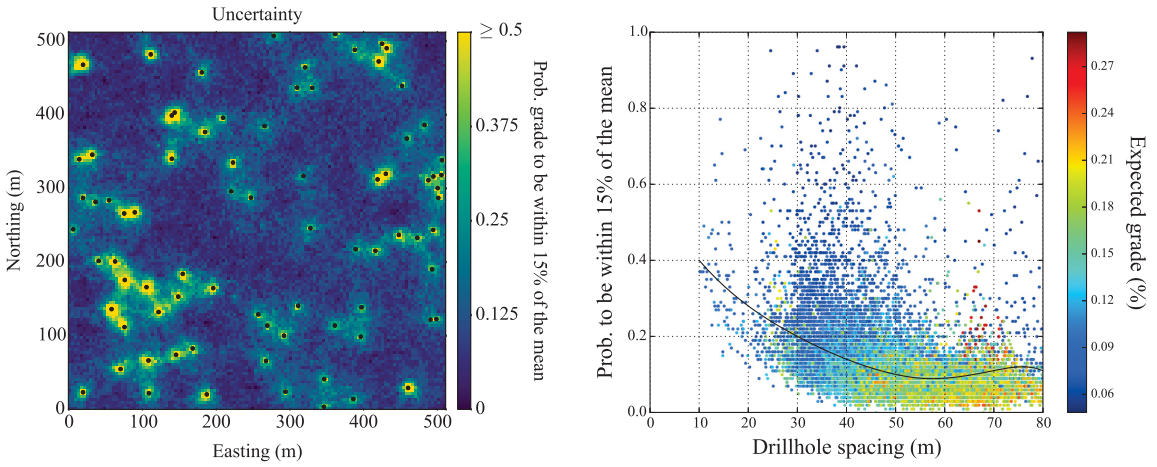

Figure 3.5: Uncertainty versus drillhole spacing curve when the precision is the measure of uncertainty. Plan view of the uncertainty on the left and uncertainty versus DHS curve on the right.

3.2 Common Confounding Factors

Common confounding factors are the proportional effect, scale and stationarity. These factors are briefly discussed in this section. Less common factors such as the number of realizations, parameter uncertainty, misclassification and data quality also disturb the relationship between uncertainty and data spacing (Wilde, 2010). Some of these factors affect uncertainty in a non predictable way (e.g. stationarity); some are predictable (e.g. proportional effect).

3.2.1 Proportional Effect

The proportional effect occurs when a random variable has a skewed distribution (Journel and Huijbregts, 1978; Pinto and Deutsch, 2014; Wilde, 2010). Positively skewed distributions will generaly result in more uncertainty in high estimates areas, whereas negatively skewed histograms will result in high uncertainty in low estimates areas (Pinto, 2015).

The proportional effect can be seen directly from the model, comparing the expected value and the standard deviation as shown in Figure 3.6. High standard deviation is seen in high mean zones. This data set has a positively skewed distribution (the histogram is not shown). Another data set with positively skewed histogram is used to demonstrate the proportional effect, in Figure 3.7.

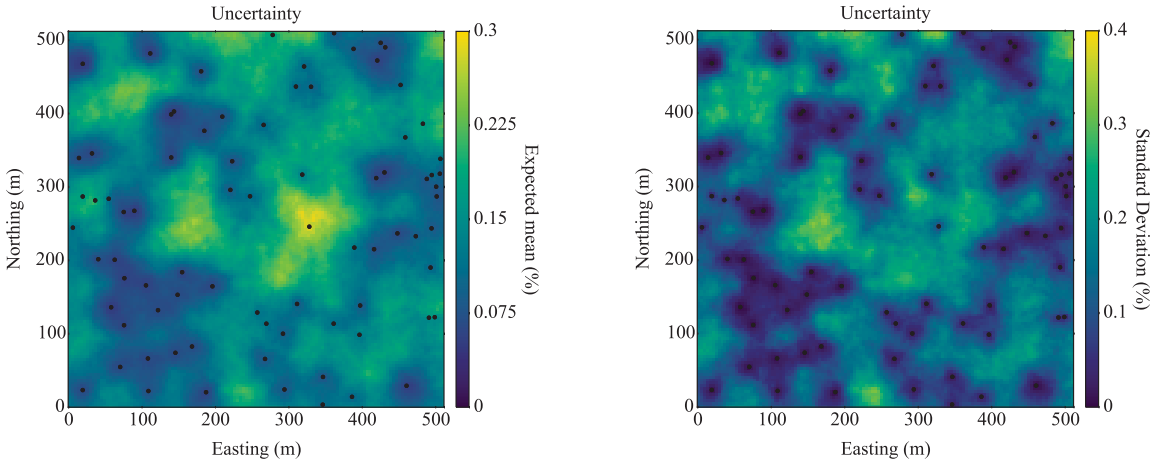

Figure 3.6: High deviated grade areas are associated with high local grade mean - conditional proportional effect. Expected mean (grade) on the left and standard deviation on the right.

The proportional effect is also related to transition zones and spatial trends, it may also indicate a lack of spatial homogeneity (Rossi and Deutsch, 2014). An example of transition zone is given in Figure 3.8. High grade and low grade zones are less uncertain, the highest variability is seen at the borders and in the zone located between the high (orange and red) grade and low (blue) grade regions, compare these regions in the figure. This behavior is typical of bimodal histograms, see Figure 3.9 for this data distribution and scatterplot. In this case the proportional effect is not associated to a high variability in high mean regions, but to the transition zone.

In some cases, high value areas are more certain, the opposite of the proportional effect. The uncertainty in these areas is closely related to the spread of the local distribution and the

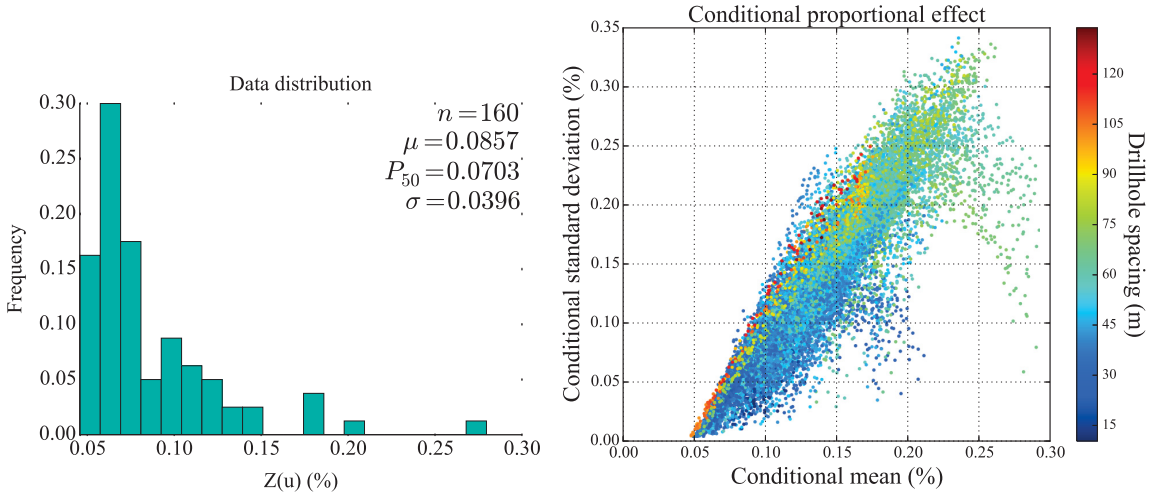

Figure 3.7: Conditional proportional effect for a positively skewed distribution. The histogram is shown on the left and the scatterplot on the right. The dots represent the SMUs, colored by the drillhole spacing.

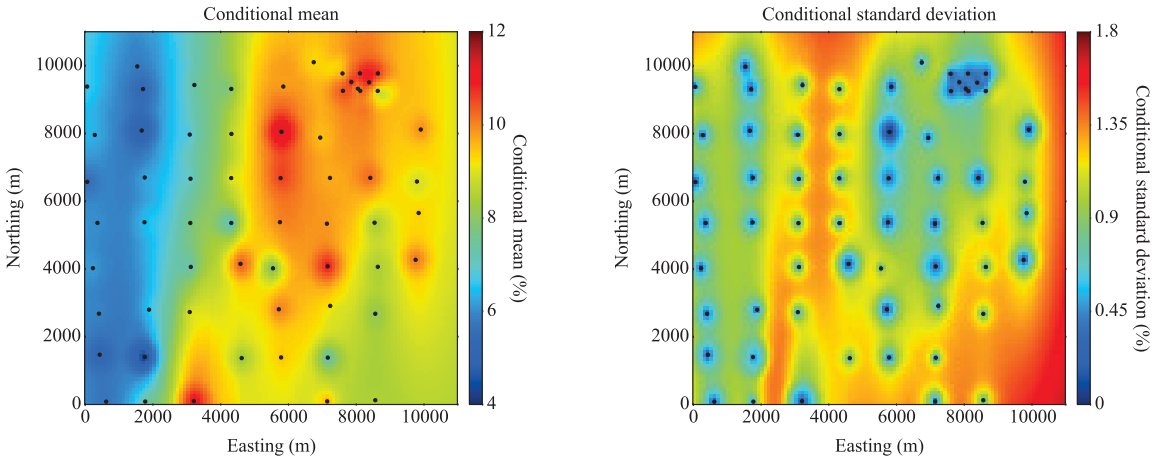

Figure 3.8: Conditional proportional effect due to a transition zone. The conditional mean in shown on the left and the conditional standard deviation on the right.

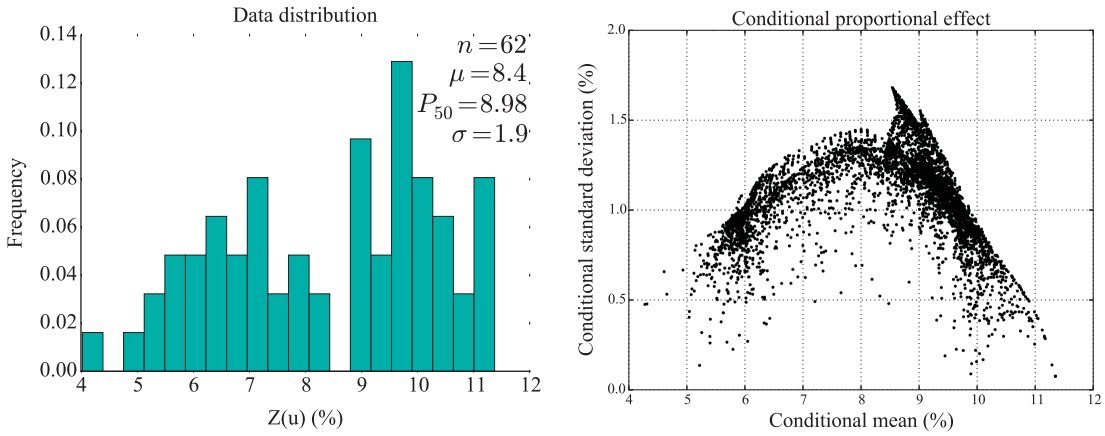

Figure 3.9: Conditional proportional effect for a bimodal distribution. The histogram is shown on the left and the scatterplot on the right.

proportion of high values around the local mean. A generalized graphical representation of the proportional effect by histogram is given in Figure 3.10 (Pinto and Deutsch, 2014). The proportional effect is mitigated when considering larger volumes because the distributions become more symmetric.

3.2.2 Scale

Pyrcz and Deutsch (2014) discuss three main considerations when choosing the grid size: a model should be built for a specific project goal, the grid size must be chosen to resolve important geological features (faults, bounds, stratigraphy) and the grid size can be scaled up to a meaningful larger scale. Journel and Huijbregts (1978) proposed a block size from 1/3 to 1/2 of the drillhole spacing to avoid artificial smoothing for very small blocks and not loosing information from data for very large blocks. The choice of the SMU size is based on production scale, mining operation and engineering factors whereas larger production scale size is considered for long-term planning (Rossi and Deutsch, 2014). The choice of scale is also dependent on computational efficiency and level of details that can be storage (Pyrcz and Deutsch, 2014).

When increasing scale, high and low values are averaged out. The effect of scale is

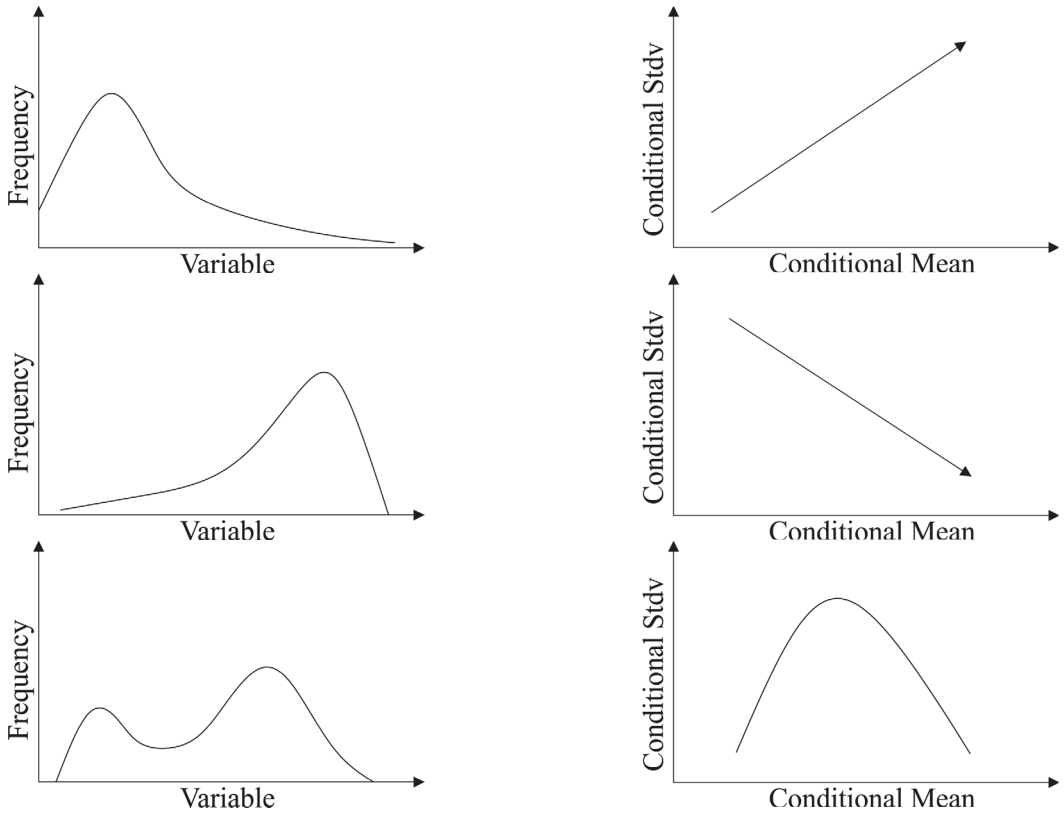

Figure 3.10: A generalization of expected conditional proportional effect scatterplots for some typical histograms. The positevely skewed histogram is shown in the upper left corner, the negatively skewed histogram is in the center left and the bimodal histogram is the bottom left graph. Their expected conditional proportional effect are shown in the graphs to the right.

demonstrated by considering uncertainty, the precision (15%), in three different production scales (Figure 3.11). The effect of scale is visually evident in the colourmaps. The change in the uncertainty distributions is seen at the histograms, as the average of the distributions increases for larger scales.

A clear representation of the effect of scale on uncertainty is shown by the EUC for the three production scale in Figure 3.12. For any DHS the SMU scale is the less certain, as the scale increases the uncertainty decreases.

3.2.3 Stationarity in the Variogram

This decision regards which data should be pooled together for further statistical analysis, since all sample statistics such as the histogram, and geostatistical tools such as the variogram, refer to a population and not to any sample in particular (Pyrcz and Deutsch, 2014). Data is usually grouped together by geological domains (facies, rock type) in which important properties are said to be constant or show little spatial variation over that domain. The decision of stationarity affects uncertainty.

The variogram is commonly assumed constant, although spatial continuity may change locally in a domain. The variogram or covariance function, measure the spatial correlation of samples spaced by a distance **h**, said lag distance. The correlation decreases for samples apart until the variogram reaches the sill, at this point the correlation is zero and samples are not linearly correlated. The variogram depends on the model decision of stationarity, the mean and variance are considered constant and independent of location (Rossi and Deutsch, 2014).

Despite the decision of stationarity, the variogram may change locally, in zones of more or less continuity inside a domain. Uncertainty is dependent on the continuity; usually uncertainty is higher in less continuity areas and lower in high continuity areas. The influence of the variogram in uncertainty versus data spacing curves was previously discussed in Chapter 2. Another example is given to show how continuity affects

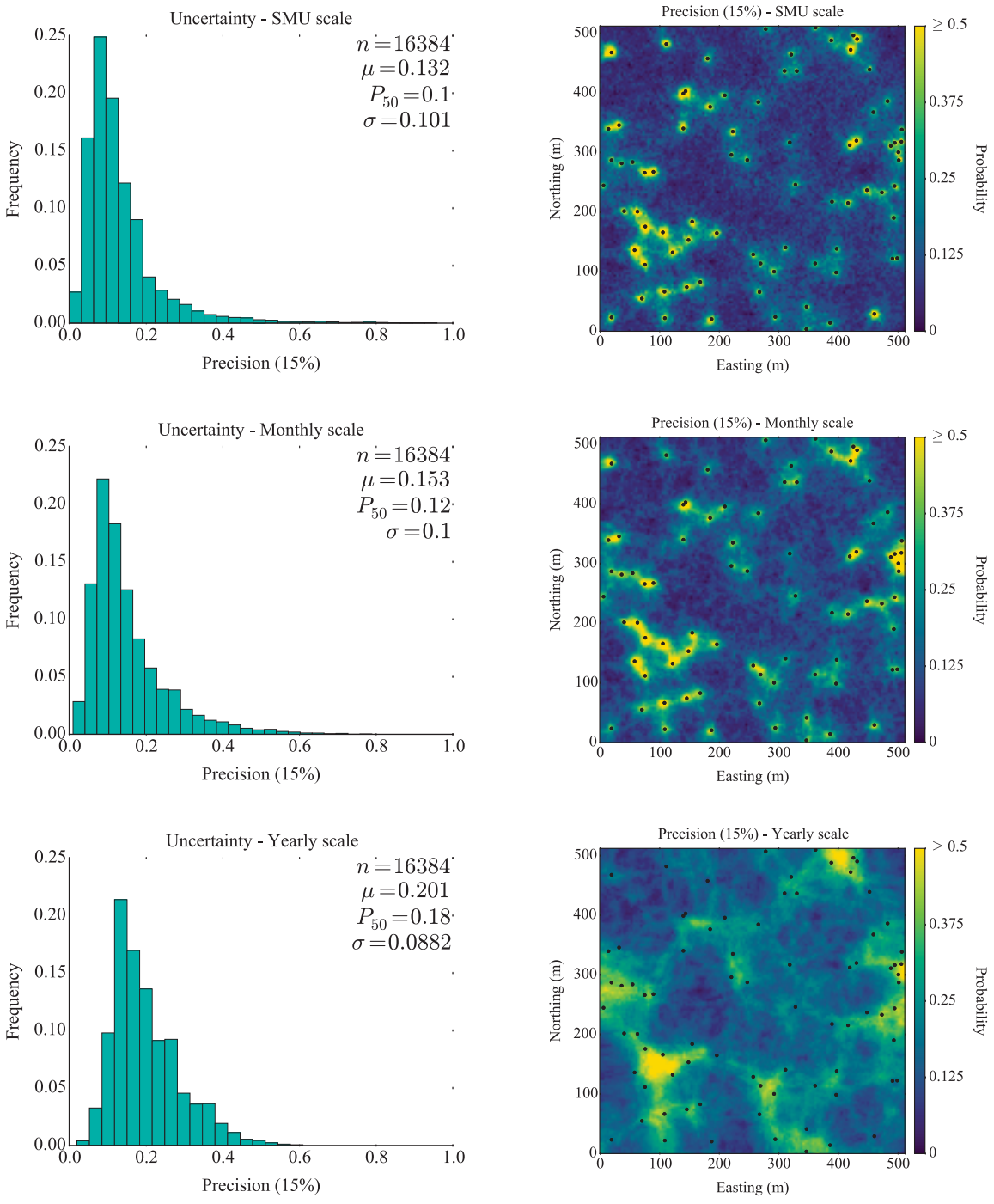

Figure 3.11: The effect of scale on uncertainty. Three different production scale and their uncertainty distribution on the left and plan view of the uncertainty on the right.

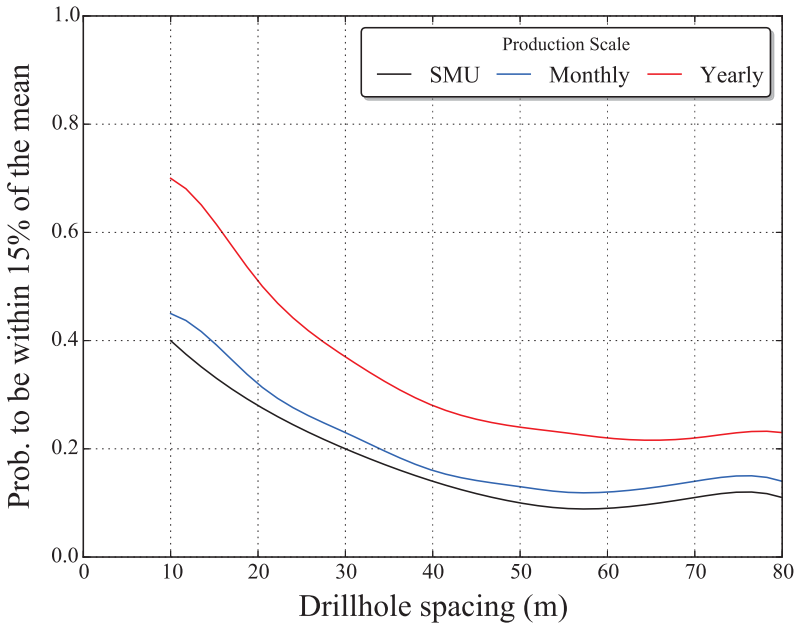

Figure 3.12: EUC for three different production scale. The yearly production scale uncertainty curve is more certain for any DHS.

uncertainty.

The data set of Section 3.1.3, is used in this example. The standard deviation (Figure 3.6) of the simulated variable was calculated for three different variogram models: spherical with no nugget effect, one nested structure and ranges of $150m$, $250m$ and $300m$, see Figure 3.13. The short range variogram results in the greatest uncertainty for any DHS. Uncertainty decreases as continuity increases. Wilde (2010) demonstrated that the distribution of uncertainty is more spread for DHS near the variogram range and less variability in uncertainty is observed for DHS less than and greater than the variogram range.

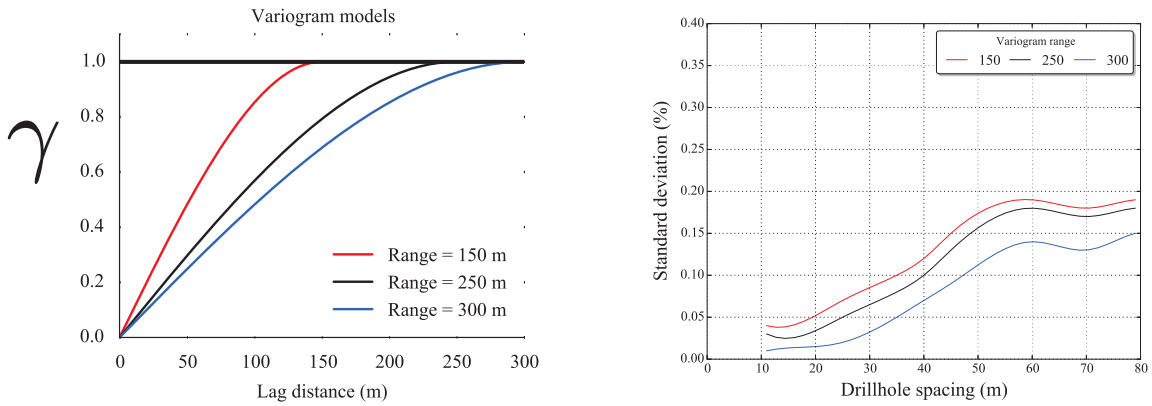

Figure 3.13: The effect of stationarity in the variogram on uncertainty.

3.3 Example - Regular Spacing Methodology

To illustrate this methodology, consider a reference true model (Figure 3.14) generated by unconditional simulation using one nested structure spherical variogram, with no nugget effect and a range of 250m. For demonstration purposes, all values are kept in normal units.

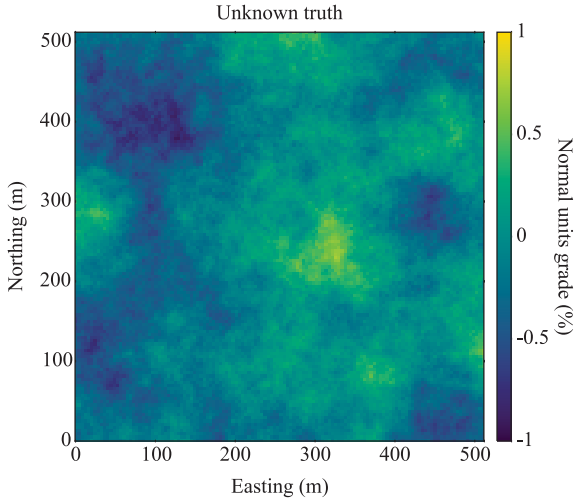

Figure 3.14: Unknown true model that will be sampled with different DHS.

This model is sampled with drillholes regularly spaced in 5 different grids, 8m x 8m, 16m x 16m, 32m x 32m, 64m x 64m and 128m x 128m, as shown in Figure 3.15 for 32m and

$128m$ spacing.

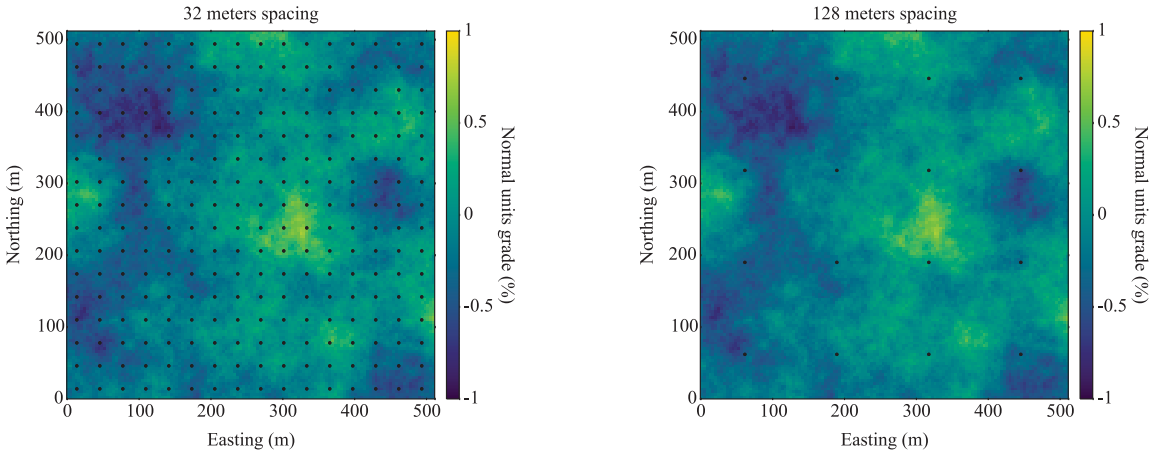

Figure 3.15: Spatial location of the drillholes. $32m$ spacing on the left and $128m$ spacing on the right.

For each one of these data configurations a total of 100 realizations of conditional simulation are generated. It is expected that the models sampled at a smaller spacing will better reproduce the true model. The simulated nodes are block averaged and a single measured of uncertainty is calculated for each model, the average standard deviation of the simulated values. The capacity to reproduce the truth and the standard deviation for the models of $32m$ and $128m$ spacing are shown in Figure 3.16.

The expected uncertainty versus DHS curve is shown in Figure 3.17. Since the spacing between drillholes is regular, and a single average measure of uncertainty is calculated per model, there is no cloud of points in the scatterplot. The EUC is connected linearly between the calculated points. The original drillholes can be left in the model and the new drillholes (from sampling) are added to the data distribution prior conditional simulation.

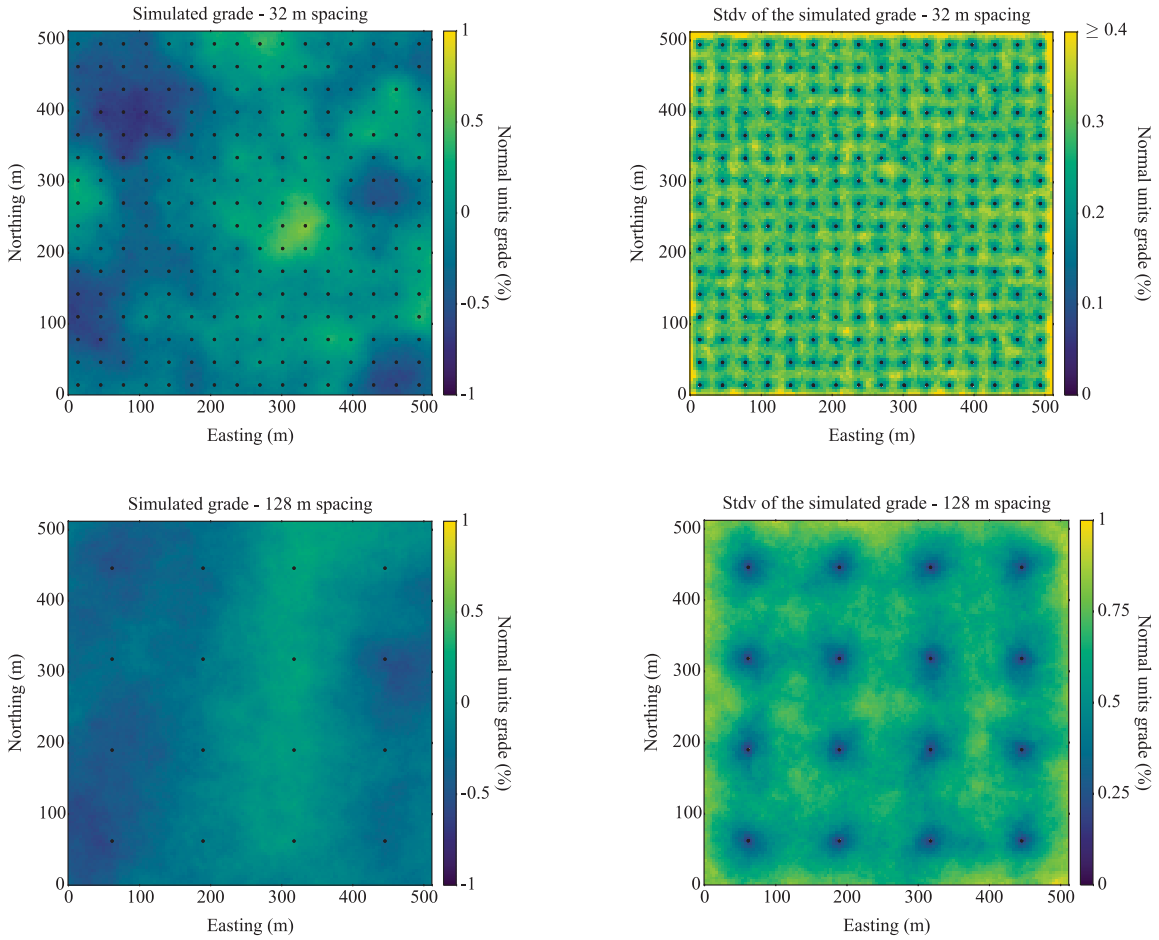

Figure 3.16: Plan view of the simulated models (on the left) and uncertainty (on the right) for the drillholes at $32m$ (on the top) and $128m$ spacing (on the bottom).

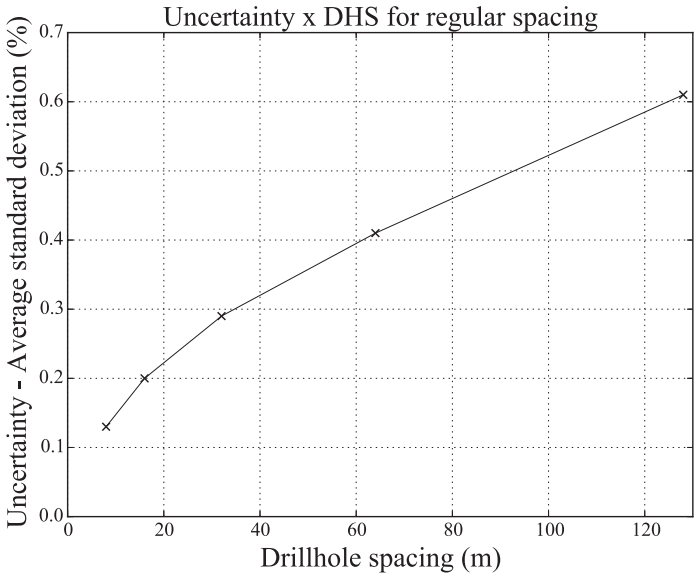

Figure 3.17: Expected uncertainty for the five different regular drillhole spacing. The average uncertainty for each drillhole spacing is shown as 'x'. The EUC is shown as the continuous line.

3.4 Explanatory Models

Uncertainty is mainly explained by factors such as data spacing, local mean, local standard deviation, and entropy. An explanatory model is a statistical model based on regression, able to quantify the amount of uncertainty in a response variable that depends on the explanatory factors (predictor variables). The predictor variables are also referred to as independent variables affecting the variance in the response variable. Three regression analysis models are presented: multiple linear regression, ACE and stepwise removal of the factors. Regression techniques or analysis of the variance (ANOVA) methods are abudant in the literature and can be used for estimating the relationship among variables.

For the explanatory factors and models explanation, consider the same data set from the example given in Section 1.3.1.

3.4.1 Explanatory Factors

Explanatory factors were introduced in Chapter 1, Section 1.3. They are calculated at the production scale being considered. The most common explanatory factors are explained below.

3.4.1.1 Drillhole Spacing

Drillhole spacing is an important explanatory factor because drilling is a direct source of information and all estimates are affected by the data. Estimates are better when close to drillholes and less precise away from the data. Drillhole spacing and density calculations were demonstrated in Section 1.2.2. When drillholes are unevenly distributed over an area, then drillhole spacing must be calculated locally.

Areas with small DHS, or more densely drilled areas, will likely be more certain, therefore the impact of this factor on uncertainty depends on the local DHS.

3.4.1.2 Conditional Mean and Standard Deviation

Conditional mean $\overline{\mu}_V(\mathbf{u})$ and standard deviation $\overline{\sigma}_V(\mathbf{u})$ are two factor calculated based on the grade estimates. Given all N SMUs grade estimates $z(\mathbf{u})$ at a production volume $V(\mathbf{u})$ over a set of realizations K (for $\mathbf{u} \in D, k = 1, ..., K; N = n_v \times K$), the conditional mean and standard deviation are calculated by Equations 3.8 and 3.9 respectively.

$$\overline{\mu}_V(\mathbf{u}) = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_v} z_k(\mathbf{u}_i) \tag{3.8}$$

$$\overline{\sigma}_V(\mathbf{u}) = \sqrt{\frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{n_v} [z_k(\mathbf{u}_i) - \overline{\mu}_V(\mathbf{u})]^2} \tag{3.9}$$

These factors define the conditional proportional effect (Section 3.2.1). The influence of

these two factors on uncertainty depends on the data distribution and the local proportion of high and low values.

3.4.1.3 Entropy

Entropy captures the uncertainty due to the mixing of rock types. "*Entropy can be seen as a measure of the uniformity in the distribution of the available categories of a variable. Entropy increases both with the uniformity in the distribution and with the number of categories*" (Darcy and Aigner, 1980). The entropy $H_V(\mathbf{u})$ at a production volume $V(\mathbf{u})$, $\mathbf{u} \in D$, is calculated based on the local rock type proportions p_c for all categories C, as defined in Equation 3.10 (Darcy and Aigner, 1980).

$$H_V(\mathbf{u}) = -\sum_{c=1}^{C}[p_c(\mathbf{u}) \times ln(p_c(\mathbf{u}))] \quad c = 1, ..., C \tag{3.10}$$

Entropy is a good measure of uncertainty for deposit where the mineralization is associated to the rock types and the border of the different rock types are uncertain. Entropy is zero (its minimum value) when there is only one category, and it is maximum when the proportions of all categories is the same (Darcy and Aigner, 1980).

3.4.2 Regression Analysis

Regression analysis is the statistical methodology to explore the relationship between two or more variables when it is known that this relationship exists; it is also used for predicting values of the response (dependent) variable from the predictors (independent) variables (Johnson and Wichern, 2007; Walpole et al., 2012).

When the relationship between the response variable (Y) and the predictor variable (X) is not known theoretically, the choice of the model is based on a scatter diagram and the regression model is said to be empirical (Montgomery and Runger, 2002). Consider the hypothetical scatterplot of Figure 3.18. Although no simple curve passes through all the

points, there is a straight line where points are randomly distributed around it.

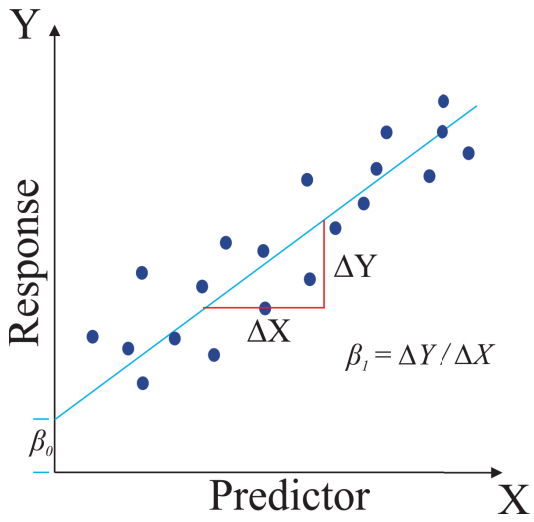

Figure 3.18: Empirical scatter diagram of the linear regression between response (Y) and predictor (X) variables. β_0 is the intercept and β_1 the slope of the line.

It is assumed that the mean of the response variable Y is related to the predictor values x by the linear Equation 3.11. The regression coefficients are defined by the intercept β_0 and the slope β_1 of the line.

$$E(Y \mid x) = \beta_0 + \beta_1 x \tag{3.11}$$

The expected value of Y is a linear function of x, however the actual observed value of y does not fall exactly on the straight line. A probabilistic way to generalize Equation 3.11 is to assume that $E(Y)$ is a linear function of x, but for a fixed value of x the actual value of Y is given by the mean plus a random error ϵ, see Equation 3.12 (Montgomery and Runger, 2002).

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{3.12}$$

The regression coefficients and the random error are estimated from the data. The slope β_1

is interpreted as the change in the mean of Y for a unit change in x; the error ϵ is a random variable that is assumed to be distributed with $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2$.

The calculation of the coefficients is done by the method of Least Squares (Johnson and Wichern, 2007). The coefficients β_0 and β_1 are such that minimizes the sum of the squares of the residuals e. A residual is the error in the fit of the model; calculated by the difference of the observed y values and the fitted $\hat{y}$ values; $e_i = y_i - \hat{y}_i; i = 1, ..., n$. The method of the Least Squares is then the minimization of the sum of squares of the errors (SSE) or the sum of squares of the residuals (SSR), as defined in Equation 3.13:

$$SSE = SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3.13}$$

The analysis of the variance in regression is done by the coefficient of determination R^2. It measures the proportion of the variability explained by the fitted model; in other words, it explains the amount of the variance of the response variable that is explained by the predictors. In a regression model, the unexplained variance is given by the SSE. The portion of the variance that would ideally be explained is calculated by the total correct sum of squares (SST), in Equation 3.14.

$$SST = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2; \quad i = 1, ..., n \tag{3.14}$$

Whereas the term $(y_i - \hat{y}_i)^2$ in Equation 3.13 reveals the variance due to the error of the fitting (hence the variance of the regression model), the term $(y_i - \bar{y}_i)^2$ in Equation 3.14 measures the variance when sampling in a non-regression scenario. The quantity of the variance explained is given by $SST - SSE$. The R^2 is defined in Equation 3.15. If $R^2 = 1$, all residuals are zero and all variability is explained by the predictors.

$$R^2 = 1 - \frac{SSE}{SST} \tag{3.15}$$

An acceptable value of R^2 depends on the application of the regression model and complexity of the data. The coefficient of determination always increases for new predictors, although the change in its value may not be much. A predictor that does not change significantly the value of R^2 should not be considered in the analysis, aiming for a simpler model is preferred. Due to the natural complexity of geological phenomena and the simplicity of the regression analysis, it is not expected a high R^2 when applying linear regression to geological data.

The adjusted R^2 is a modification of the R^2 to account for the number of predictors in the model. R^2 will always increase for more predictors, but the adjusted R^2 increases only if a new predictor improves the model more than would be expected by chance. The adjusted R^2 is always lower than R^2.

3.4.2.1 Simple Linear Regression

The simplest linear regression is given by one response and one predictor (Equation 3.11). Consider the example case from Section 1.3.1; the conditional local mean is the predictor variable and the probability of the grade to be within 15% of the mean is the response variable.

The fitted model is shown in Figure 3.19, with $\beta_0 = 0.207$ and $\beta_1 = 0.276$. A "unit" increase in local mean is associated with 0.276 "unit" of the uncertainty and more precise the grade would be.

If an increase in local mean was associated to an increase in uncertainty (less precision), then β_1 would be negative. Regression models are high bias and low variance models; under repeated sampling the fitted line will likely stay in the same place (low variance), but the average of the models will not capture the true relationship (high bias) (James et al., 2013). A way to check the confidence in regression models is by hypothesis testing. The null hypothesis is when there is no relationship between predictor and response ($\beta_1 = 0$); the alternative hypothesis is when there is a relationship and β_1 is not equal to zero. The null hypothesis is rejected if the confidence interval of calculations does not include zero. The

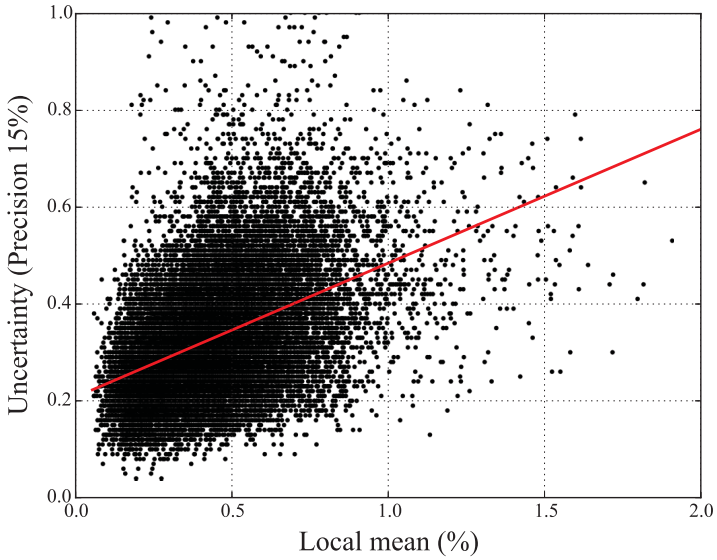

Figure 3.19: Scatterplot of local mean against uncertainty. Each dot represents a SMU. The red line is the fitted model.

statistical *p-value* represents the probability that the coefficient is actually zero. If the 95% confidence interval does not include zero, then *p-value* < 0.05 and there is a relationship between the predictor and response. This is important to define whether linear regression fits or not the model for that predictor.

The simple regression model of the local mean is summarized in Table 3.1. Using only local mean as predictor, the model is able to quantify only 19,6% of the variability in the uncertainty, $R^2 = 0.196$. The other statistics validate the model.

Table 3.1: Simple linear regression summary for the fitted model.

	Coefficient	**Std. error**	**p-value**	**95% Confidence interval**
β_0	0.2079	0.001	0.000	0.206 ; 0.210
Local mean	0.2766	0.002	0.000	0.272 ; 0.281

Response variable: Uncertainty; $R^2 = 0.196$; Adj. $R^2 = 0.196$

A higher portion of the total variability is explained when other predictors are added to the regression model.

3.4.2.2 Multiple Regression

When the complexity of the uncertainty cannot be explained by a sole predictor, a multiple regression model is considered for P number of predictors. A predictor can be discarded when its contribution explained the variability is not significant. A multiple linear regression is given by Equation 3.16.

$$E(Y \mid x_1, ..., x_p) = Y = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p; \quad p = 1, ..., P \tag{3.16}$$

The method of least squares can also be used for estimating the coefficients in linear models that one believes the mean does not fall on a straight line, but on a polynomial equation that fits better the model, see Equation 3.17 below for the case of a single predictor:

$$E(Y \mid x) = Y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_r x^r \tag{3.17}$$

Walpole et al. (2012) explain that a linear model is a model which parameters occur linearly, regardless of how the independent variables enter the model. In some cases, with proper transformation, a non-linear function can be expressed as a straight linear; these models are referred as intrinsically linear (Montgomery and Runger, 2002).

Consider the multiple linear regression with four predictors: local mean, local standard deviation, entropy and drillhole spacing (see Equation 3.18). The R^2 is a statistical measurement from the model and it cannot be splitted between the predictors in a multiple regression. Although the coefficients $\beta_1, ..., \beta_p$ are good indicators of the importance of each predictor when explaining uncertainty, the R^2 cannot be calculated separately. In these cases, the R^2 is used to compare different models.

$$Y = \beta_0 + \beta_1 mean + \beta_2 stdv + \beta_3 DHS + \beta_4 entropy \tag{3.18}$$

The fitted linear regressions for drillhole spacing and local standard deviation are shown in Figure 3.20. For a comparison between the non-parametric fitting with the expected uncertainty curve for drillhole spacing and the linear fitting, see Figure 1.24.

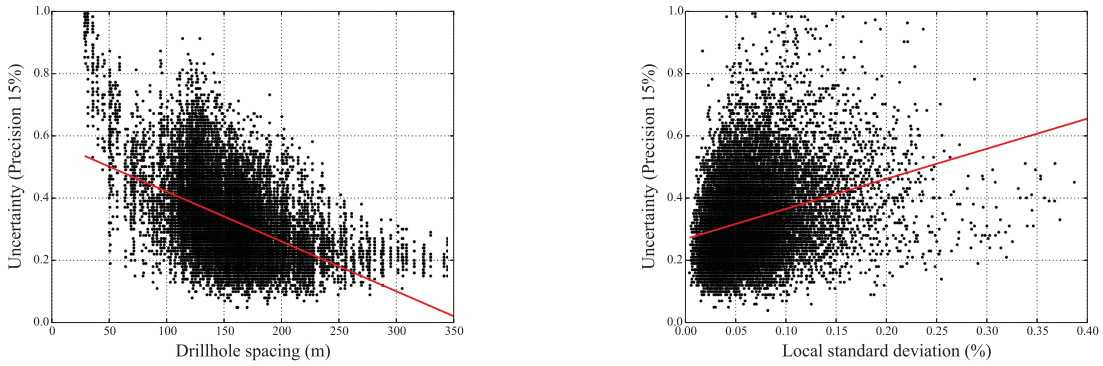

Figure 3.20: The fitted linear regression (red line) for drillhole spacing (on the left) and local standard deviation (on the right).

The multiple linear regression model summary is shown below in Table 3.2. The model with multiple predictors explains more of the uncertainty, $R^2 = 0.43$, when compare to the simple linear model, $R^2 = 0.196$. This model fits better the uncertainty data than the previous model. Local standard deviation, DHS and entropy are negatively associated with uncertainty, whereas local mean is positively associated with uncertainty. The expected uncertainty can be estimated for any value of the predictors, provided that the values used are in the range of the data.

From the coefficients analysis, for a given local mean, standard deviation and entropy, a decrease of 0.0016 m in DHS is associated to an increase of 1% of the precision. The $p - value < 0.05$ for all predictors reject the null hypothesis (that there is no correlation between predictor and response).

The summary of a regression is better visualized in a tornado chart, see Figure 3.21 below (Zagayevskiy and Deutsch, 2011). The vertical bars are scaled by the standardized sensitivity coefficient (numbers in blue); bars are yellow for positive coefficient and green

Table 3.2: Multiple linear regression summary for the fitted model.

	Coefficient	Std. error	p-value	95% Confidence interval
β_0	0.5058	0.002	0.000	0.501 ; 0.510
Local mean	0.2800	0.002	0.000	0.275 ; 0.285
Local stdv	-0.3352	0.013	0.000	-0.360 ; -0.310
DHS	-0.0016	0.000	0.000	-0.002 ; -0.002
Entropy	-0.0647	0.002	0.000	-0.068 ; -0.061

Response variable: Uncertainty; $R^2 = 0.43$; Adj. $R^2 = 0.43$

for negative. The right side of the bars plots the coefficients with the confidence interval (95%), accounting for uncertainty (the small box at the end of the bars). The prediction percentage measures the power of the model to predict a new value. High standardized coefficients are important for model quality and the capacity of explain the total variability (Zagayevskiy and Deutsch, 2011). Based on the standardized coefficients, the local standard deviation has the lowest power when explaining the total variability of the uncertainty, whereas DHS explains the most of the uncertainty.

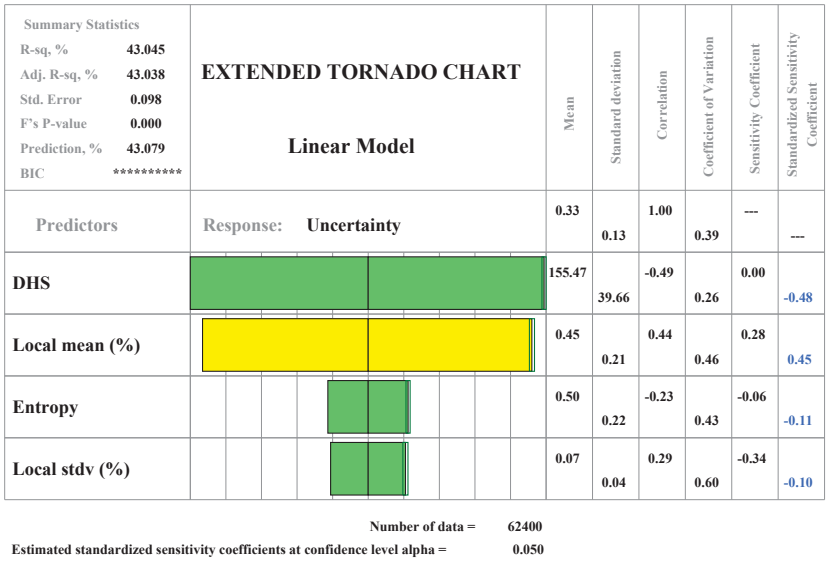

Figure 3.21: Tornado chart of the multiple linear regression summary for the fitted model. The standardized coefficients are plotted. The vertical bars are scaled by the coefficients; yellow for positive coefficient and green for negative.

When the same multiple linear regression analysis is done without considering the local

standard deviation, $R^2 = 0.424$. The small difference between R^2 calculated with this predictor and without it shows that the local standard deviation can be excluded from the model, only for the sake of simplicity. The importance of DHS is shown in a $R^2 = 0.218$ when the model is analysed without this predictor.

Consider the multiple regression model using a quadratic polynomial fitting for predictors and response. The fitted line for DHS is shown in Figure 3.22. The quadratic model appears to fit better the data in comparison to the linear model (Figure 3.20).

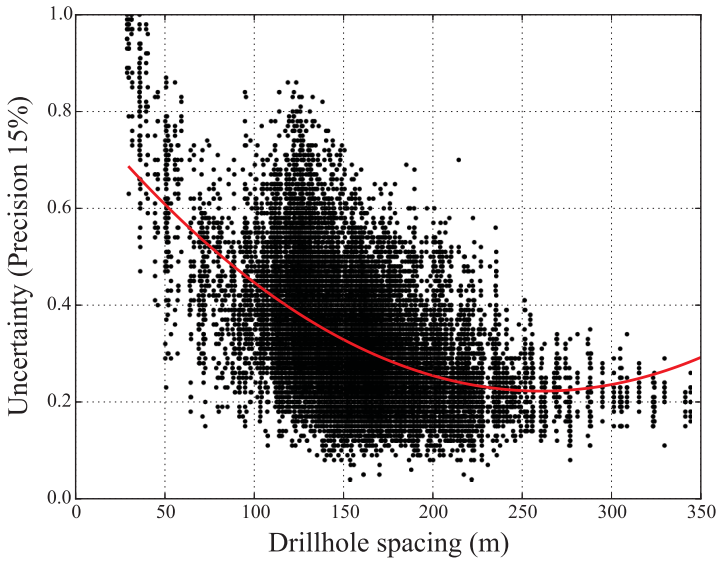

Figure 3.22: Scatterplot of drillhole spacing agains uncertainty with the quadratic polynomial fitted line (in red).

Geological data is often too complex to be explained by linear regression, a higher coefficient of determination is expected when high dimensional polynomial is used to fit the data. Attention is necessary when defining the order of the polynomial used in regression. When the dimensionality increases, the space (volume) of the data increases even faster, resulting in a series of phenomena that do not occur in lower dimensions; this is also referred to "*Curse of Dimensionality*" (Keogh and Mueen, 2010).

When the quadratic model is used, the percentage of the uncertainty explained by the four

predictors is 50.8%, 7.8% more when compared to the multiple linear regression, see the tornado chart of Figure 3.23. The small bars on the left represent the coefficient between predictor-predictor (local mean x local stdv, local mean x entropy, ...) and are colored orange when positive and blue when negative. The drillhole spacing is the most important predictor, followed by local mean, local standard deviation and entropy. Although the entropy has the smallest standardized coefficient, the R^2 calculated from a model without this predictor is of 46.53%.

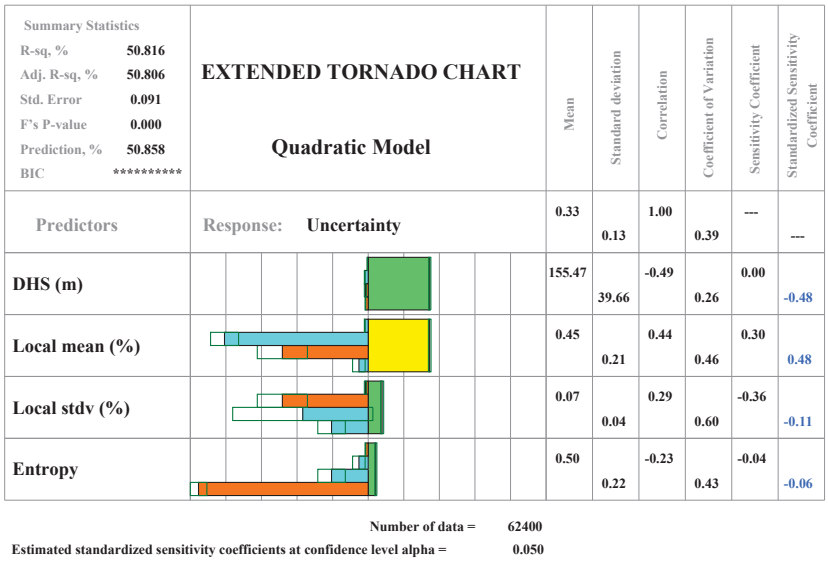

Figure 3.23: Tornado chart of the multiple linear regression summary for the fitted quadratic model. The standardized coefficients are plotted. The vertical bars are scaled by the coefficients; on the right side, yellow for positive coefficient and green for negative. On the left side, orange for positive coefficient and blue for negative.

One of the pitfalls of linear regression is the consideration that the response variable is dependent of the predictors that are considered independent variables. Moreover, the optimal scenario is the one that the error have the same scatter regardless of the value of X (homoskedastic). When this is not the case (heteroskedasticity), the points in a scatter diagram gets more dispersed in some intervals of X values, and the SSE is wrong. Regression analysis, even when high order polynomials are used to fit the data are not able to capture all the complexity of the data.

3.4.3 Alternating Conditional Expectations

Alternating Conditional Expectations (ACE), (Breiman and Friedman, 1985), is a non-parametric regression technique able to capture complex relationships between the response and the predictor variables. Linear regression can yield to erroneous analysis when this relationship is unknown or inexact and forced to be parametrically linear (Wang and Murphy, 2004).

Let Y be the response variable and the predictor variables expressed by $X_1, ..., X_p$. The goal in ACE is to find the transformations in the response and predictors that lead to the best fitting additive model (Breiman and Friedman, 1985). The variables are replaced by functions $\theta(Y)$ and $\phi_1(X_1), ..., \phi_p(X_p)$ with arbitrary measurable mean-zero functions. The general ACE regression model is given by Equation 3.19 below:

$$\theta(Y) = \alpha + \sum_{i=1}^{P} \phi_i(X_i) + \epsilon \tag{3.19}$$

The ACE algorithm estimates those functions that maximizes the correlation of their additive regression and minimizes the variance not explained ϵ^2 making minimal assumptions concerning the data and the form of the solutions (Barnett and Deutsch, 2013b; Breiman and Friedman, 1985; Wang and Murphy, 2004). The transformations are performed until the linear relationship between $\theta(Y)$ and the sum of the transformed predictors $\sum_{i=1}^{P} \phi_i(X_i)$ is maximized. In ACE the correlation between the transformed response and the sum of the transformed predictor variables is the amount of the variance explained. The variance of the transformed predictors is related to the capacity of that predictor to explain the uncertainty. The higher the variance of a transformed predictor is, the more variability of the response it explains.

Consider the ACE model with four predictors: local mean, local standard deviation, entropy and drillhole spacing. In the previous regression models, the amount of the variance explained for a linear and quadratic fitting are 43% and 50.8% respectively. The

predictors are plotted against their transformed function in Figure 3.24. The standard deviation σ_y of the transformed predictors (Y-axis) represents their capacity to explain the uncertainty. DHS ($\sigma = 0.536$) is the most important predictor, followed by the local mean ($\sigma = 0.374$), entropy ($\sigma = 0.228$) and local standard deviation ($\sigma = 0.091$).

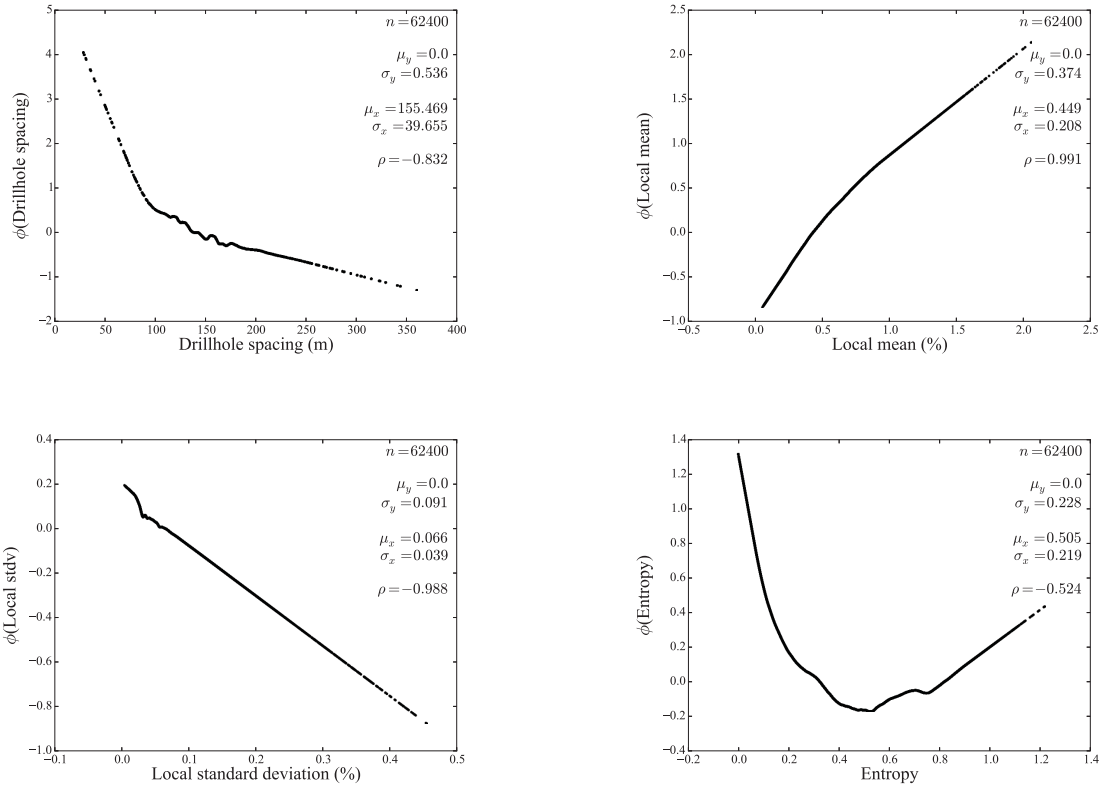

Figure 3.24: Scatterplots of the predictors (X-axis) against their transformed functions (Y-axis). Drillhole spacing is in the upper left corner, local mean is in the upper right corner. Local standard deviation is the bottom left plot and entropy is the bottom right plot.

The total variability explained is given by the correlation of the transformed response variable $\theta(Y)$ and the sum of the transformed predictors $\sum_{i=1}^{P} \phi_i(X_i)$, see Figure 3.25. The ACE model explains 73.5% of the uncertainty; its prediction is greater than both linear regression models (linear and quadratic).

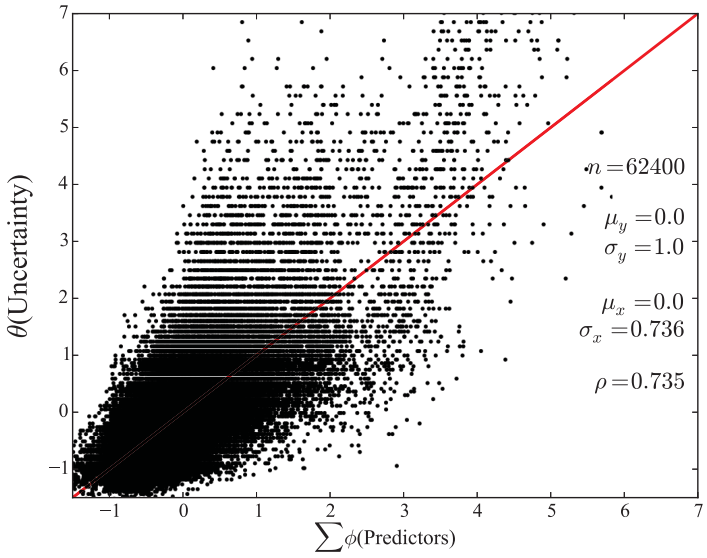

Figure 3.25: Transformed response variable (Y-axis) against the ACE predicted values (X-axis). The capacity of the ACE model to explain total variability is given by the correlation. This model explains 73.5% of the uncertainty.

ACE has the advantage to handle continuous and categorical variables in the same model, not accounting for the form of the transformations and fitting the data in a non-parametric way. This provides ACE the powerful to handle complex features and relationship in the data. One of the disadvantages of using ACE it is the risk of overfitting. The risk of overfitting is proportion to the ratio of number of predictors and the number of samples (Barnett and Deutsch, 2013b). The algorithm of ACE *per si* does not provide a measure of uncertainty when functions are overfitted and unreliable; this uncertainty can be captured by bootstrap, although this is not implemented in the original algorithm (Barnett and Deutsch, 2013a; Breiman and Friedman, 1985). The transformed function is not unique, ACE results depend on the order the predictors are entered into the analysis, although the model correlation does not vary significantly (Wang and Murphy, 2004).

3.4.4 Stepwise Removal of the Factors

Another non-parametric regression analysis is the stepwise removal of the predictors (Draper and Smith, 2014). The order of importance of the predictors must be known, a previous regression model is used as reference. To illustrate this method, consider the previous ACE regression model and the order that predictors are removed given by the standard deviation of their transformed functions: drillhole spacing, local mean, entropy and local standard deviation. The total variability in the uncertainty (response variable) is given by its variance $Var\{Uncertainty\}$.

The first predictor variable removed is the drillhole spacing. The left plot of Figure 3.26 shows the expected uncertainty curve as function of the drillhole spacing, the dots represent the SMUs in the model. The uncertainty in each SMU is subtracted from the EUC, see the right plot for the $\Delta\{Uncertainty \mid DHS\}$. The variance explained by drillhole spacing is given by Equation 3.20 below.

$$\sigma^2(\text{due to DHS}) = \frac{Var\{Uncertainty\} - Var\{Uncertainty \mid DHS\}}{Var\{Uncertainty\}} \tag{3.20}$$

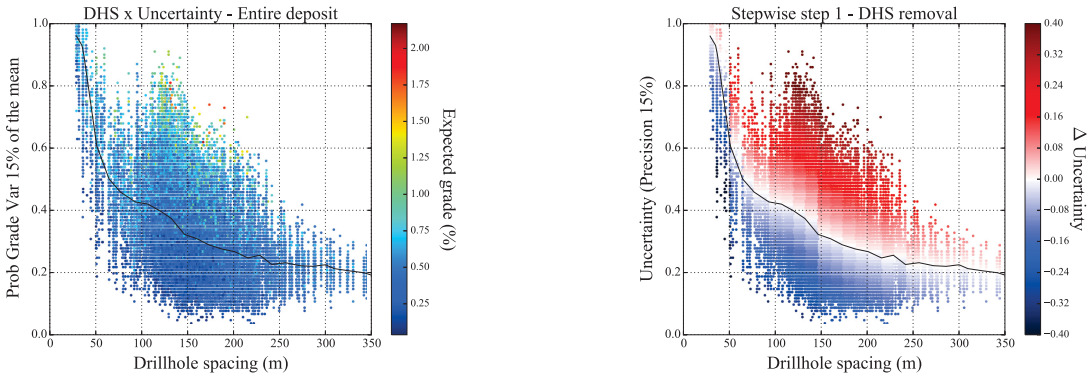

Figure 3.26: Stepwise removal of the predictors, step 1. Drillhole spacing is removed considering the expected uncertainty curve (on the left). The difference between the uncertainty and the expected uncertainty for all SMUs is shown on the right.

In order to remove the local mean, the $\Delta\{Uncertainty \mid DHS\}$ is plotted against the local mean and the expected curve is fitted. The removal of the local mean is done with respect to the fitted regression curve in a similar way to the drillhole spacing. Figure 3.27 shows the steps for local mean and entropy removal. The standard deviation as the last predictor cannot be removed by itself.

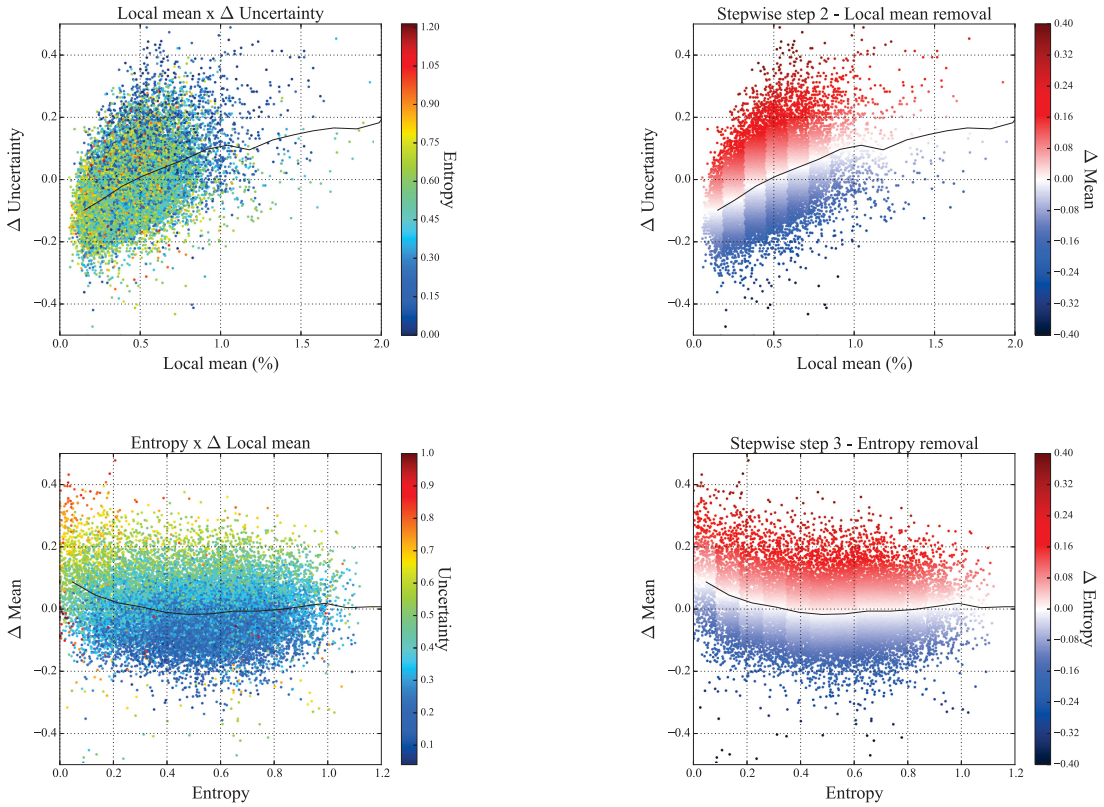

Figure 3.27: Stepwise removal of the local mean and entropy. The regression line must be fitted for each predictor.

This method explains 55.8% of the total variance in the response variable. For each predictor removed, a lesser amount of the uncertainty is explained. The drillhole spacing explains 29.4%, the local mean 21.0% and the entropy explains 5.4%. In this method the amount of the total uncertainty explained by the predictors is known numerically in a non-parametrically way. The disadvantage of using it is the fact that a prior model of regression to determine the order that predictors are removed is needed.

Chapter 4: Case Study Part I - Uncertainty Modelling

The first part of the case study discusses the practical application of geostatistical modelling for uncertainty assessment. Models of uncertainty are constructed at the SMU block scale and at monthly, quarterly and yearly production scales. The drillhole spacing versus uncertainty curves and explanatory models depend strongly on the uncertainty model. Three grade variables are modelled: Variable A (Var A), Variable B (Var B) and Variable C (Var C). The results of the case study are shown in terms of Var A in different lithologies.

This case study is developed using real data provided by a major Canadian mining company for research purposes. The coordinates and units of variables were changed for confidentiality reasons. The workflow and nature of the results are not affected.

4.1 Motivation

To account for uncertainty in the geological model, a stochastic model of the geology is required. The rock types (RTs) are modeled with sequential indicator simulation (SIS) (Goovaerts, 1997). Assessing uncertainty with sequential Gaussian simulation (SGS) requires the bivariate distribution to be multivariate Gaussian after the univariate transform of the variable (normal scores transformation). When the multivariate distributions are non-Gaussian after such transformation the complex relationship between variables must be handled prior simulation with geostatistical multivariate techniques such as Projection Pursuit Multivariate Transform (PPMT) and Minimum/Maximum Autocorrelation Factors (MAF) (Barnett et al., 2014). Lastly, each realization of the grade is run with a different realization of the rock type.

Uncertainty at different production scales is calculated after back-transformation of the simulated values.

4.1.1 Workflow Description

The workflow for uncertainty modelling is summarized as follows:

1. Define the stationary domains (the rock types);
2. Data analysis and declustering of the rock types;
3. Model the indicator variograms;
4. Model the rock types with SIS;
5. Model the normal scores variograms for data imputation;
6. Data imputation and declustering of continuous variables;
7. PPMT and MAF transformations;
8. Model the PPMT variograms;
9. Simulate the grade with SGS and back-transform the simulated values;
10. Assess uncertainty at different production scales;

The geological model generated with SIS is used to transfer the uncertainty in the geological model to the grade model. Moreover, this model allows the analysis of entropy as one explanatory factor. Data imputation is required due to the unequal sampling of the variables. PPMT and MAF are used to decorrelate the variables; after such transformations the variables are simulated independently. The simulated model is averaged up from the data scale to the SMU scale; uncertainty is then calculated at a SMU resolution within the different production volumes.

4.2 Data

The data provided for this case study consists of 75,980 samples, collected from 764 near vertical drillholes. Three variables, unequally sampled over the deposit, are modelled from

a total of 54,640 samples of Var A, 54,085 of Var B, and 54,855 of Var C. The drillholes are located over a volume of 20,800 m East, 14,000 m North and 3,520 m Elevation as shown in Figure 4.1.

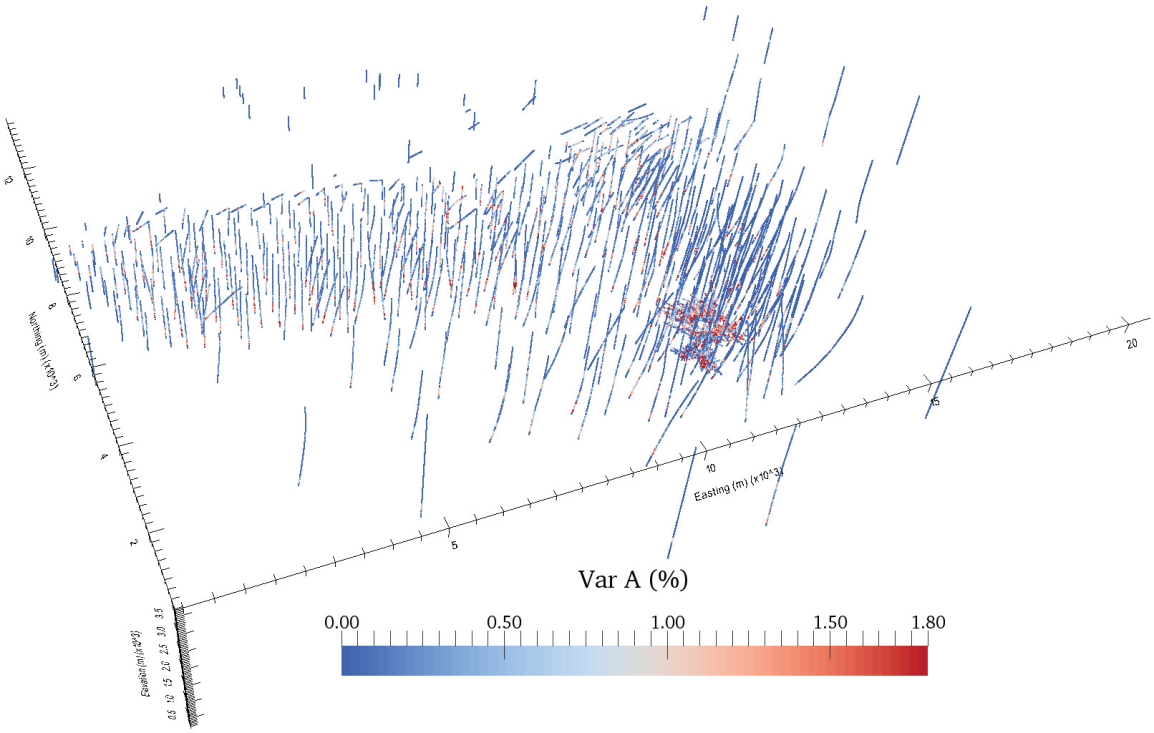

Figure 4.1: Oblique view of the drillholes over the deposit, colored by the grade of the Variable A. The grade of 1.8 % represents the P95 of the grade distribution.

The length of the drillholes is greater in the south of the deposit and higher grades of Var A are observed towards the bottom of the deposit. The clustered distributions of the variables are shown in Figure 4.2. A positively skewed histogram is observed for all variables, variables A and B contain a large amount of low grade samples.

4.2.1 Stationarity and Domaining

The stationary domains were chosen based on first and second order statistics, spatial proximity and similar geological features. The decision of stationarity was reviewed through the workflow, specially in the variography steps, in order to ensure geological and

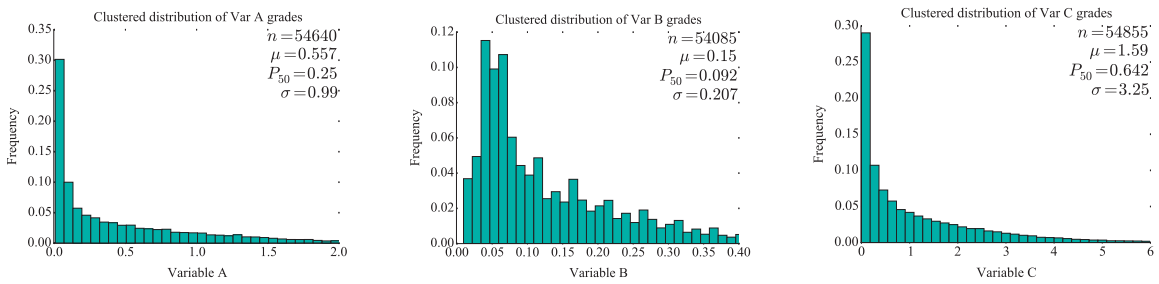

Figure 4.2: Distribution of variables A, B and C with no declustering weights. A positively skewed histogram is observed for all variables.

grade estimates consistently. Seven domains were defined from the original 18 rock types composing the data. The rock types are labeled 200, 211, 311, 321, 400, 401 and 500, as shown in oblique view of the deposit in Figure 4.3 and side-view in Figure 4.4.

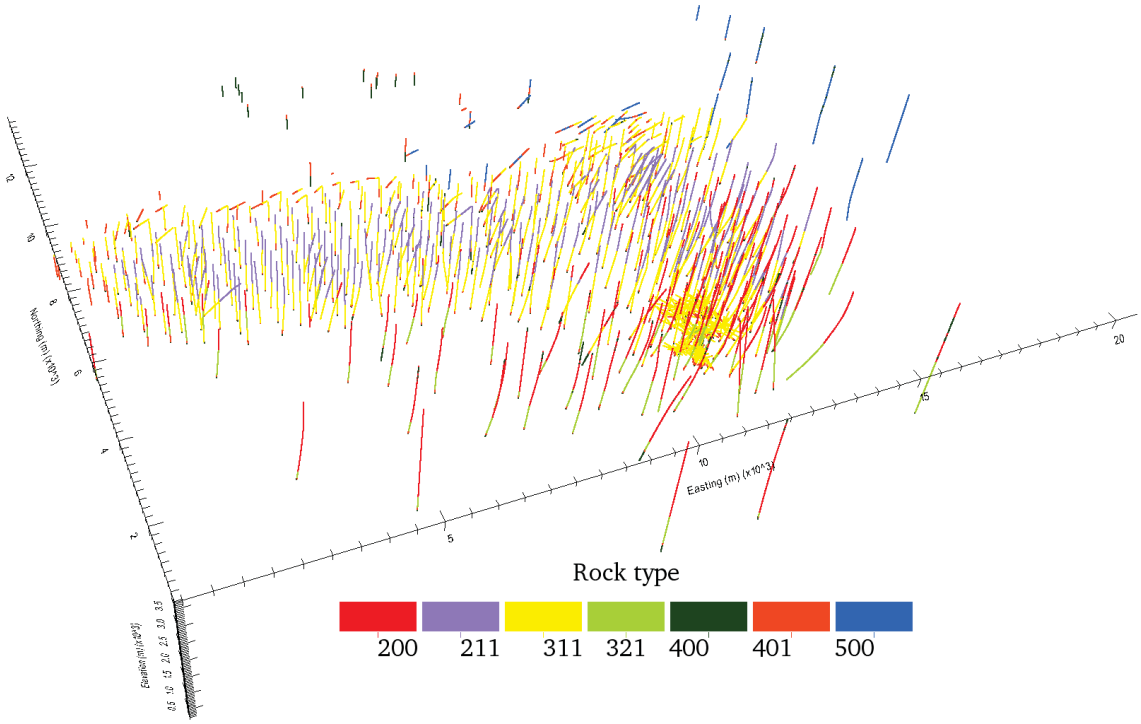

Figure 4.3: Oblique view of the drillholes over the deposit, colored by the lithology.

Due to the irregular drillhole spacing and the large volume covered by drilling, a convex hull (Figure 4.5) is applied to limit simulation; which decreases run time, optimizes storage,

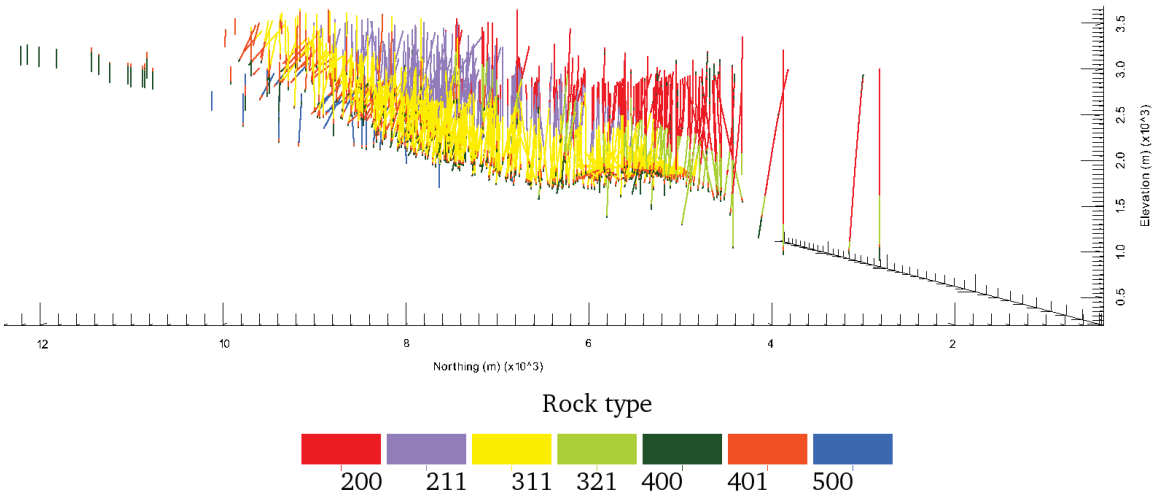

Figure 4.4: Side-view of the drillholes over the deposit, colored by the lithology.

and limit estimates inside a relevant volume.

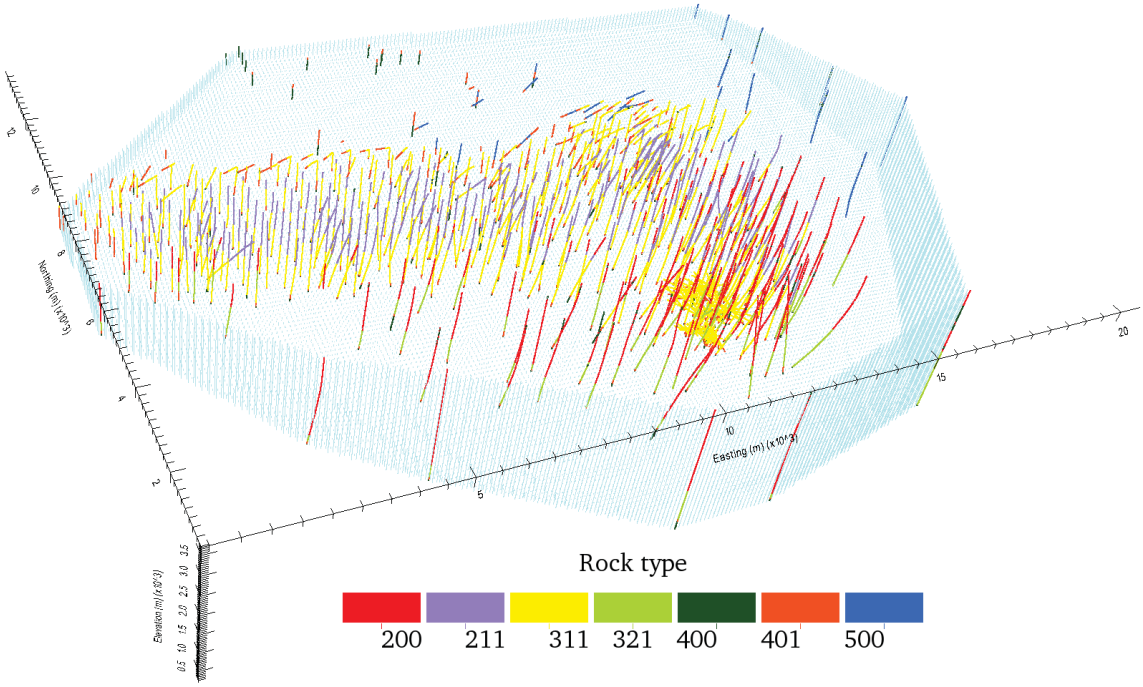

Figure 4.5: Convex hull (in light blue) applied to the drillhole locations in order to simplify the model.

4.2.2 Basic Statistics

The data have been averaged to twenty meters fixed length composites which represents the SMU height. Original data was provided at 5-12 m intervals. The averaged drillhole spacing over the deposit is 1,630 m and the average drillhole density is 0.02 $drills/ha$. Declustering of the continuous variables is done after imputation.

4.2.2.1 Categorical Variables

Although the drillhole spacing is nearly regular in the center of the deposit, areas of large spacing exist and declustering is required. The cell size for declustering depends on the rock type and is based on the average drillhole spacing in sparsely sampled areas. The cell sizes used are: 2,500 x 2,500 m for RT 200; 1,000 x 1,000 m for RTS 211 and 311; 2,000 x 2,000 m for RTs 321, 400 and 401; and a cell size of 2,000 x 2,000 m for RT 500. The clustered and declustered rock type proportions are shown in Table 4.1. RTs 211 and 311 (see Figure 4.3) have the smallest and most regular drillhole spacing and have their proportions lowered after declustering.

Table 4.1: Clustered and declustered rock type proportions.

Rock Type	Clustered Proportions	Declustered Proportions
200	0.161	0.231
211	0.194	0.077
311	0.449	0.201
321	0.053	0.102
400	0.037	0.141
401	0.081	0.106
500	0.023	0.142

4.2.2.2 Continuous Variables

The distributions of the grade assays in the rock types are shown in Table 4.2. The highest grades of Var A and Var B are observed in RTs 311 and 321, whereas the highest grades of Var C are in RTs 311, 321 and 401.

Table 4.2: Basic statistics of the variables in the deposit and per domain.

Rock Type	Variable	Samples	Min (%)	Max (%)	Mean (%)	Stdv (%)
All	A	54640	0.01	50.00	0.56	0.99
	B	54085	0.01	6.24	0.15	0.21
	C	54855	0.01	72.74	1.59	3.25
200	A	5962	0.01	2.45	0.19	0.31
	B	5964	0.01	0.85	0.09	0.07
	C	6028	0.01	21.95	0.43	0.82
211	A	11503	0.01	3.20	0.20	0.31
	B	11484	0.01	0.60	0.09	0.07
	C	11556	0.01	15.80	0.37	0.60
311	A	30692	0.01	50.00	0.80	1.20
	B	30358	0.01	6.24	0.19	0.25
	C	30730	0.01	72.74	2.20	3.93
321	A	2783	0.01	3.90	0.55	0.50
	B	2736	0.01	2.07	0.15	0.12
	C	2759	0.02	25.42	1.52	1.68
400	A	355	0.01	1.91	0.19	0.27
	B	383	0.01	0.38	0.08	0.06
	C	473	0.01	5.93	0.42	0.61
401	A	2633	0.01	14.53	0.28	0.91
	B	2484	0.01	4.49	0.08	0.24
	C	2601	0.05	46.02	3.03	4.21
500	A	712	0.01	1.85	0.23	0.23
	B	676	0.01	0.60	0.09	0.07
	C	708	0.02	8.62	0.86	1.05

Variables A and B are highly correlated in all rock types, see Table 4.3 for the correlation in original units of the variables in all domains, prior imputation. The correlations of Var A x Var C and Var B x Var C are low only in RT 500. The scatterplots in Figure 4.6 show the correlation between Var A and the other variables in RT 211 before data imputation. A compositional constraint is seen in both plots. This constraint is also observed in other rock types and should be reproduced by the simulation.

Table 4.3: The correlation between variables in original units per rock type prior to imputation.

Rock Type	ρ_{AB}	ρ_{AC}	ρ_{BC}
200	0.877	0.632	0.597
211	0.839	0.782	0.668
311	0.743	0.686	0.842
321	0.892	0.540	0.585
400	0.851	0.683	0.649
401	0.815	0.529	0.607
500	0.880	0.388	0.370

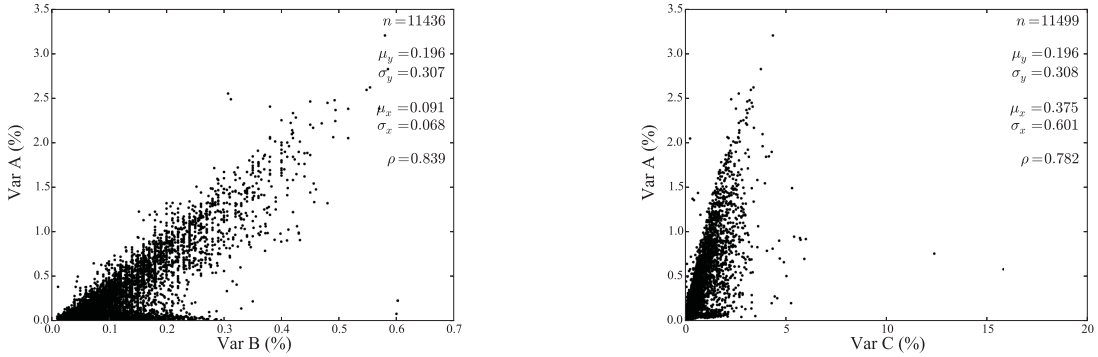

Figure 4.6: Correlation between Var A x Var B (on the left) and Var A x Var C (on the right) in RT 211 prior imputation.

4.3 Rock Type Modelling

A three dimensional trend is modelled by moving window averaging, see Figure 4.7 for the vertical trend representation. The values in the windows are weighted by their distance from the moving center using a Gaussian kernel to produce a smooth trend. The local declustered proportions calculated from the trend are used in the simple kriging equations when the residuals are kriged in SIS (Deutsch, 2005; Journel, 1983).

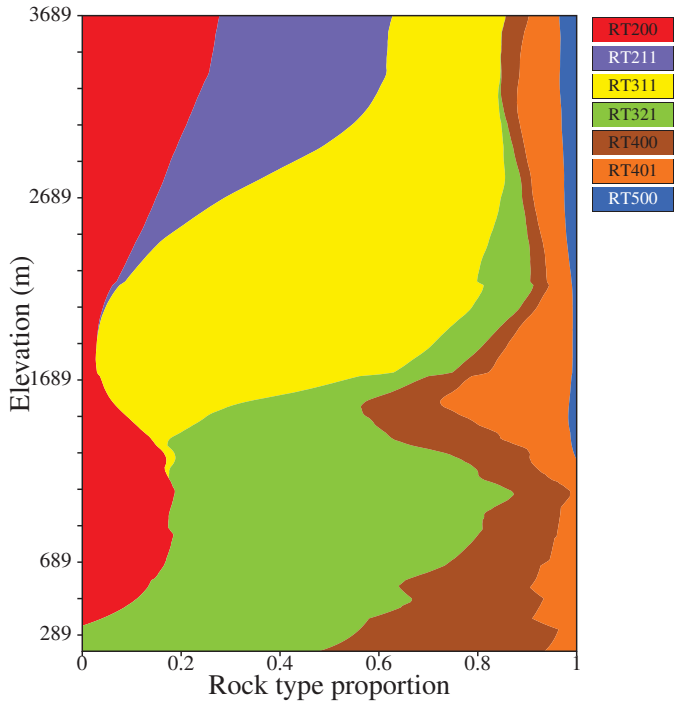

Figure 4.7: Vertical rock type proportions. The 3D trend is calibrated to the declustered proportions.

Indicator variograms were modelled with three directions whenever horizontal anisotropy was observed (2 horizontal and the vertical directions). All variograms were modelled with zero nugget effect and three spherical structures. An azimuth of 0 degrees is considered for the minimum horizontal direction of continuity, and 90 degrees azimuth for the maximum horizontal direction of continuity. Table 4.4 summarizes the indicator variogram models. The variogram models for RTs 211 and 321 are shown in Figure 4.8.

Table 4.4: Indicator variograms summary. Horizontal minor and major directions of continuity are shown.

RT	Minor H.	Major H.	Comments
200	-	-	Omnidirectional in hor. direction with trend in horizontal and vertical
211	N0°E	N90°E	Normal anisotropy with a trend in vertical direction
311	N0°E	N90°E	Zonal anisotropy with a trend in vertical direction
321	-	-	Omnidirectional in hor. direction with trend in horizontal and vertical
400	-	-	Omnidirectional in hor. direction with trend in horizontal direction
401	N0°E	N90°E	Weak anisotropy with high continuity in vertical direction
500	N0°E	N90°E	Normal anisotropy with high continuity in vertical direction

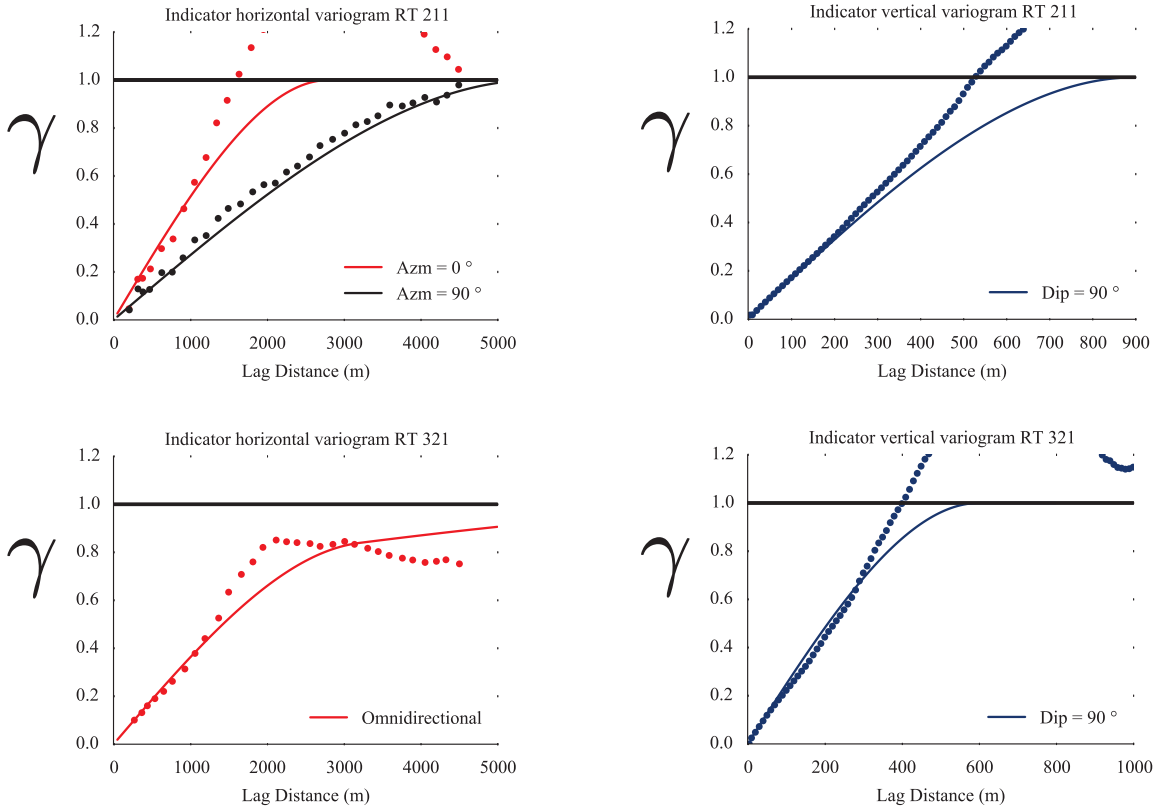

Figure 4.8: Indicator variogram models for RTs 211 (at the top) and 321 (at the bottom). The horizontal variograms are shown on the left and the vertical on the right. RT 211 shows geometric anisotropy in the horizontal direction, whereas RT 321 was modelled with an omnidirectional variogram in the horizontal.

One hundred realizations of the geology were generated with SIS, respecting the trend model, the variograms and the declustered proportions. The P50 of the realizations, at the SMU scale, is shown for RTs 211 and 321 in Figure 4.9.

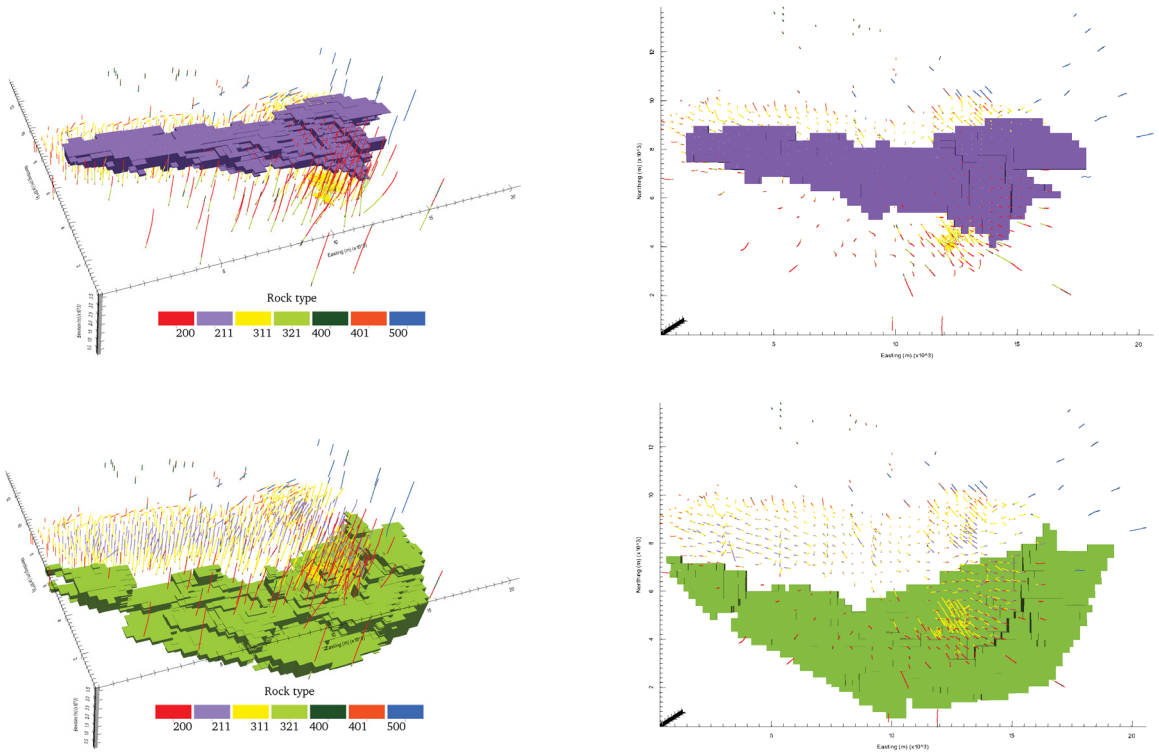

Figure 4.9: Oblique and plan view of the P50 based on global proportions in the final realization of RTs 211 (at the top) and 321 (at the bottom). Simulated SMUs and the drillholes are shown.

4.4 Multivariate Grade Modelling

Gaussian co-simulation would be considered if the variables were multi-Gaussian after normal scores transformations. The three most important sources of non-Gaussianity are nonlinearity, heteroscedasticity and constraints (Leuangthong and Deutsch, 2003). The scatterplots of the bivariate distributions, see Figure 4.6 for RT 211, show constraints in all rock types. In order to check the bivariate distributions, the data are transformed to

Gaussian units and the bivariate distributions are compared to a perfect multivariate normal distribution. The averaged deviation between data points in these two distributions can be measured for different density contours intervals and inform about the Gaussianity of the variables (Deutsch and Deutsch, 2011).

Strong non-Gaussianity is observed in RTs 200, 211, 311, and 321, see Figure 4.10 for the kernel normal bivariate distributions of Var A x Var B and Var A x Var C in rock type 211. RTs 400, 401 and 500 show weak non-Gaussianity. The lack of Gaussianity motivates the use of PPMT to decorrelate variables prior simulation (Barnett et al., 2014). The proper back transformation of the simulated values reintroduces the complexity between the variables.

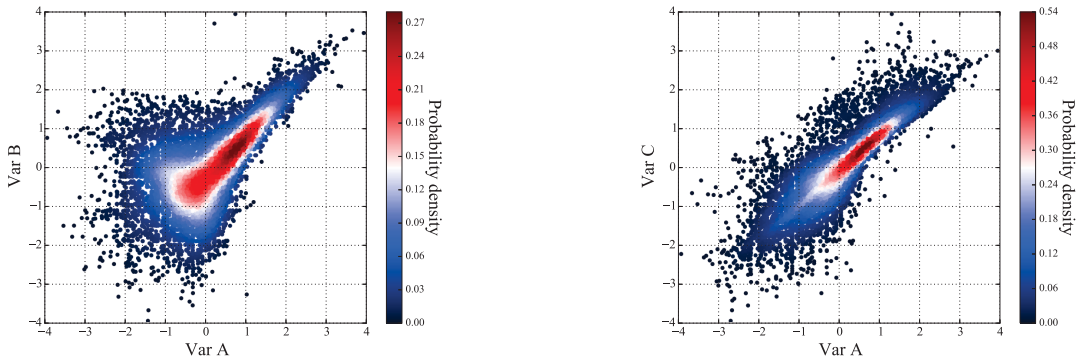

Figure 4.10: Normal scores bivariate distributions of Var A x Var B (on the left) and Var A x Var C (on the right) in RT 211, colored by the bivariate kernels density estimators. The complex distributions should be reproduced after simulation and special data transformations are required to decorrelate variables prior simulation. Only homotopic samples are used for plotting.

The detailed multivariate grade modelling workflow is summarized as follows:

1. Model the normal scores variograms for data imputation;
2. Data imputation with Gaussian mixture models (GMM) (Silva and Deutsch, 2015);
3. Declustering weights;
4. PPMT and MAF transformations;

5. Model the Variograms;
6. Independet grade simulation with SGS and back-transformation of the simulated values;

4.4.1 Data Imputation

The PPMT transformation can only be applied to homotopic data, that is, the variables must be equally sampled at data locations. Data imputation was performed with GMM and non-parametric Bayesian updating (Silva and Deutsch, 2015). The normal scores variograms are needed for data imputation, see Figure 4.11 for the normal scores variograms for Var B in RT 211 and Var A in RT 321. The total number of data imputed in each rock type is summarized in Table 4.5.

Table 4.5: Number of data imputed in each rock type.

Rock Type	Var A	Var B	Var C
200	66	64	-
211	57	76	4
311	56	390	18
321	8	55	32
400	119	91	1
401	10	159	41
500	-	36	4

4.4.1.1 Gaussian Mixture Models and Imputation

The joint probability distributions defining the likelihood distributions are calculated from Gaussian mixture fitted to the multivariate data. The likelihood distribution are calculated as the marginal of the conditional Gaussian mixture models. The prior and likelihood distributions are combined with Bayesian updating; the updated distribution is sampled to generate multiple realizations of the missing data (Silva and Deutsch, 2015).

The number of multivariate Gaussian components used to fit the conditional distributions varies from 2 to 10, according to the rock types. The density plots of Figure 4.12 show the

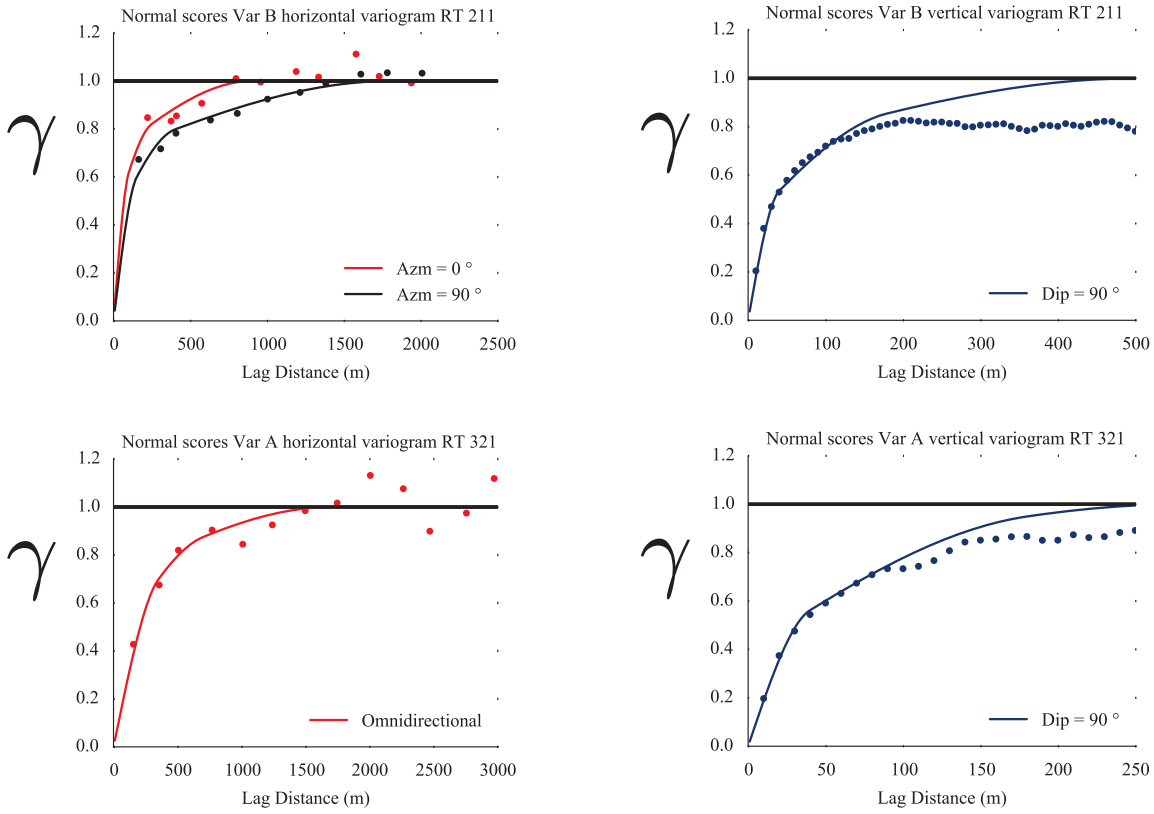

Figure 4.11: Normal scores variogram models used in data imputation for Var B in RT 211 (at the top) and for Var A in RT 321 (at the bottom). The horizontal variograms are shown on the left and the vertical on the right. The variogram of Var B shows geometric anisotropy in the horizontal direction, whereas Var A was modelled with an omnidirectional variogram.

marginal bivariate distributions of Var A x Var B, and Var A and Var C in rock type 211 fitted with 8 components. A threshold of 1% of the joint probability density function (PDF) is used in the plot. Although the joint distribution populates the entire range of normal values, a better visualization is achieved plotting 99% of the density values.

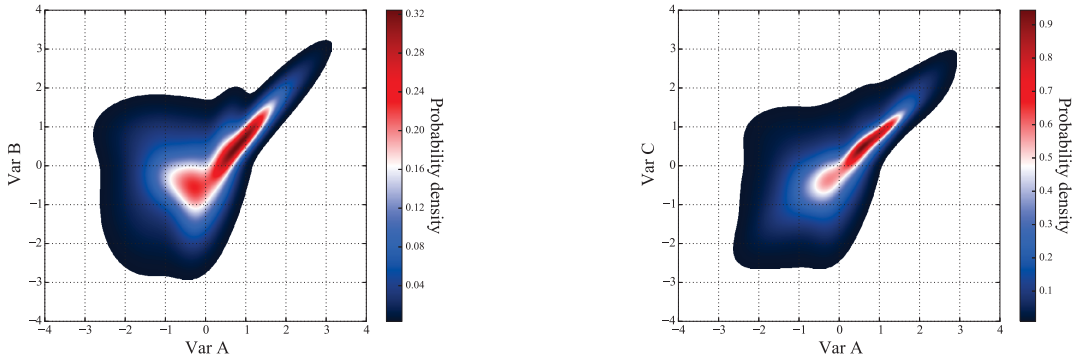

Figure 4.12: Marginal bivariate distributions of the GMM fitted for Var A x Var B (on the left) and Var A x Var C (on the right) in RT 211 colored by the bivariate kernels density estimators. The visualized values represent 99% of the joint PDF.

The imputed data in RT 211 is plotted over the bivariate distribution in Figure 4.13. This check is performed in all rock types.

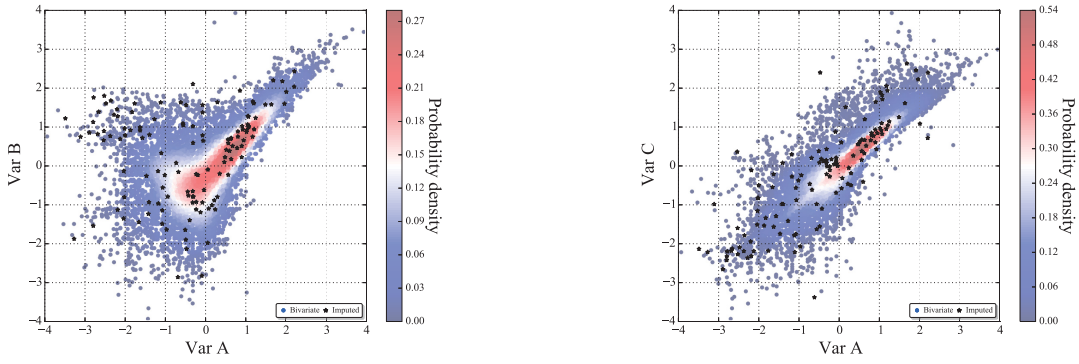

Figure 4.13: Imputed data (black dots) plotted over the normal scores bivariate distributions (colored dots) in RT 211.

4.4.1.2 Declustering

Declustering weights are calculated after imputation with similar window sizes used for declustering of the categorical variables. The bivariate distributions are recalculated after data imputation. The declustered histograms are calculated and summarized in Table 4.6. The correlation coefficients after imputation are shown in Table 4.7. The change in the histograms and correlations depends on the number of data imputed and their values.

Table 4.6: Declustered statistics after imputation of the variables per domain.

Rock Type	Variable	Samples	Min (%)	Max (%)	Mean (%)	Stdv (%)
200	A	6028	0.01	2.45	0.20	0.29
	B	6028	0.01	0.85	0.09	0.07
	C	6028	0.01	21.95	0.46	0.89
211	A	11560	0.01	3.20	0.20	0.31
	B	11560	0.01	0.60	0.09	0.07
	C	11560	0.01	15.80	0.37	0.60
311	A	30748	0.01	50.00	0.80	1.20
	B	30748	0.01	6.24	0.19	0.25
	C	30748	0.01	72.74	2.20	3.93
321	A	2791	0.01	3.90	0.54	0.48
	B	2791	0.01	2.07	0.16	0.16
	C	2791	0.02	25.42	1.47	1.63
400	A	474	0.01	1.91	0.19	0.26
	B	474	0.01	0.38	0.08	0.07
	C	474	0.01	5.93	0.62	0.86
401	A	2643	0.01	14.53	0.20	0.58
	B	2643	0.01	4.49	0.07	0.16
	C	2643	0.05	46.02	2.37	3.38
500	A	712	0.01	1.85	0.27	0.25
	B	712	0.01	0.60	0.11	0.08
	C	712	0.02	8.62	0.87	0.97

4.4.2 PPMT and MAF

PPMT and MAF are multivariate techniques used to model complex and high dimensional geologic data, transforming the data to an uncorrelated multiGaussian distribution (Barnett et al., 2014; Friedman, 1987). PPMT decorrelates the variable at a zero lag

Table 4.7: The correlation between variables in original units per rock type after imputation.

Rock Type	ρ_{AB}	ρ_{AC}	ρ_{BC}
200	0.876	0.633	0.594
211	0.835	0.782	0.668
311	0.744	0.686	0.841
321	0.893	0.540	0.590
400	0.853	0.704	0.589
401	0.815	0.522	0.598
500	0.880	0.392	0.351

covariance, whereas MAF decorrelates at a non-zero lag; the back-transformation reintroduces the complexity of the data (Barnett, 2011). See Figure 4.14 for the correlation at zero lag covariance for variables A and B in RT 211 after PPMT. Compared the bivariate plot of Var A and Var B from Figure 4.14 and Figure 4.10 (before PPMT).

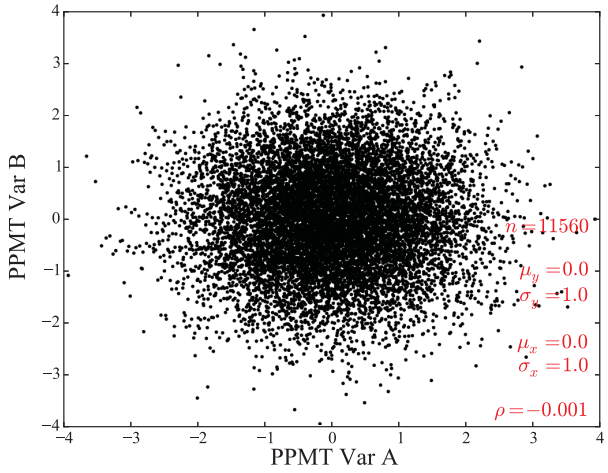

Figure 4.14: Correlation between Var A and Var B after PPMT transformation at zero lag covariance in RT 211.

The PPMT cross variograms were checked in all rock types and MAF was used whenever a correlation greater than 0.2 was observed for any lag. Short distances were preferred to avoid consideration of trend-like features. In Figure 4.15, the omnidirectional cross variogram between the PPMT transforms of Var A and Var C is shown for before and after

MAF. The variables are decorrelated at lag zero by PPMT, but correlation remains for larger lag distances. In most cases, MAF decorrelates the variables at all lags, although in the presence of a trend, correlation may remain (Manchuk and Deutsch, 2015). In Figure 4.15 MAF is performed with a lag distance of 80 m with lag tolerance of 50 m. Although correlation is still observed for some lags after MAF, the short ranges are decorrelated or have their correlation decreased.

MAF was considered in RTs 400, 401 and 500.

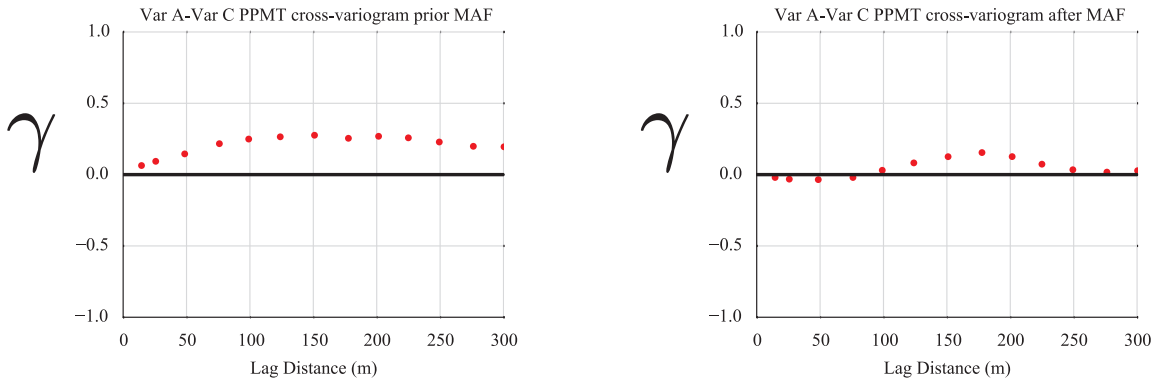

Figure 4.15: Cross variograms for Var A and Var C prior MAF (on the left) and after MAF (on the right) in RT 401. A lag distance of 80 m and a lag tolerance of 50 m are used for decorrelation.

4.4.2.1 Variograms

The variograms were modelled after PPMT and MAF, see the summary in Table 4.8. All variograms were modelled with three spherical structures and no nugget effect. Horizontal anisotropy is observed for Var A and Var B in RT 200, and for Var B and Var C in RT 211; a horizontal omnidirectional variogram is considered otherwise. See Figure 4.16 for the variogram models of Var B in RT 211 and for Var A in RT 321. These variograms are similar to the normal score variograms; in fact, some recent research (Manchuk and Deutsch, 2015) has shown that the normal scores variograms are preferred in this workflow.

Table 4.8: Continuous variograms summary. Horizontal minor and major directions of continuity is shown. Down-hole variograms are calculated with DIP of 90° since drillholes are mostly vertical. Variograms are calculated after PPMT and MAF.

RT	Var	Minor H.	Major H.	Comments
200	A	N0°E	N90°E	Normal anisotropy in hor. dir.
	B	N0°E	N90°E	Normal anisotropy in hor. dir.
	C	-	-	Omnidirectional in hor. dir.
211	A	-	-	Omnidirectional in hor. dir. with trend in vertical dir.
	B	N0°E	N90°E	Normal anisotropy in hor. dir.
	C	N0°E	N90°E	Normal anisotropy in hor. dir. with trend in vertical
311	A	-	-	Omnidirectional in hor. dir. with trend in vertical
	B	-	-	Omnidirectional in hor. dir. and high continuity in vertical
	C	-	-	Omnidirectional and high continuity in hor. dir.
321	A	-	-	Omnidirectional in hor. direction
	B	-	-	Omnidirectional in hor. direction
	C	-	-	Omnidirectional in hor. direction
400	A	-	-	Omnidirectional in hor. dir. and high continuity in vertical
	B	-	-	Omnidirectional in hor. direction
	C	-	-	Omnidirectional in hor. dir. and high continuity in vertical
401	A	-	-	Omnidirectional and high continuity in hor. direction
	B	-	-	Omnidirectional in hor. dir. and high continuity in vertical
	C	-	-	Omnidirectional in hor. dir. and high continuity in vertical
500	A	-	-	Omnidirectional in hor. dir. and high continuity in vertical
	B	-	-	Omnidirectional in hor. dir. and high continuity in vertical
	C	-	-	Omnidirectional in hor. direction

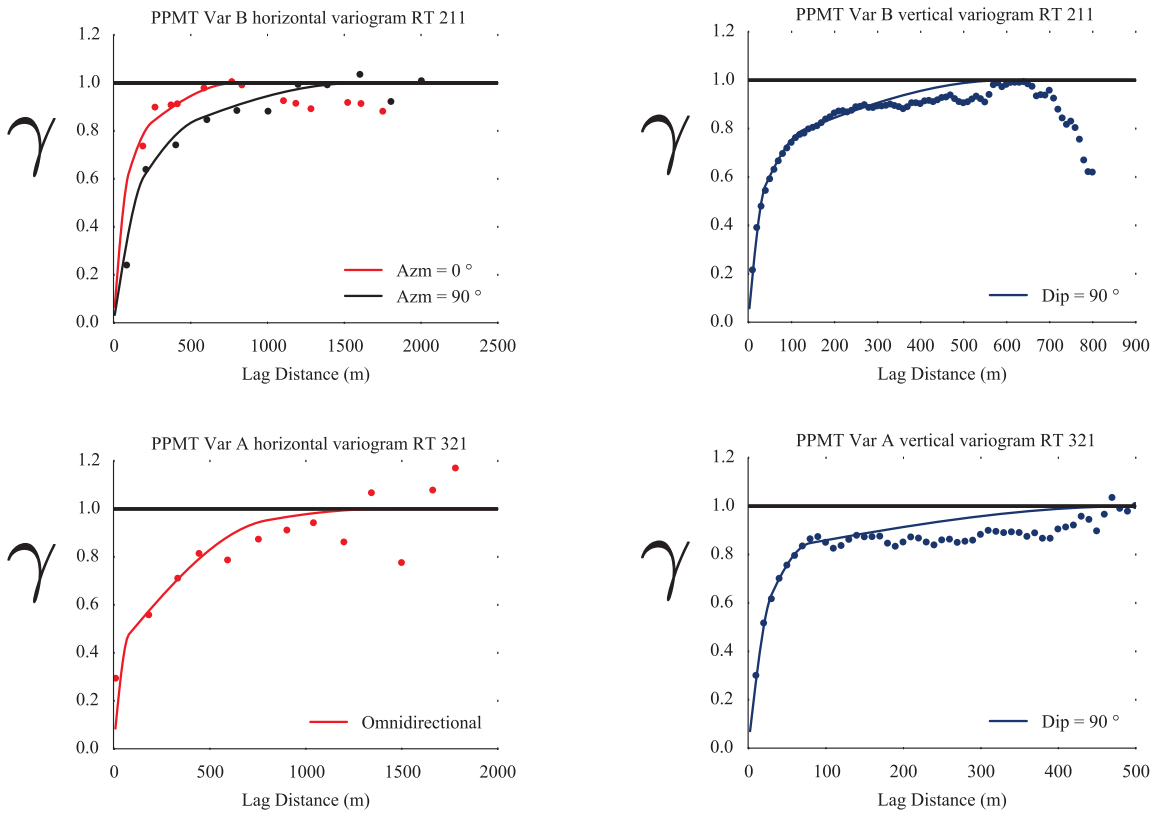

Figure 4.16: PPMT variogram models for Var B in RT 211 (at the top) and for Var A in RT 321 (at the bottom). The horizontal variograms are shown on the left and the vertical on the right. The variogram of Var B shows geometric anisotropy in the horizontal direction, whereas Var A was modelled with an omnidirectional variogram.

4.4.3 Conditional Simulation

The transformed variables were simulated independently within the rock types, with the modelled PPMT variograms. All bivariate distributions were checked against the reference distributions for any discrepancy. In Figure 4.17, the reference normal scores bivariate distributions, calculated after data imputation, are checked against the simulated distributions in RT 211. The complexity seen in the data is reasonably reproduced.

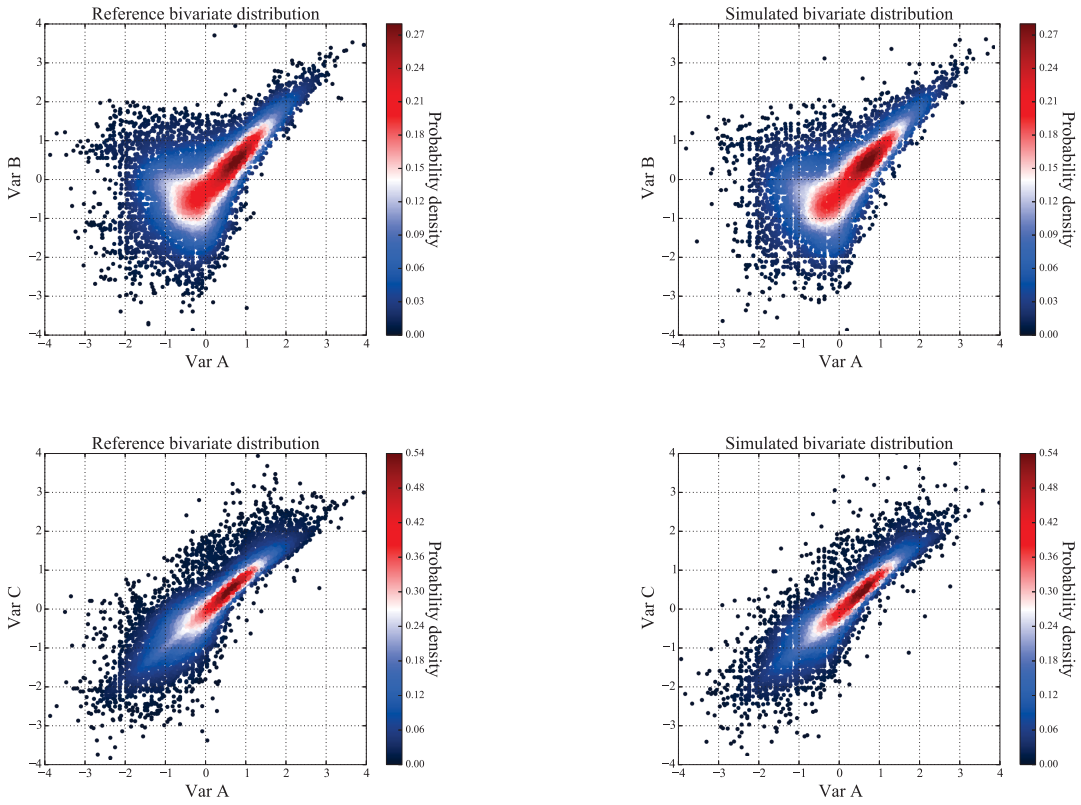

Figure 4.17: Normal scores bivariate distribution of Var A x Var B after imputation in the upper-left corner, and the simulated distribution in the upper-right corner in RT 211, colored by the bivariate kernels density estimators. The reference and simulated distributions for Var A and Var C in the same RT are shown at the bottom.

4.4.3.1 Model Checking

Uncertainty is assessed and post-processed in original units. The simulated variables are back-transformed to original units following the reverse order of the workflow transformations. The histogram reproduction is checked for all variables in all rock types, see Figure 4.18 for the histograms of Var A in RT 211 and Var B in RT 321. Good histogram reproduction is obtained in RTs 211, 321 and 500. The simulated histograms are affected by trends in the continuous and categorical models, declustering weights, geological model and rock type boundaries. The reference declustered histograms mean, the simulated histograms and their relative differences are shown in Table 4.9 for all rock types.

The correlation between variables is also checked. The proposed workflow reproduced the complexity and the constraints observed in data in Gaussian and original units. As reference, the correlation between variables calculated after data imputation (the reference) and after simulation, in RT 211, are plotted in Figure 4.19.

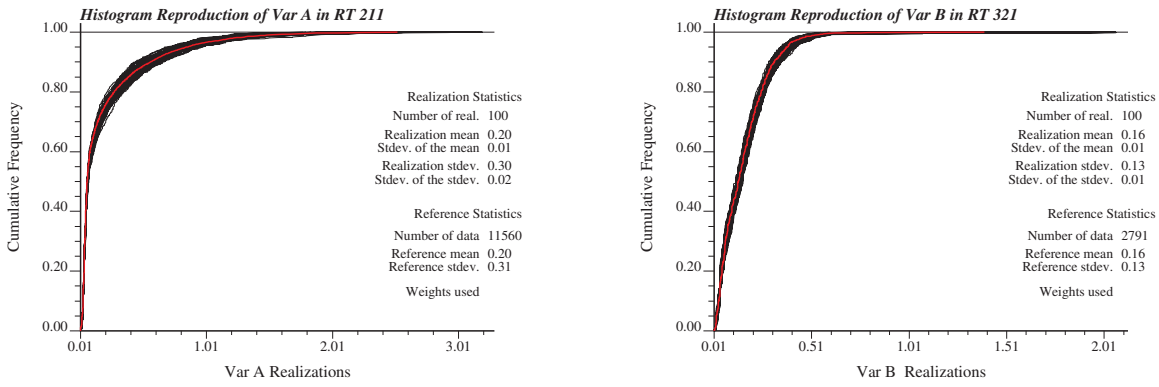

Figure 4.18: Histogram simulation of Var A in RT 211 (on the left) and Var B in RT 321 (on the right). The red line is the reference declustered distribution. The 100 simulated CDFs are shown as black lines.

Table 4.9: Histogram summary statistics. The reference declustered histogram mean is compared with the simulated histogram mean.

Rock Type	Variable	Reference mean	Simulated mean	Difference (%)
200	A	0.20	0.18	- 10.0
	B	0.09	0.08	- 11.1
	C	0.46	0.40	- 13.0
211	A	0.20	0.20	-
	B	0.09	0.09	-
	C	0.37	0.37	-
311	A	0.80	0.75	- 6.2
	B	0.19	0.18	- 5.2
	C	2.20	1.96	- 10.9
321	A	0.54	0.57	+ 5.5
	B	0.16	0.16	-
	C	1.47	1.47	-
400	A	0.19	0.21	+ 10.5
	B	0.08	0.07	- 12.5
	C	0.62	0.64	+ 3.2
401	A	0.20	0.17	- 15.0
	B	0.07	0.07	-
	C	2.37	2.47	+ 4.2
500	A	0.27	0.27	-
	B	0.11	0.10	- 9.0
	C	0.87	0.87	-

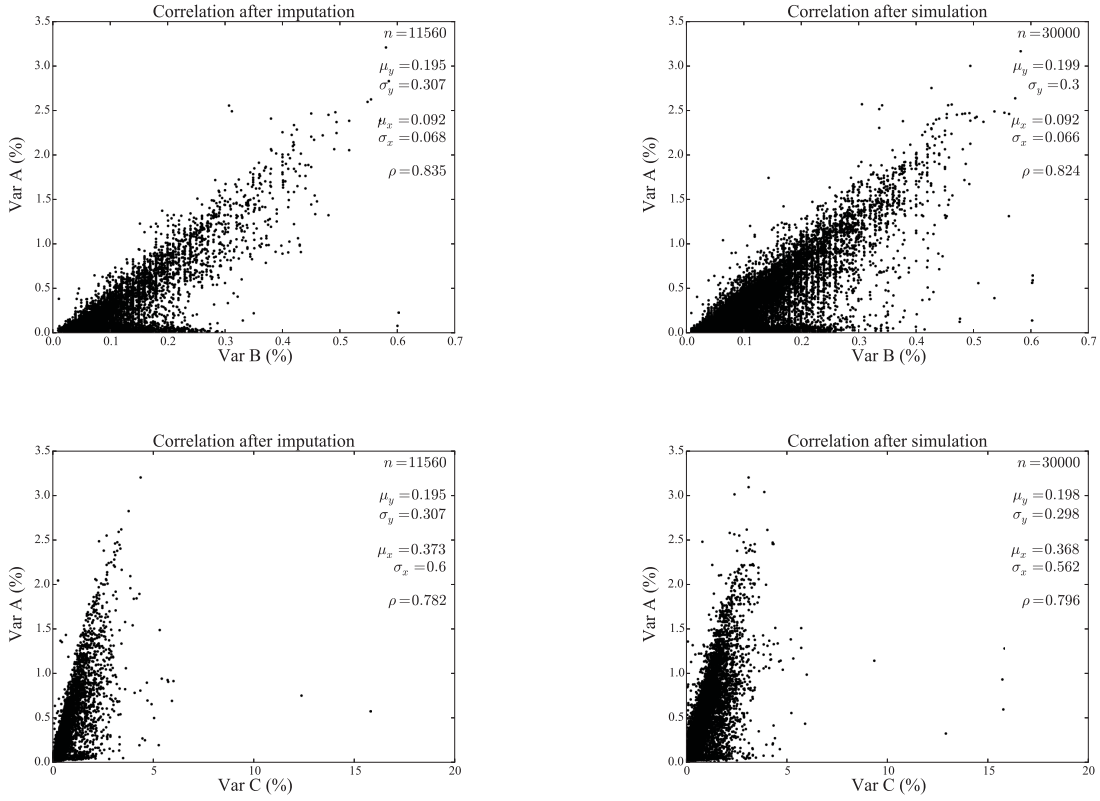

Figure 4.19: Correlation between Var A x Var B after imputation in the upper-left corner, and the correlation after simulation in the upper-right corner in RT 211. The reference and simulated correlation for Var A and Var C in the same RT are shown at the bottom. The simulated distribution is sampled to generate a subset of 30,000 values used in the plots.

4.5 Modelling Scale

After back-transformation, the simulated values are block-averaged to the SMU scale. 9 simulated values are averaged to compose a SMU. There are 8 SMUs in nominal monthly volumes, 24 SMUs in nominal quarterly volumes and 96 SMUs in nominal yearly production volumes. The drillhole spacing study takes into consideration these scales.

The uncertainty at the production volume is calculated with a zero cut-off and fixed volume approach (Section 1.2.3). Uncertainty is reported in terms of the grade of Var A. The probability of the simulated grade of Var A to be within 15% of the mean is calculated for different production scales in the entire deposit and within rock types. The precision (15%) is plotted for a slice of the deposit in Figure 4.20.

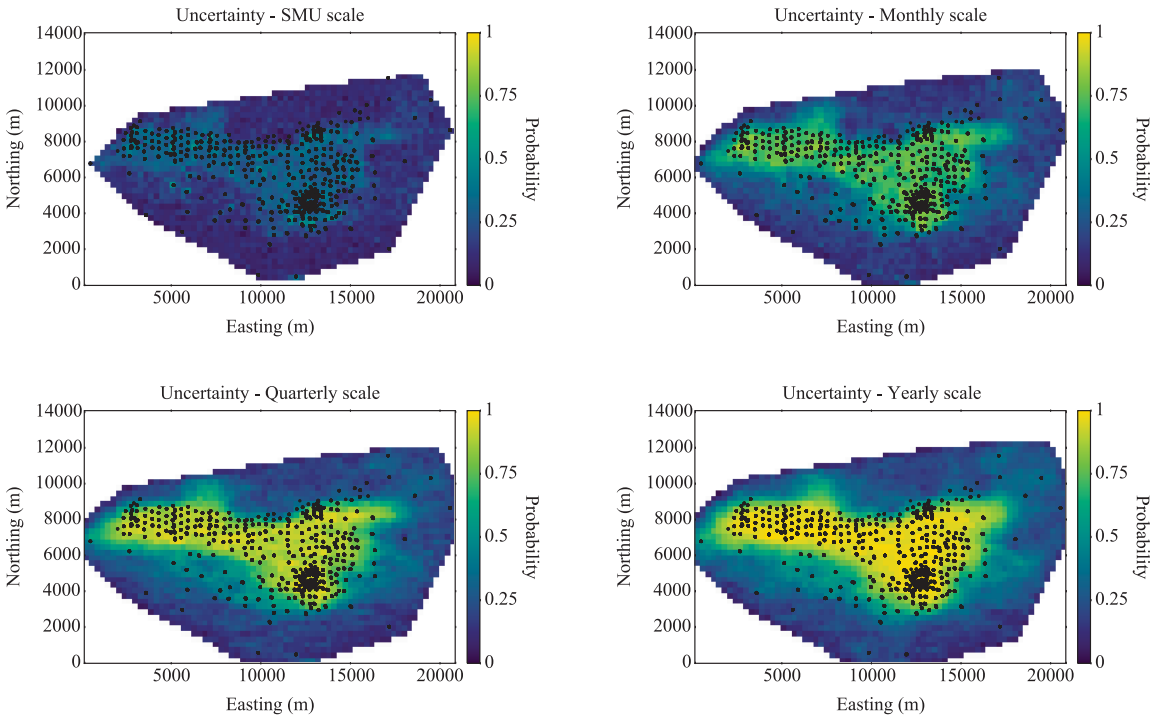

Figure 4.20: The probability of the simulated grade of Var A to be within 15% of the mean for different production volumes in the entire deposit. The dots represent the drillholes, the white part of the plot is outside the convex hull. The plan view of a slice of the deposit is shown.

4.6 Comments

The chosen workflow for uncertainty modelling demonstrated the capacity of PPMT to model complex multivariate relationships. PPMT is considered when the multivariate distributions are not Gaussian. After decorrelation, the variables are simulated independently and the complexity is reintroduced with back-transformation. An alternative workflow for highly correlated variables would be more conventional Gaussian co-simulation. In this specific case study, modelling three variables would require the independent simulation of one of the variables. A second variable would be co-simulated with the previously simulated variable with an intrinsic collocated cokriging as model of coregionalization to avoid variance inflation. The third variable would be co-simulated with the super secondary of the two previously simulated variables. This workflow would not necessarily reproduce the complex relationships between variables and would have change the uncertainty assessment.

Chapter 5: Case Study Part II - Explanatory Factors and Data Spacing

The second part of the case study discusses the drillhole spacing study, the explanatory factors and presents a series of plots to support DHS decisions. The explanatory factors are calculated at a monthly scale for the entire deposit and within rock types. The uncertainty versus drillhole spacing curves and explanatory models of three rock types are shown with more detail. These are the most important rock types when considering the zoning of the deposit. The measure of uncertainty, and the response variable, is the probability of the grade of Var A to be within 15% of the mean.

Economic analysis based on a cut-off and the equivalent grade is done to support zoning of the deposit. When targeting drillhole spacing, the quarterly and yearly uncertainty curves are used for resources classification. The Learning Curve is calculated for the entire deposit and practical recommendations are given.

5.1 Explanatory Factors

The explanatory factors are calculated based on a monthly production scale for the entire deposit. The multiple linear regression and ACE models are shown. The response variable (the uncertainty) is plotted for a slice of the deposit in Figure 5.1.

5.1.1 Factors

All factors are calculated inside the convex hull of the data. Predictor variables are plotted against the response variable for a better understanding of their influence on uncertainty. Unless specified, all drillhole versus uncertainty curves are calculated at a monthly production scale.

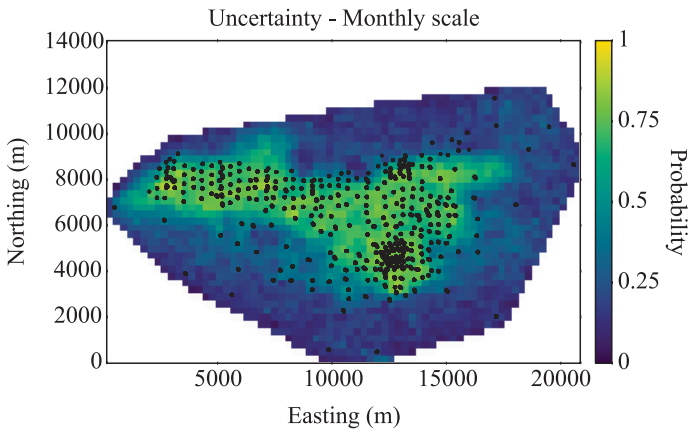

Figure 5.1: The plan view of the uncertainty at a monthly production scale for a slice of the deposit. The probability of the grade of Var A to be within 15% of the mean is the response variable. The black dots represent the drillholes.

5.1.1.1 Drillhole Spacing

The average drillhole spacing is 1,630 m. Uncertainty versus drillhole spacing plots are shown for DHS below 1,500 m. The expected uncertainty is very high and does not change considerably for DHS greater than 1,500 m. There are 162,110 SMUs within this spacing with an average DHS of 761 m. The DHS inside the convex hull and below 1,500 m are shown in Figure 5.2.

The uncertainty versus drillhole spacing curve for the entire deposit is shown in Figure 5.3. The uncertainty shows little change for DHS greater than 1,000 m. The EUC is steeper in the DHS interval of 100 to 1,000 m. Due to the large amount of SMUs plotted, information such as the interval of DHS that contains more SMUs is not clearly represented. Another representation of the same plot is given by the kernel density, in Figure 5.4. SMU density is greater in the DHS intervals of 300-400 m and 900-1400 m.

5.1.1.2 Local Mean and Standard Deviation

The conditional local mean and standard deviation are shown in Figure 5.5. The local standard deviation is high in local mean areas and in the transitions zones. The conditional

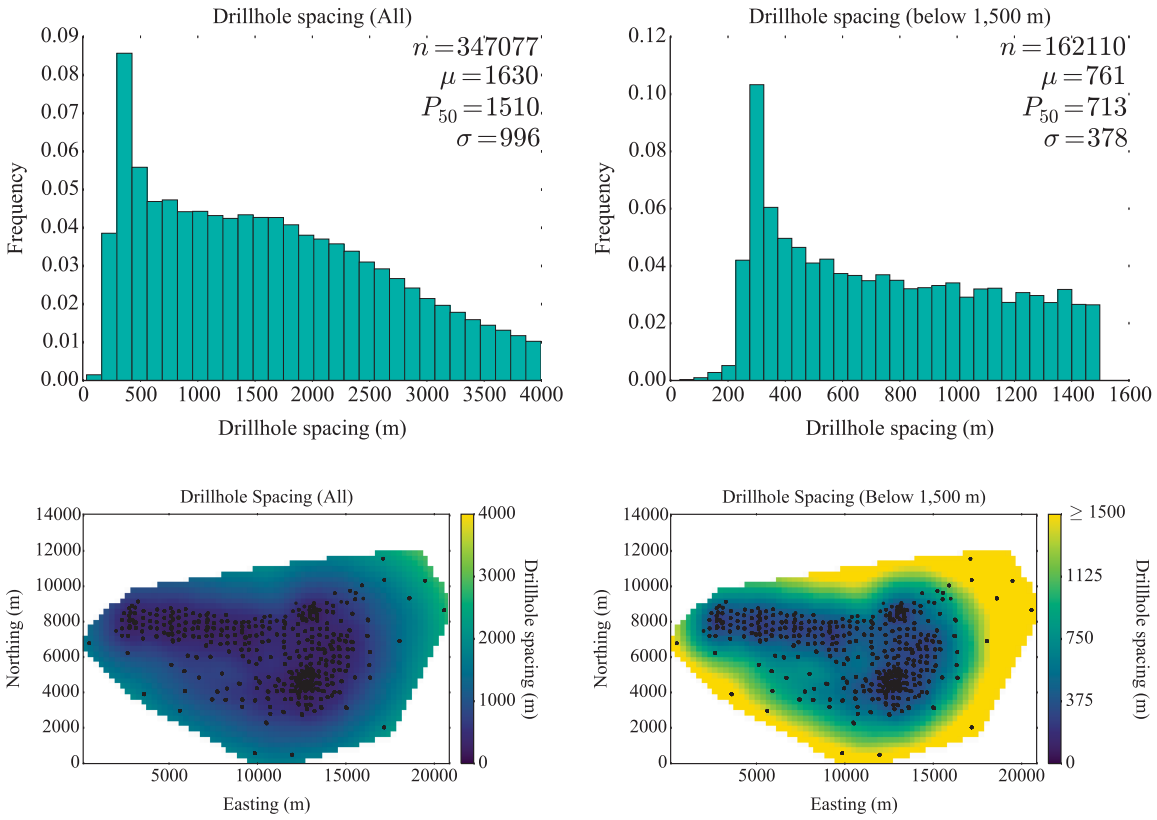

Figure 5.2: The histograms of the DHS inside the convex hull (on the left) and within a spacing of 1,500 m (on the right). The areal coverage of the two plan view plots are the same, but the scale is different to highlight the area covered by the practical maximum DHS considered. The drillhole versus uncertainty curves are plotted with a maximum DHS of 1,500 m.

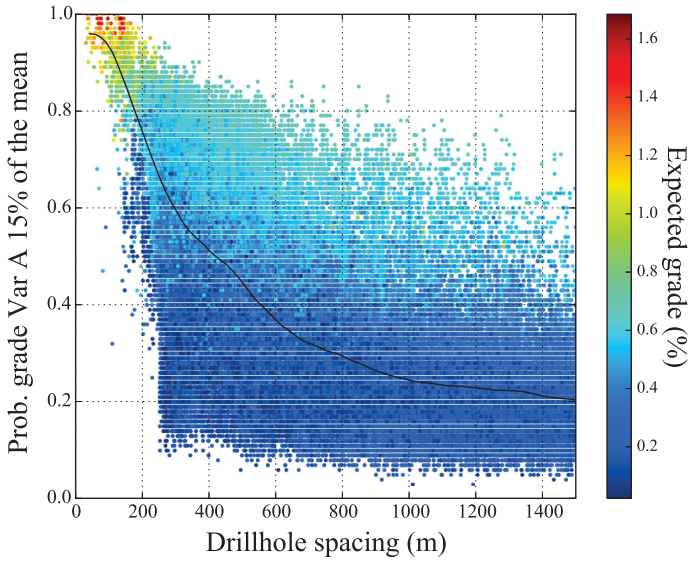

Figure 5.3: Uncertainty versus DHS curve at a monthly production scale for the entire deposit. The dots represent the SMUs, colored by the expected grade. The continuous line is the expected uncertainty curve (EUC).

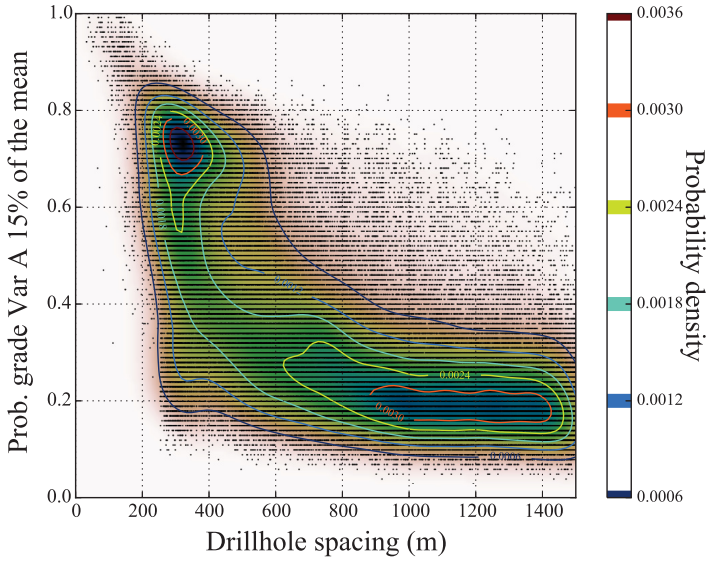

Figure 5.4: Uncertainty versus DHS scatterplot at a monthly production scale for the entire deposit. The small dots represent the SMUs. The kernel density is calculated and the iso-probability density contours are plotted.

proportional effect is shown in Figure 5.6. High variance SMUs are more commonly associated to high grade SMUs, moreover, high grade is related to small DHS. The conditional mean and standard deviation are plotted against the uncertainty in Figure 5.7. Uncertainty is lower for high values of the mean and standard deviation, although the uncertainty does not change considerably for standard deviation greater than 0.25%.

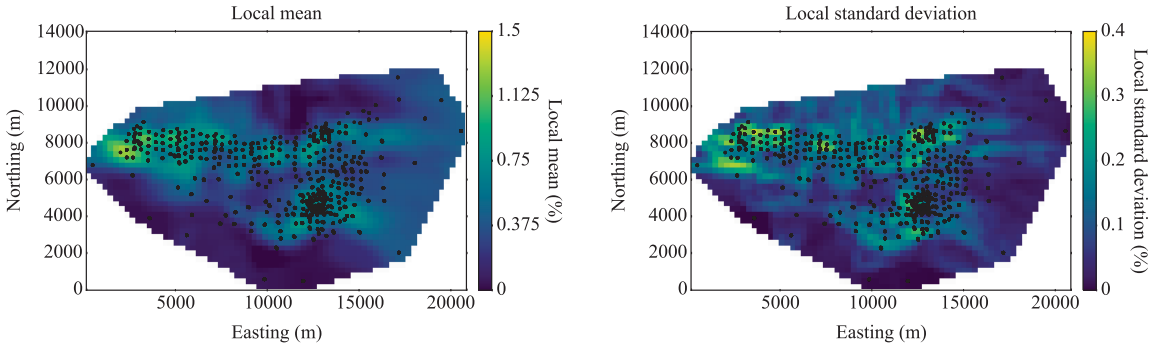

Figure 5.5: The plan view of the local mean (on the left) and standard deviation (on the right) for a slice of the deposit. The local standard deviation is high in local mean areas and in the transitions zones.

5.1.1.3 Entropy

The entropy and the rock type model are shown in Figure 5.8. Entropy is zero inside a domain and higher at the contact between different rock types. The entropy shows litte correlation with the uncertainty, as seen in Figure 5.9. The fact of uncertainty being slightly lower when entropy increases is due to the geological model. Some arbitrary volumes in the contact of rock types may have SMUs from high grade rock types; the local entropy and local grade are then high. Arbitrary volumes with high grade are more certain (Figure 5.7), what explains the relationship between entropy and uncertainty.

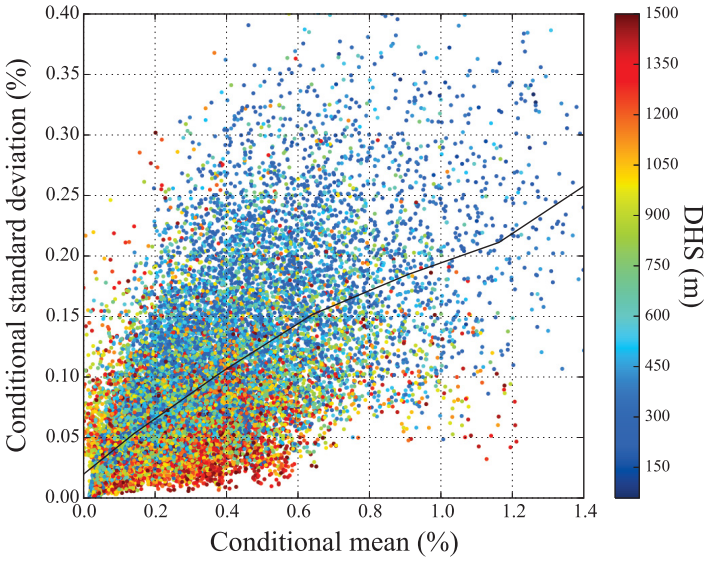

Figure 5.6: The conditional proportional effect in the entire deposit. High variance SMUs are related to high grade SMUs and small DHS. The non-parametric regression curve is shown as the continuous line.

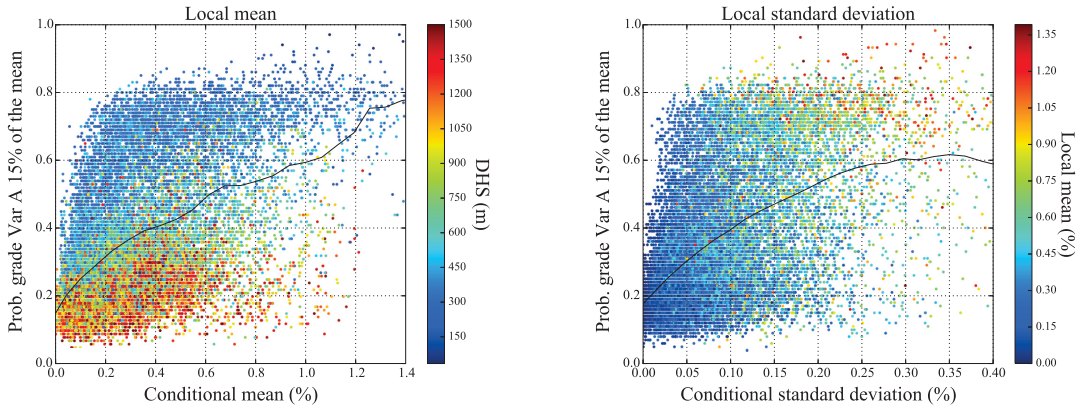

Figure 5.7: Conditional local mean (on the left) and conditional standard deviation (on the right) versus uncertainty. The non-parametric regression curve is shown as the continuous line.

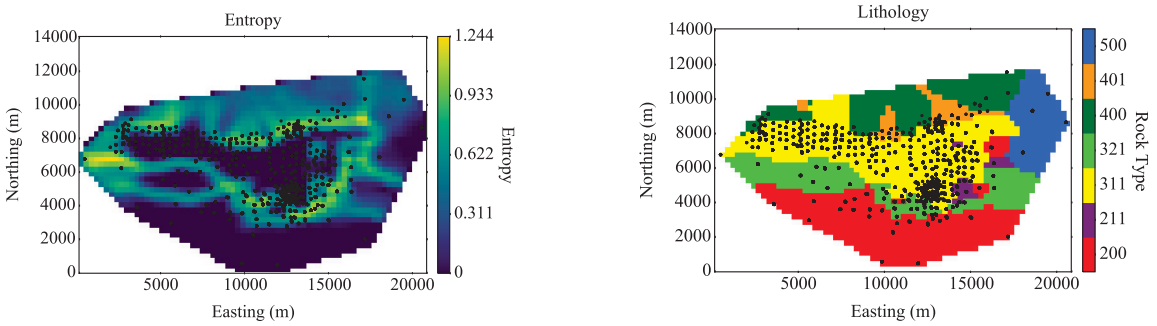

Figure 5.8: The plan view of the entropy (on the left) and rock type model (on the right) for a slice of the deposit. Entropy is higher in the contact of different rock types.

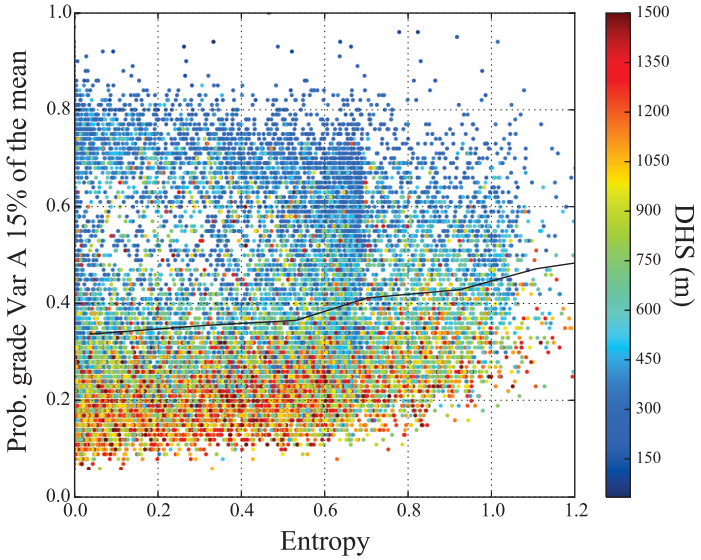

Figure 5.9: Entropy versus uncertainty curve. Entropy shows little influence on the uncertainty. The non-parametric regression curve is shown as the continuous line.

5.1.2 Regression Models

Multiple linear regression and ACE are considered to explain the relationship between the uncertainty and the explanatory factors. ACE explains more of the variability and is considered for supporting decisions.

5.1.2.1 Multiple Linear Regression

The summary of the multiple linear regression for the linear and quadratic fittings are shown in Figure 5.10. The linear model explains 59.3% of the uncertainty, whereas the quadratic model explains 68.5% of the variability in the response variable. DHS and local mean are the most important factors. Entropy and local standard deviation have little impact in the models.

Figure 5.10: Multiple linear regression summary for a linear fitting (on the left) and a quadratic fitting (on the right). The linear and quadratic models explain 59.3% and 68.5% of the uncertainty respectively. DHS and local mean are the most important factors.

5.1.2.2 ACE

The ACE transformations are plotted against the predictors in Figure 5.11. The most important predictors from ACE, with their respective transformed standard deviations are DHS (0.641), local mean (0.353), entropy (0.040) and local standard deviation (0.014).

The low influence of the local standard deviation and entropy on uncertainty is the reason of the noisy and overfitted curves seen for these factors. They are kept for demonstration reason but their influence on the ACE regression is minimal and rejecting them could be considered. Note the small units associated to their transforms. Ace explains 81.6% of the uncertainty, as shown in the regression plot of the summation of all predictor transformations against the transformed response variable in Figure 5.12.

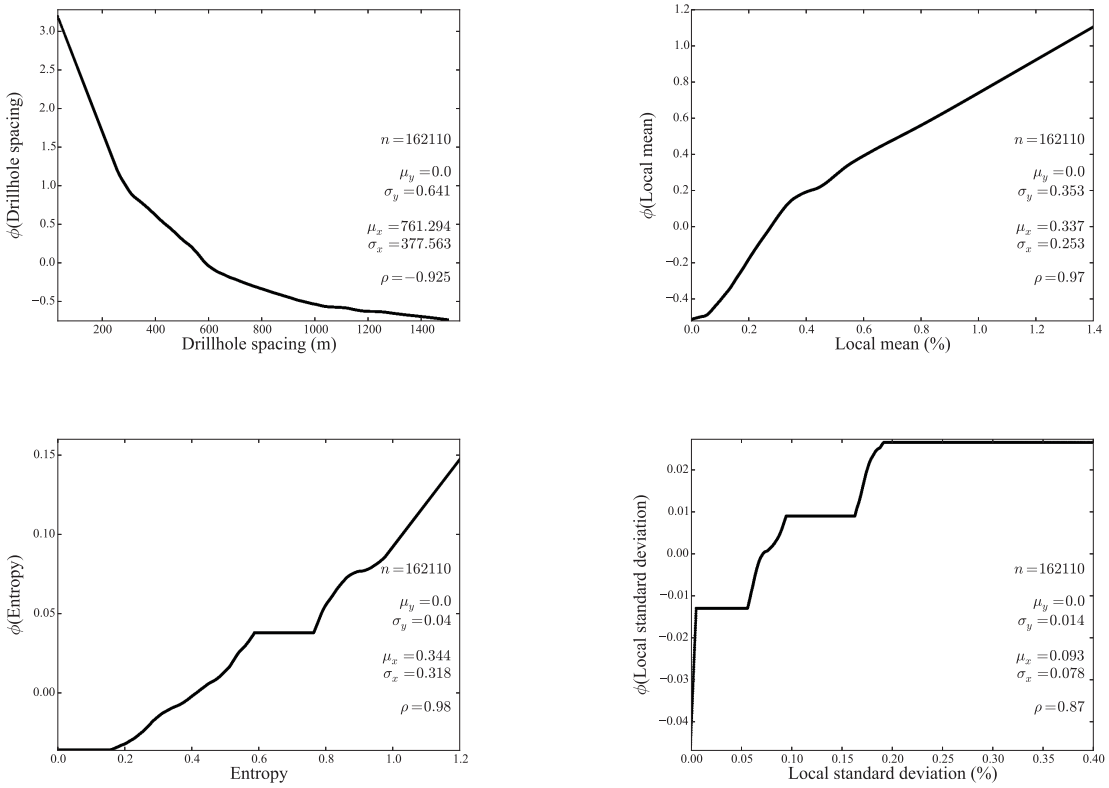

Figure 5.11: Scatterplots of the predictors (X-axis) against their transformed functions (Y-axis). Drillhole spacing is in the upper left corner, local mean is in the upper right corner. Entropy is the bottom left plot and the local standard deviation is the bottom right plot.

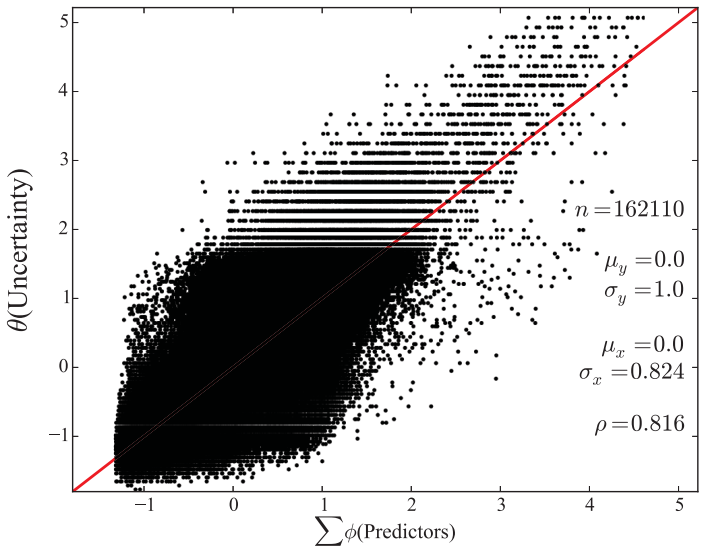

Figure 5.12: ACE regression plot. The high ACE correlation demonstrates the capacity of ACE to model the relationship between predictors and response variables.

5.2 Main Rock Types

The main rock types in the deposit are RTs 311, 321 and 401. These RTs have the highest average grades of variables A, B and C. A more comprehensive study of these rock types is presented. The P50 of the rock type realizations generated by SIS of RT 321 is shown in Figure 4.9, whereas the P50 of RTs 311 and 401 are shown in Figure 5.13. RT 211 is another important rock type, stratigraphically located above RT 311. RT 400 is located below RT 401, whereas RT 321 is located below RT 200. The importance of the RTs 311, 321 and 401 is also related to their locations. New vertical drilling through these rock types will also drill other rock types. RT 500 is an isolated rock type at the northeast region of the deposit and it is analyzed separately.

5.2.1 Uncertainty Versus DHS Plots

The uncertainty versus drillhole spacing curves and the kernel density are plotted for the main RTs in Figures 5.14 to 5.16. The combined kernel densities contours of these RTs

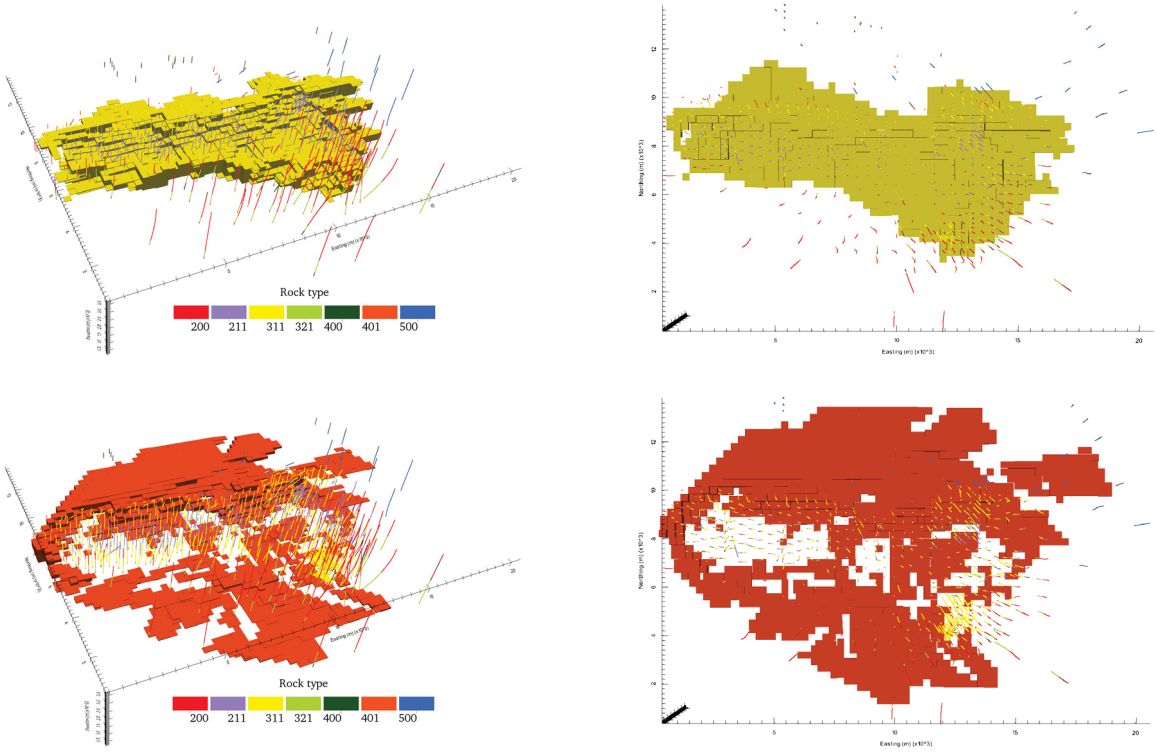

Figure 5.13: Oblique and plan view of the P50 based on global proportions in the final realization from SIS of RTs 311 (at the top) and 401 (on the bottom). Simulated SMUs and the drillholes are shown.

explain the density contours of Figure 5.4, calculated for the entire deposit. RT 311 has the smallest average DHS and lower uncertainty. Although the average DHS is greater in RTs 321 and 401, the EUC in these rock types can be calculated with confidence for small DHS values. In these RTs the EUC can be calculated for a level of uncertainty greater than 80%.

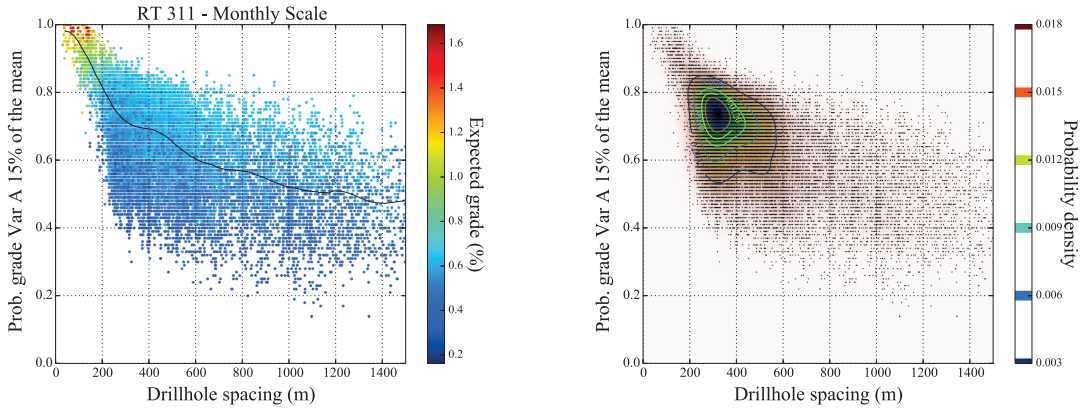

Figure 5.14: Uncertainty versus DHS curve (on the left) and the kernel density plot (on the right) for RT 311 calculated at a monthly scale.

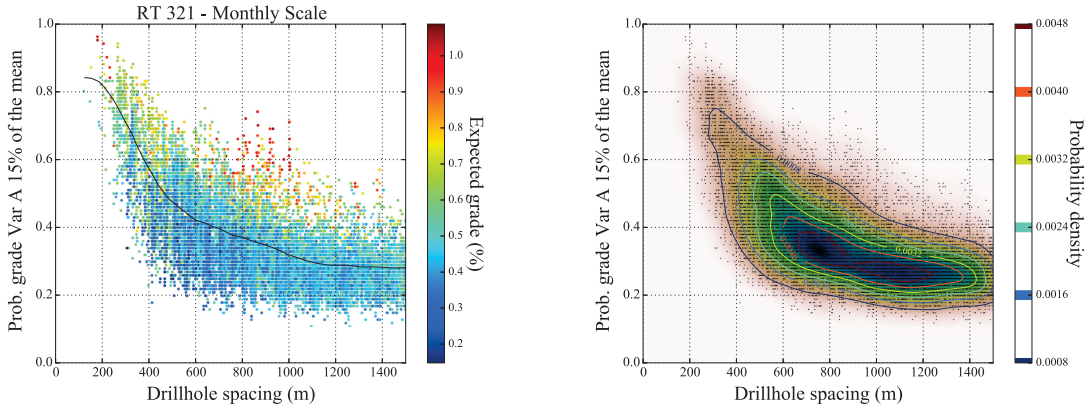

Figure 5.15: Uncertainty versus DHS curve (on the left) and the kernel density plot (on the right) for RT 321 calculated at a monthly scale.

The EUC of the main rock types are plotted together in Figure 5.17. The expected uncertainties in RTs 311 and 321 for the DHS interval of 200-300 m are similar. For other

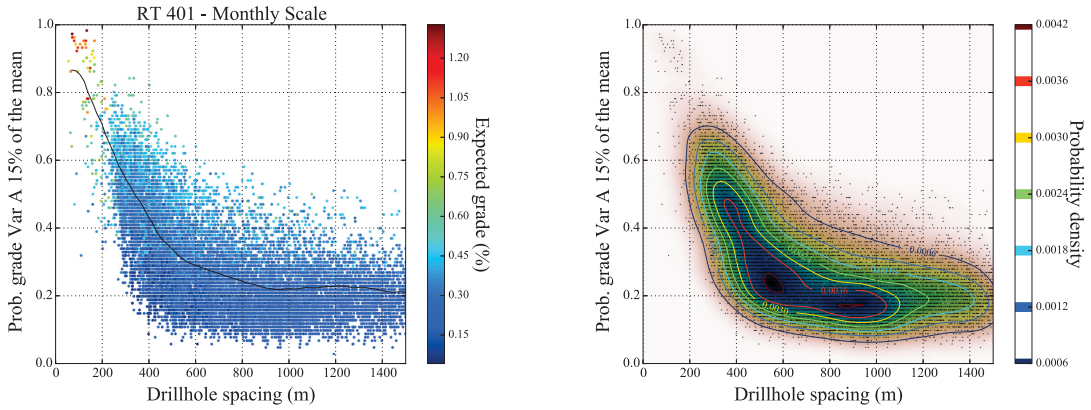

Figure 5.16: Uncertainty versus DHS curve (on the left) and the kernel density plot (on the right) for RT 401 calculated at a monthly scale.

DHS, the expected uncertainty in RT 311 is lower than RTs 321 and 401.

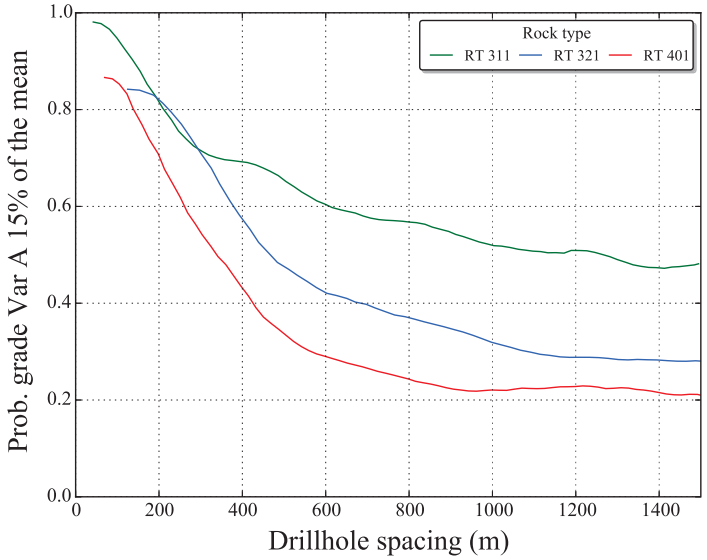

Figure 5.17: Uncertainty versus DHS curve for RTs 311 (green), 321 (blue) and 401 (red) calculated at a monthly scale.

5.2.2 Uncertainty Charts

A comparison of the ACE factors between the main rock types is shown in the tornado chart of Figure 5.18. The DHS is the most important factor in the three rock types. Entropy is an important explanatory factor in RTs 311 and 401. RT 311 is located in the center of the deposit, making contact to all other rock types. RT 401 is a thin rock type located in the bottom of the deposit, making contact to RT 400 (the bed rock) and other rock types located above it. Local mean is the second most important factor in RT 321.

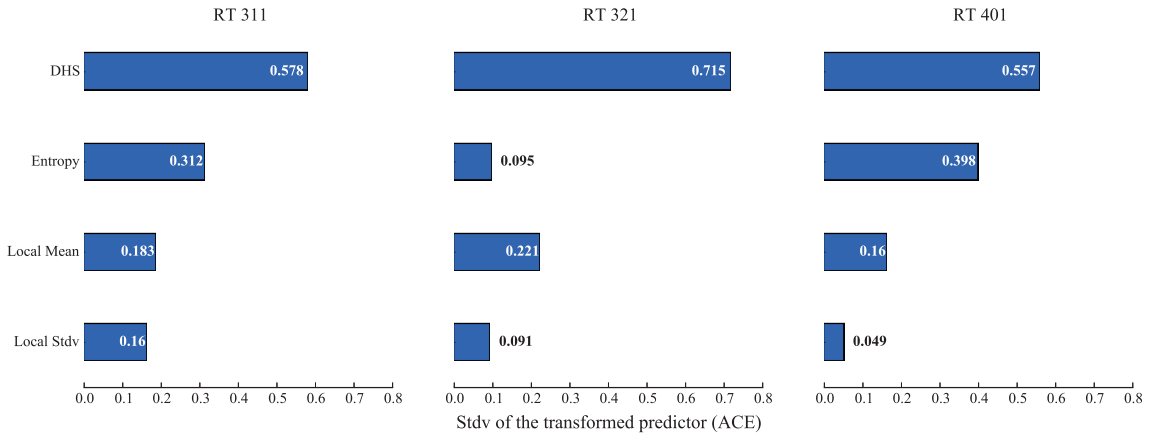

Figure 5.18: Tornado chart of the main rock types based on ACE transforms. DHS is the most important factor for RTs 311, 321 and 401. The location of RTs 311 and 401 relative to the other rock types explains why entropy as an important explanatory factor in these rock types.

The standard deviation of the ACE predictor transforms in all rock types, as well as the ACE correlations are summarized in Table 5.1. DHS is the most important factor in 5 of the seven rock types. Practically, the DHS and local mean have the same contribution explaining the uncertainty in RT 500. Local standard deviation is the least important factor in 6 of the seven rock types. A high ACE correlation is obtained in all rock types but RT 500. RT 400 covers the bottom of the entire deposit and has high proportions to the north of the deposit, where drillholes are sparse; which explains the high entropy in this RT.

Table 5.1: ACE summary for all rock types. The most important factor in each rock type is colored in blue, whereas the least important factor is colored in red. DHS and local standard deviation are likely the most and least important factors.

Rock Type	DHS	Entropy	Local mean	Local stdv	ACE Correlation
All	0.641	0.353	0.040	0.014	0.816
200	0.460	0.310	0.206	0.122	0.747
211	0.569	0.225	0.143	0.227	0.694
311	0.578	0.312	0.183	0.160	0.841
321	0.715	0.095	0.221	0.091	0.815
400	0.228	0.599	0.161	0.038	0.782
401	0.557	0.398	0.160	0.049	0.878
500	0.355	0.270	0.358	0.102	0.613

5.2.3 Production Scale Plots

The uncertainty versus drillhole spacing curves for the different production volumes in the entire deposit are shown in Figure 5.19. Higher rates that uncertainty changes are seen in the DHS interval of 200-800 m. All EUC show an inflection point approximately at 200 m spacing. The DHS associated to a same level of uncertainty increases from the monthly to the yearly production scales.

The uncertainty versus drillhole spacing curve calculated based on a quarterly production volume for RT 311 is shown in Figure 5.20. A comparison with the monthly scale plot for the same RT (Figure 5.14) shows a lower overall uncertainty and a higher expected grade in the SMUs. The uncertainty versus DHS curve calculated at different production volumes for RT 311 are shown together in Figure 5.21.

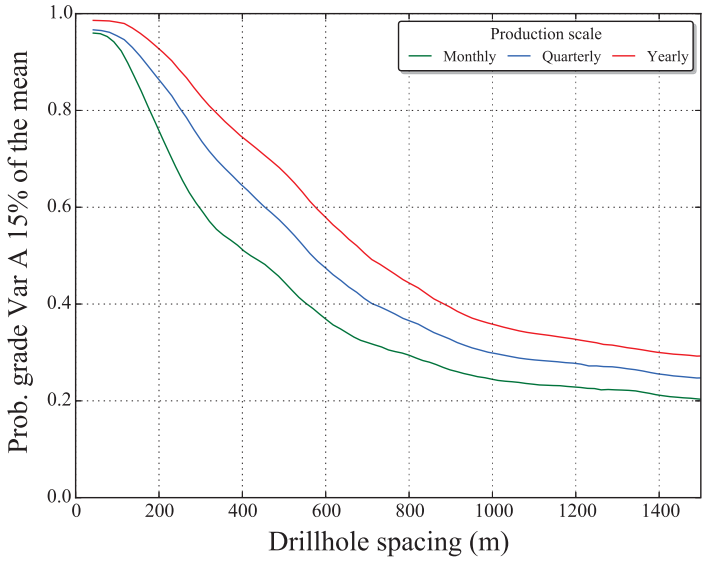

Figure 5.19: Uncertainty versus DHS curves at different production scales for the entire deposit. The monthly curve is shown in green, the quarterly in blue and the yearly in red.

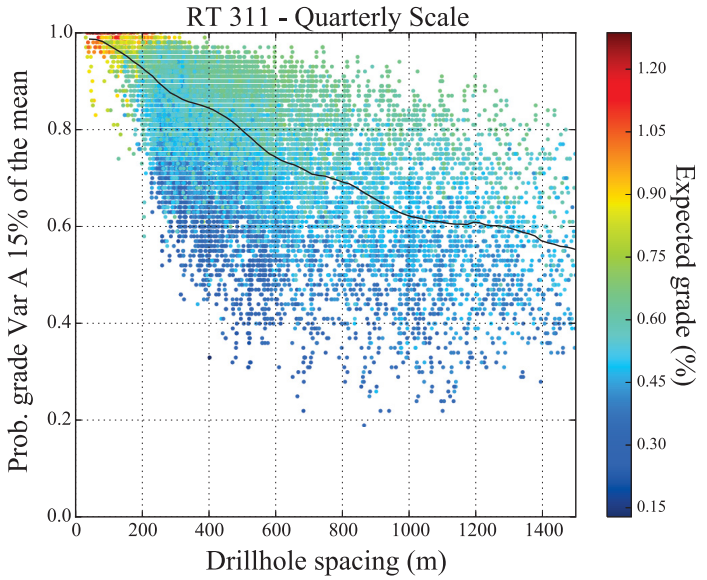

Figure 5.20: Uncertainty versus DHS curve at a quarterly production volume for RT 311.

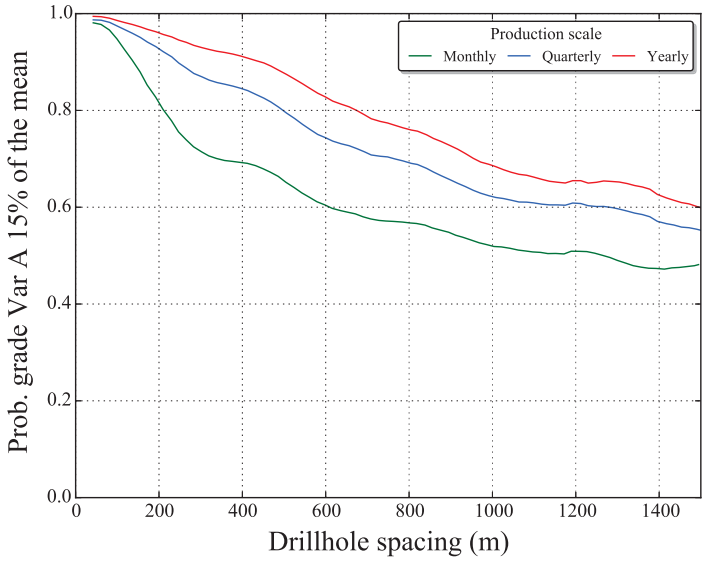

Figure 5.21: Uncertainty versus DHS curves at different production scales for RT 311.

5.3 Decision Support

An economic analysis is done to support the zoning of the deposit for the purpose of drillhole spacing. The analysis of the explanatory factors and uncertainty versus DHS curves in these zones are used for classification, uncertainty visualization and understanding of local uncertainty.

5.3.1 Factors Supporting Zoning

Besides uncertainty versus DHS curves, the zoning of the deposit is also supported by technical expertise, engineering and geological features, in addition to other economic factors. An economic analysis based on the equivalent grade (EG) at the deposit is done to support zoning. The equivalent grade, Equation 5.1, is calculated based on the simulated grades of Var A, B and C and summarize the information of the three variables in a unique economic model. Variable C is considered a contaminant, hence it is subtracted from the equivalent grade. A cut-off of 0.25% is applied to define ore and waste SMUs. All calculations are done at the SMU scale.

$$EG(\mathbf{u}) = VarA(\mathbf{u}) + VarB(\mathbf{u}) - 0.1 \times VarC(\mathbf{u}) \tag{5.1}$$

The economic value, stripping ratio, expected grade of the variables and the predominant rock type are calculated. With the ore and waste model, the 3D grid is transformed to a plan, with all calculations done vertically in the grid; see *3d2plan* (in the Appendix) for the explanation of the calculations. This analysis is compatible with the nearly vertical drilling of the deposit. The economic analysis is presented in the section below. Next section discusses the practical application of the zoning when targeting DHS.

5.3.1.1 Economic Value

The economic value (EV), Equation 5.2, is calculated by the summation of the SMUs equivalent grade; where the ore SMUs (n_o) receive a weight of 4 and the waste SMUs (n_w) receive the fixed value of -1. If only two SMUs are considered, one ore SMU with EG equals to the cut-off value and one waste SMU, then the economic value is zero ($EV = -1 + 4(0.25) = 0$).

$$EV(\mathbf{u}) = -1 \sum_{j=1}^{n_w} n_w(\mathbf{u}) + 4 \sum_{i=1}^{n_o} EG_o(\mathbf{u}) \tag{5.2}$$

The ore and waste location maps are shown in Figure 5.22. The majority part of the deposit is considered ore (left plot). In the ore zones, the highest economic values are located in the central, northeast and southeast regions of the deposit.

5.3.1.2 Stripping Ratio

The stripping ratio is calculated dividing the number of waste SMUs by the number of ore SMUs (going to the bottom of the model in each vertical column). The stripping ratio map in Figure 5.23 shows the lowest stripping ratios in the central, northeast and southeast zones. Relatively low stripping ratio is seen in the south, closer to the central region.

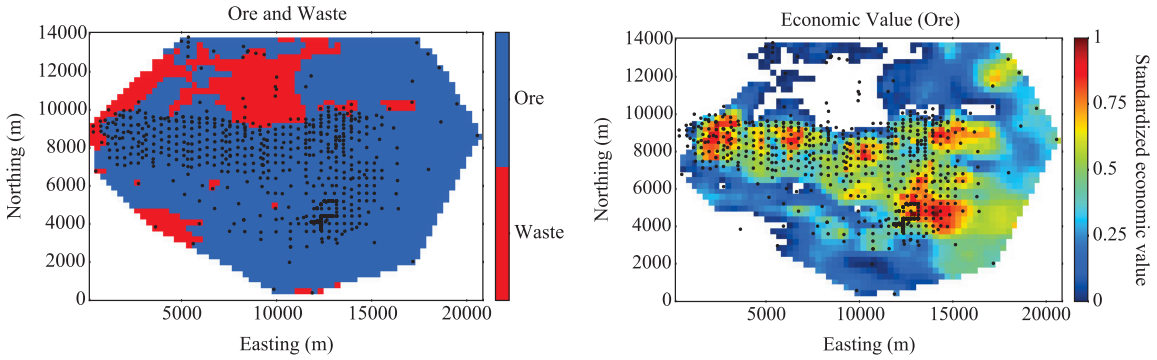

Figure 5.22: Economic value map for ore and waste regions (on the left) and only for ore (on the right). The zones with highest economic values are revealed when only ore in plotted (on the right). The dots represent the collar of the drillholes.

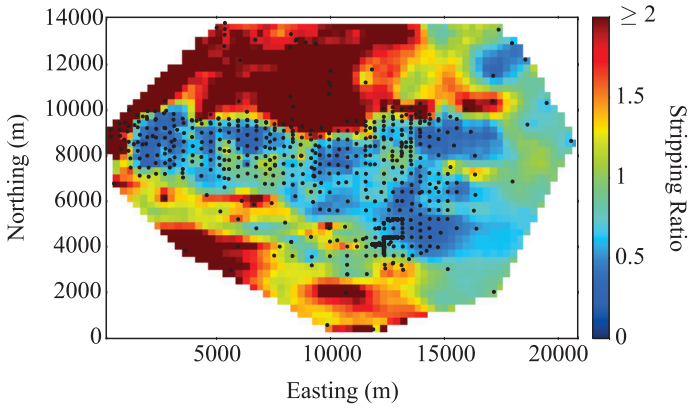

Figure 5.23: Stripping ratio map. The zones with lowest ratios are central, northeast and southeast. Relative low stripping ratio is seen in the south, closer to the central region.

5.3.1.3 Expected Grade

The expected grade is calculated by the average of the SMUs grades for the entire vertical column. The expected grade maps are shown in Figure 5.24. Variables A and B have the highest grades in the central, northeast and southeast zones. Relative high grade of Variable B is also seen in the south. Variable C has the highest grades at the contact of the RTs 311 and 401 and to the north of the deposit.

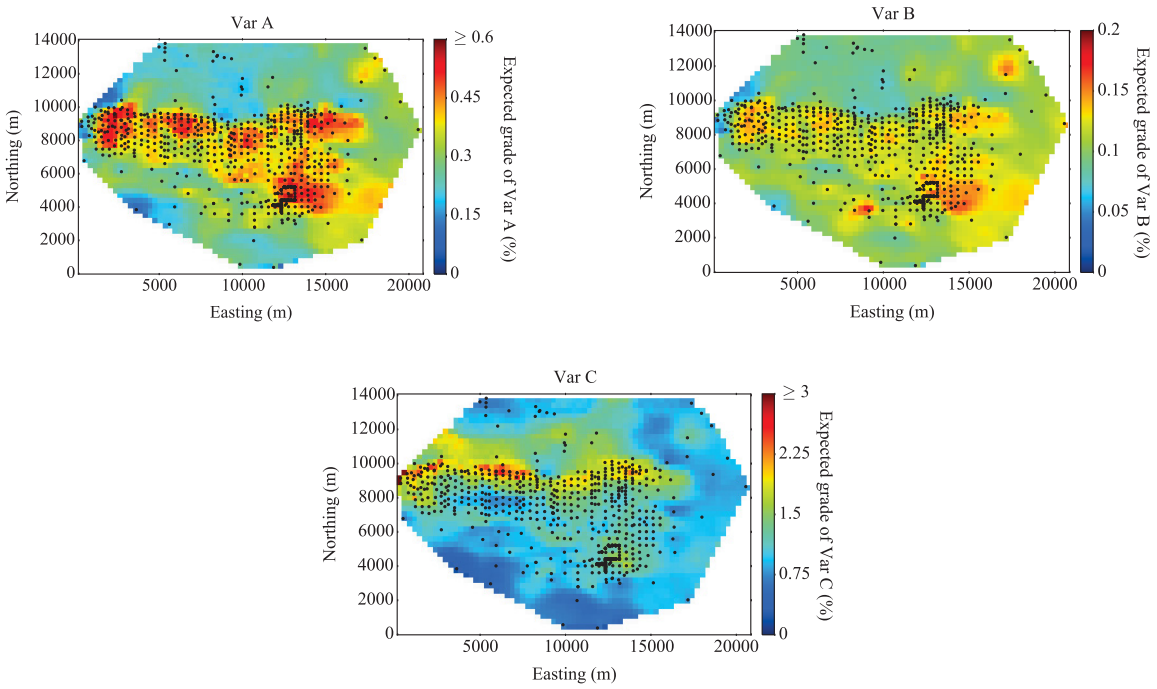

Figure 5.24: Expected grade maps of the variables. Variables A and B have the highest grades in the central, northeast and southeast zones. Variable C has the highest grades to the north of the deposit.

5.3.1.4 Predominant Rock Type

The predominant rock type is calculated by the statistical mode of the RTs. After defining the zones based on the economical factors, the uncertainty versus drillhole spacing curves are analyzed for the predominant rock type inside the zones. The predominant rock type

map is shown in Figure 5.25.

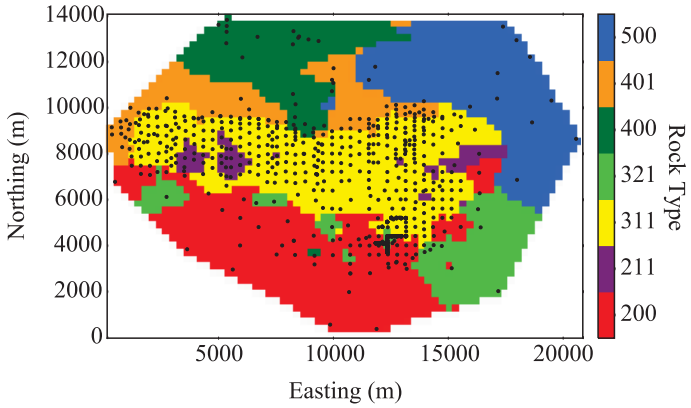

Figure 5.25: Predominant rock type map. The most common rock types over the deposit per region are RT 311 in the center, RT 321 in the southeast, RT 200 in the south, RT 500 in the northeast, and RTs 400 and 401 in the north.

5.3.2 Zoning

Based on the economic analysis and the predominant rock types, four zones are defined to target DHS, see Figure 5.26. The zones with their respectively predominant rock type are: main or central (RT 311), south (RT 200), southeast (RT 321) and northeast (RT 500). This zoning was chosen for illustration purposes. In practice, more detailed knowledge of the geological setting and mine plan would be considered.

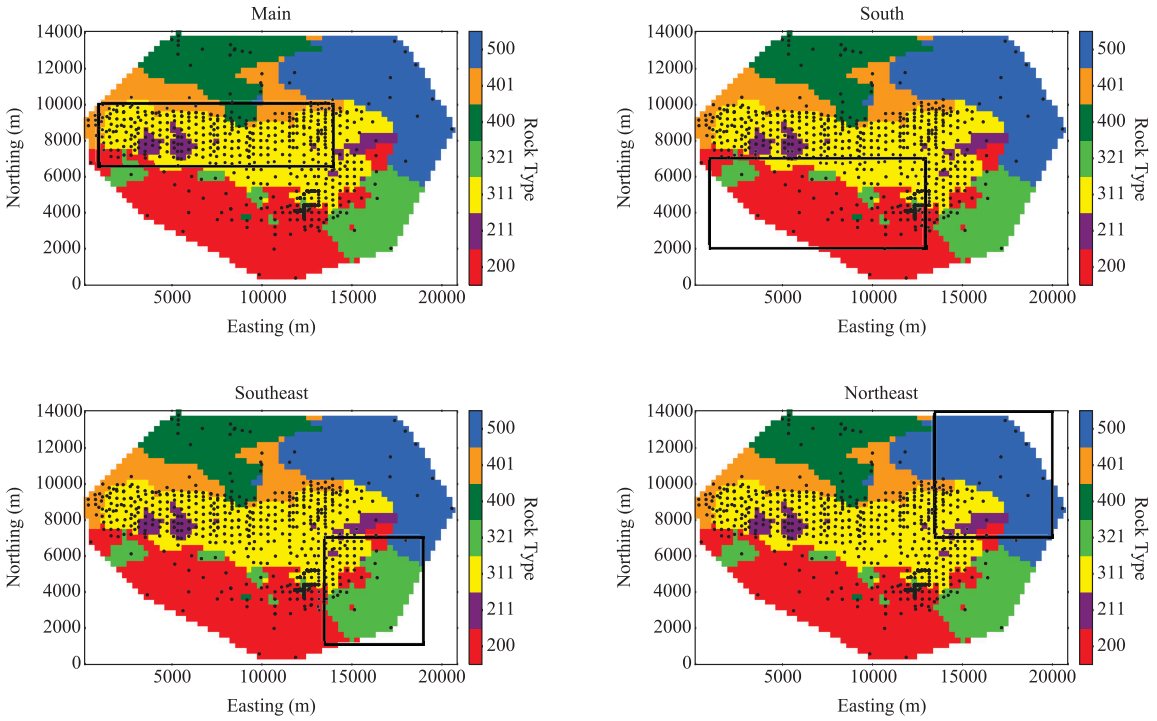

Figure 5.26: The four defined zones are shown in the rectangles. The main or central zone has RT 311 as predominant rock type. South, southeast and northeast zones have RTs 200, 321 and 500 as respective predominant rock types.

5.4 Practical Applications and Decision Support

One application of uncertainty versus DHS curves is for classification. Other application include targeting DHS to reduce local uncertainty, understanding the local variability and better planning of further drilling campaign.

5.4.1 Classification and Targeting DHS

Consider targeting DHS for classification, where SMUs are classified as measured or indicated (JORC, 2012) based on, for example, the probability of the grade of Var A to be within 15% of the mean at least 80% of the times for a quarterly and yearly production scales respectively. Consider the four zones previously defined and the uncertainty versus DHS curves calculated at these two production scales for the predominant rock type in each zone.

5.4.1.1 Main Zone

In the main zone, Figure 5.27, a DHS of 500 m classifies SMUs as measured, whereas a DHS of 670 m, 34% greater, classifies SMUs as indicated.

5.4.1.2 Southeast Zone

RT 321 is the predominant rock type in the southeast zone. Similar analysis made for the main zone is done for RT 321. Based on the EUC of Figure 5.28, a DHS of 300 m classifies SMUs as measured, whereas a DHS of 390 m, 30% greater, classifies SMUs as indicated.

5.4.1.3 Northeast Zone

The predominant rock type in the northeast zone is RT 500. The EUC calculated at a quarterly production volume does not reach the required uncertainty level for classification, see Figure 5.29. Although a DHS can be defined for classification purpose

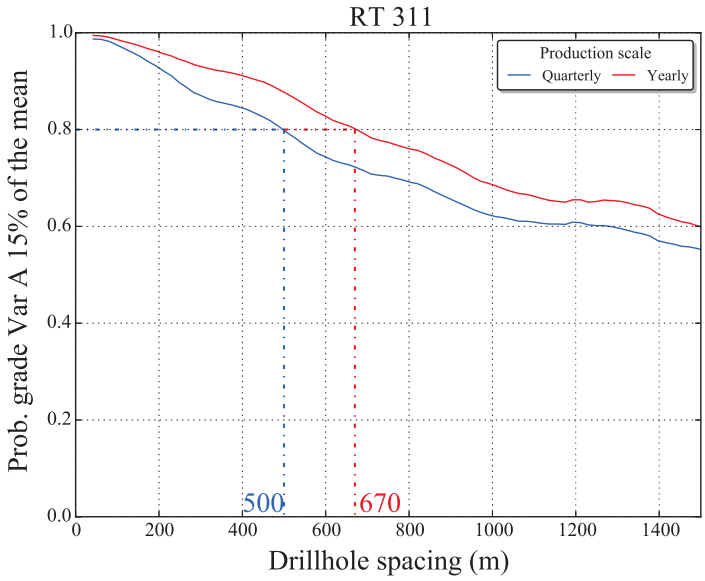

Figure 5.27: DHS required for classification in the main zone, based on the uncertainty versus DHS curves calculated for RT 311 at a quarterly and yearly production volumes.

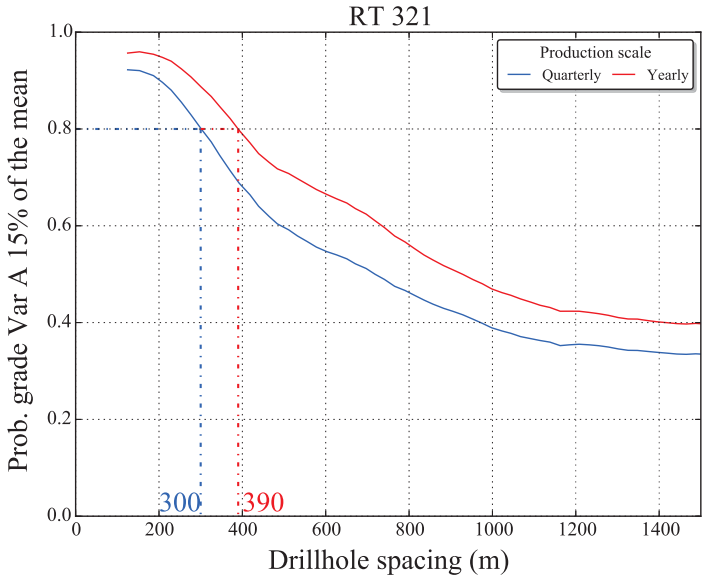

Figure 5.28: DHS required for classification in the southeast zone, based on the uncertainty versus DHS curves calculated for RT 321 at a quarterly and yearly production volumes.

at a yearly production scale, the uncertainty is high in this rock type. The EUC is plotted for a monthly scale in Figure 5.30. There is no many SMUs at DHS less than 400 m to support a reliable calculation of the expected uncertainty curve at that DHS. The re-sampling approach (Section 1.4.1) should be considered for a better decision making in this zone.

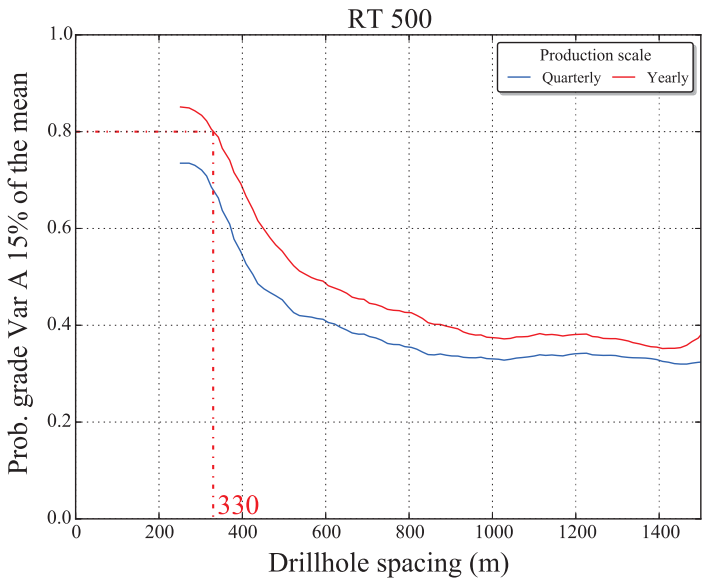

Figure 5.29: DHS required for classification in the northeast zone, based on the uncertainty versus DHS curves calculated for RT 500 at a quarterly and yearly production volumes. The required uncertainty level of 80% is not achieved at a quarterly production scale.

5.4.1.4 South Zone

Despite the large number of SMUs in RT 200 and the fact that some SMUs show low uncertainty, the average uncertainty is high and the EUC does not reach the required uncertainty level for classification; see Figure 5.31 for the EUC and the uncertainty distribution in this rock type. If any decision must be based on RT 200, then re-sampling should be considered. Otherwise, due to the geological location of RTs 200 and 321 (Figure 5.32), the uncertainty curves of RT 321 can be used as reference.

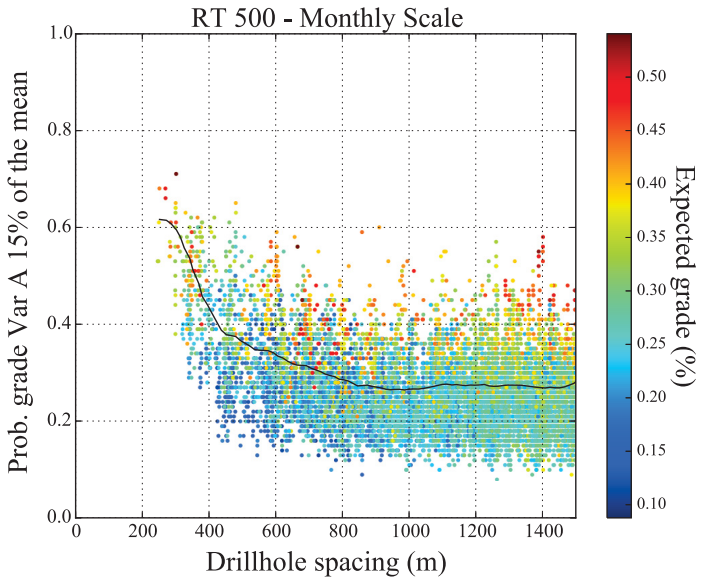

Figure 5.30: Uncertainty versus DHS curve at a monthly production volume for RT 500.

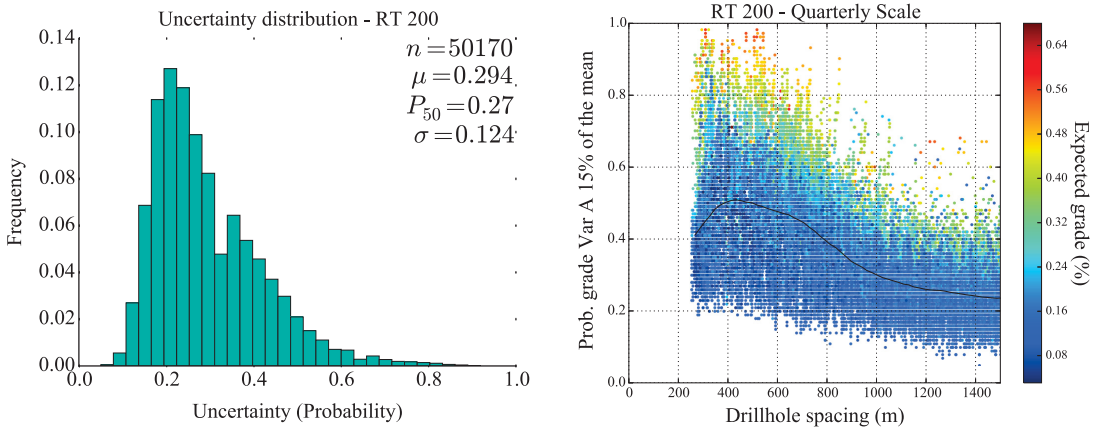

Figure 5.31: Uncertainty distribution (on the left) and the uncertainty versus drillhole spacing curve (on the right) for RT 200, calculated based on a quarterly production scale.

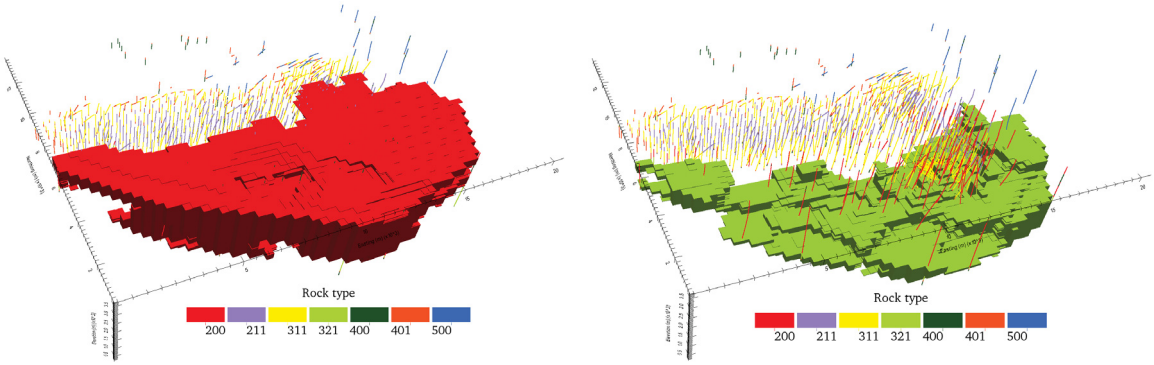

Figure 5.32: Oblique view of the P50 based on global proportions in the final realization from SIS of RTs 200 (on the left) and 321 (on the right). Simulated SMUs and the drillholes are shown.

5.4.2 Uncertainty Visualization

Uncertainty versus DHS curves are useful to visualize uncertainty, providing more information to support decision. Consider the monthly and the yearly production scale curves for the entire deposit in Figure 5.33. SMUs located above the EUC are more certain, and SMUs below it are less certain. The delta uncertainty for all SMUs is calculated and plotted for a slice of the deposit. The SMUs with a level of uncertainty greater or equal than 80% based on a yearly production scale are also plotted.

5.4.3 Learning Curve

The Learning Curve is calculated for the entire deposit at different production scales, see Figure 5.34. The LCs show low rates for DHS greater than 1000 m and higher rates at the DHS of 200-800 m for all scales. Below the DHS of 200 m the gain of information diminishes and the LC drops at all scales.

The EUC and LC calculated at all production scales are plotted together in Figure 5.35. The yearly production scale resolves faster a scale of variability for DHS greater than 500 m. The steeper inclination of the monthly scale EUC for DHS below 300 m reflects in the LC,

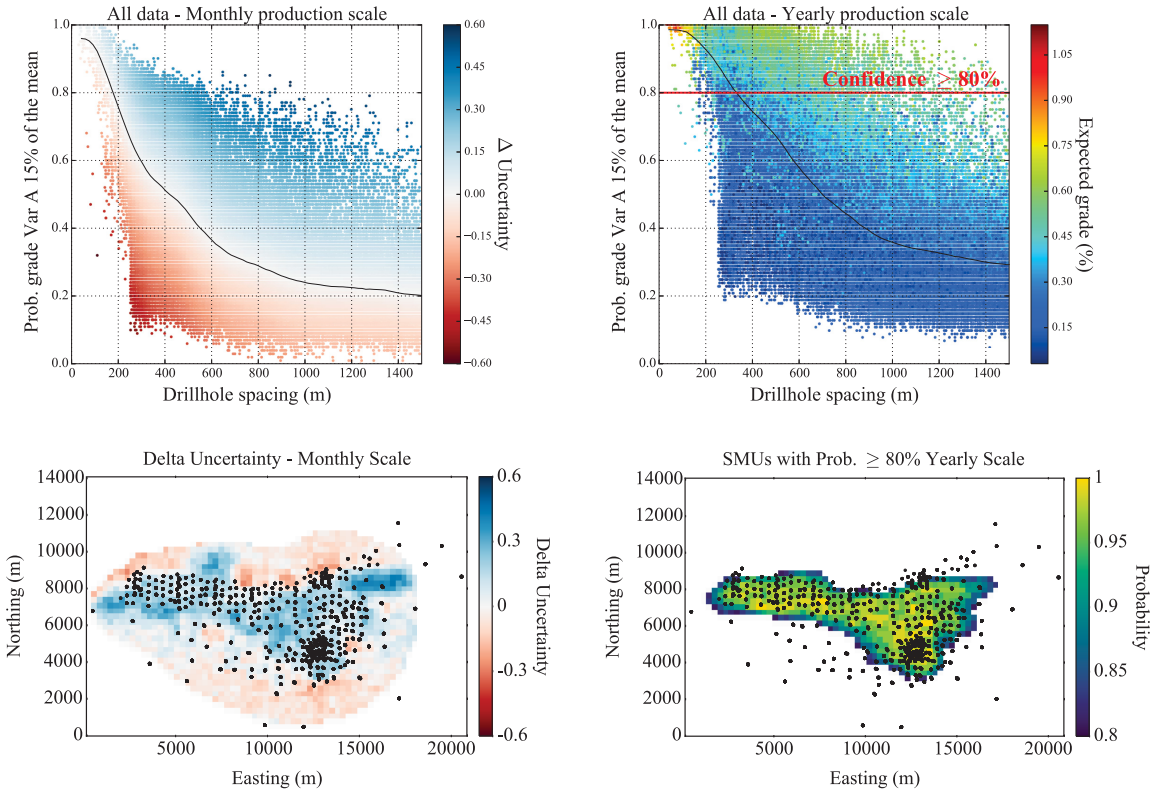

Figure 5.33: Plan view of the uncertainty for a slice of the deposit (at the bottom) based on the monthly and yearly production scale uncertainty curves (at the top). More and less certain SMUs are plotted for a monthly production scale at the bottom-left corner. SMUs with a level uncertainty greater or equal than 80% for a yearly production scale are shown at the bottom-right corner.

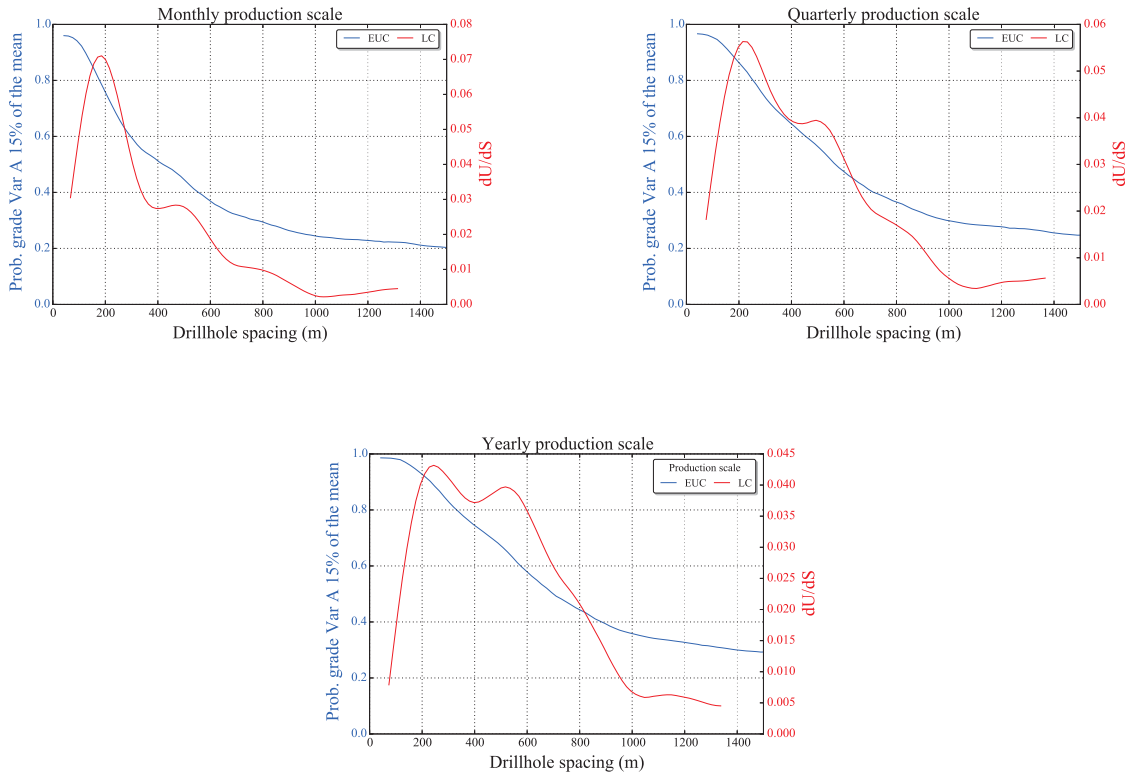

Figure 5.34: The EUC and the Learning Curve calculated at different production scales for the entire deposit.

and the variability is resolved faster at this scale for this interval of DHS. Three regimes of uncertainty are defined from the analysis of the EUC and LC. Large spacing regime is defined for DHS greater than 800 m. Little uncertainty is explained in this regime. The regime that the uncertainty is resolved faster is in the DHS interval of 200-800 m. Below DHS of 200 m the contribution of extra drilling diminishes.

In order to understand the behavior of the EUC and LC in the DHS of 300-500 m, the re-sampling approach may be considered. Drilling the simulated model at this DHS and re-calculating the EUC and LC again would provide more information to understand the flatness of the LC in that DHS interval.

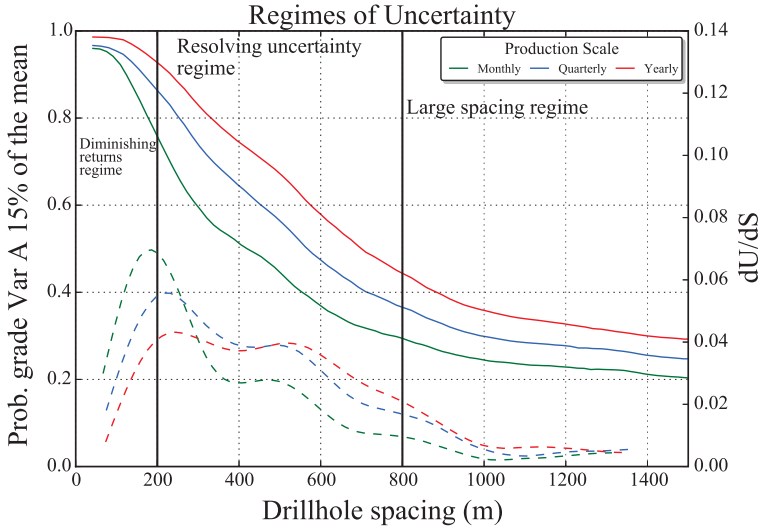

Figure 5.35: Regimes of uncertainty based on the EUC (continuous lines) and the LC (dashed lines) calculated at different production scales. The scale of the Y-axis of the Learning Curve is modified to improve visualization. The actual LC values are not changed.

Chapter 6: Conclusions and Future Work

Drilling the correct amount is an important goal for any mining company. Drilling is the most direct approach to access subsurface information of the grade continuity and geological variability of a mineral deposit. Uncertainty is often evaluated by geostatistical simulation of multiple realizations of the grade and rock types. Decisions are taken based on the available data and level of uncertainty of the simulated models. Decisions can be in terms of classification, understanding of local uncertainty and improvement of estimates. Uncertainty is mainly affected by the amount of data, the local grade variability and the local geology. Improved understanding of these explanatory local factors would improve decisions on the amount of drilling required.

A relatively large production scale is relevant for resource evaluation, mine planning and for planning future drilling campaign. Measures of uncertainty and the uncertainty versus drillhole spacing curve are calculated for a relevant production scale. There is a level of uncertainty for drillhole spacing and the analysis of the expected uncertainty curve and the Learning Curve, in addition to the explanatory factors, lead to better decisions when targeting the regular drillhole spacing.

The influence of the explanatory factors on uncertainty, as well as the effect of scale, have been presented. The concept of the Learning Curve has been presented and practically applied in the case study. The case study is important to demonstrate the concepts and details of the uncertainty versus drillhole spacing study. The main contributions of this thesis are discussed below.

6.1 Topics Covered and Contributions

The focus of this thesis is not optimizing infill drillhole placement, but to support the choice of the regular drillhole spacing that is related to a level of uncertainty. Important concepts involving regular spacing are discussed in Chapter 1. Geostatistical models have a capacity to assess how uncertainty improves with more data. They are generated at the data scale and averaged-up to a relevant scale such as an SMU scale relevant for selectivity. Assessing uncertainty at a production scale is desirable for long term planning and drilling campaigns. Uncertainty is scale dependent and different approach to assess uncertainty in different production scales are presented. The zero cut-off and fixed volume approach makes the analysis of uncertainty less based on economics, focusing more on the grade uncertainty. In this approach the uncertainty is assessed by using a moving window with size representing an arbitrary production volume, centered at the SMU. This approach has been used in the case study. Uncertainty and drillhole spacing are analyzed in terms of the expected uncertainty curve (EUC). For a given DHS there is an average uncertainty. The shape of the EUC changes with the measure of uncertainty and scale considered. When the expected uncertainty versus drillhole spacing curve cannot be reliably calculated for an interval of DHS, a re-sampling approach should be considered. One model is sampled to create new artificial drillholes and uncertainty is re-simulated to fill the gaps in the EUC.

The concept of the Learning Curve (LC) and the regimes of uncertainty are presented in Chapter 2. The LC calculates the rate at which the uncertainty is resolved with additional data; the LC is the derivative of the EUC. The LC methodology is developed based on kriging. The effect of scale and variogram model on the LC is demonstrated. The interpretation of the LC is in terms of resolving a scale of variability for different intervals of drillhole spacing. In the conceptual calculation of the Learning Curve, stationarity is assumed, the variogram model is fixed and only the drillhole spacing changes. Conceptually, the variability is resolved faster at larger scales, although the rate that uncertainty is resolved at a given DHS depends on the variogram model. However, in the

second part of the case study, Chapter 5, it has been shown that the LC calculated based on a monthly production scale is higher than the quarterly and yearly production scales for relative small DHS. Practically, the analysis of the Learning Curve must be always in terms of the calculated EUC. Although the expected uncertainty is lower for larger volumes (higher probability to be within a tolerance of the mean), there may be intervals of DHS that the EUC is steeper at a smaller scale. The LC provides tools to clarify the information contained in the expected uncertainty curve. The calculation of the LC is done analytically, when the LC is noisy the EUC should be used instead as reference for decision-making.

Different measures of uncertainty, confounding and explanatory factors are discussed in Chapter 3. The influence of the proportional effect, variogram stationarity and scale on uncertainty are presented. The proportional effect depends on the histogram and its skewness. Uncertainty is lower at large scale because high and low values are averaged out. Uncertainty depends on the variogram continuity and it is usually higher in less continuous areas. Relevant explanatory factors such as drillhole spacing, entropy, local mean and local standard deviation are explained. These factors explain most of the local uncertainty and their impact on the total variability can be calculated by statistical regression analysis. Due to its capacity of detecting underlying relationships between variables, ACE is the model that explains most of the uncertainty.

The first part of the case study, in Chapter 4, discusses the rock type and grade modelling. Modelling the rock types with SIS allows transferring the uncertainty of the geological model to the grade model. The multivariate grade modelling with PPMT is considered because the multivariate distributions are non-Gaussian. The uncertainty assessed with the multivariate workflow has a great impact on the drillhole spacing study, hence the importance of a good model of uncertainty. In Chapter 5, ACE explains the local variability due to the explanatory factors in the entire deposit and within rock types. Drillhole spacing is the most important factor in most rock types. Uncertainty versus drillhole spacing curves are calculated for the main rock types at different production scales. The economic analysis of the equivalent grade, stripping ratio and expected grade

along with the predominant rock type, support the zoning of the deposit for targeting DHS. Four zones are defined and the EUC calculated at the quarterly and yearly production scales are used to suggest classification of SMUs as measured and indicated respectively. The re-sampling approach is an alternative for the northeast zone, once the EUC is noisy and does not reach the required level of uncertainty for classification. The use of uncertainty versus drillhole spacing curves for uncertainty visualization and understanding of local uncertainty has been demonstrated. The Learning Curve is calculated for the entire deposit and the regimes of uncertainty are defined.

The main contribution of this work is to present important concepts about drillhole spacing and uncertainty to support decision regarding further drilling campaign and which is the regular drillhole spacing required to reach an expected level of uncertainty. The methodology applied for assessing uncertainty at different production scales, the analysis of the explanatory factors to explain the local uncertainty and the interpretation of the expected uncertainty curves and the Learning Curve can be used and applied to real deposits. These concepts have been discussed and practically demonstrated in this thesis.

6.2 Future Work

Uncertainty versus drillhole spacing is a vast subject for research. Some contributions have been made; however future work remains.

A study demonstrating the effectiveness of assessing uncertainty at an arbitrary production volume from different approaches, such as those presented in Chapter 1, would be valuable for this subject. The assessment of the uncertainty at a production volume requires a grid in which calculations are done. This grid represent the arbitrary production volume and its geometry could be made more or less rigid to match a specific criterion such as the tonnes of ore processed monthly. For instance, a period based uncertainty could be analyzed. In this case, the SMUs mined in the same period but from different locations of the mine could be combined into a production volume for that period, and the uncertainty could be based on the period instead of a fixed grid location.

The expected uncertainty curves are calculated based on a regular drillhole spacing. The applied drillhole spacing might be corrected based on the local anisotropy. For example, if the regular drillhole spacing that classifies SMUs as measured is 300 m and the anisotropy in this rock type is 2:1 at N90°E azimuth direction, then the drillhole spacing should be 600 m at the azimuth direction and 300 m at the direction perpendicular to it. The directly consideration of the anisotropy on the EUC would be valuable.

A number of factors impact the capacity of a model to represent the true uncertainty such as data quality and quantity. Parameter uncertainty has not been considered in the case study. If there are too few data to get a good variogram then uncertainty in the variogram should be considered. Uncertainty in the data should also be added to the workflow. Getting a good uncertainty model is required to avoid bias in the expected uncertainty curves. In the same context, non stationarity affects local and global uncertainty. Trends will likely change the uncertainty versus drillhole spacing curves.

The time scale of mining is also a source of uncertainty. The larger production volumes may not be well defined at early stages of the mine, which results in uncertainty in the grid definition (Silva, 2015). The drillhole spacing defined from EUC calculated at a certain grid might change if the arbitrary production volume changes its configuration, even if the volume or the number of SMU within that volume remain the same.

The impact of different data types on uncertainty versus drillhole spacing curves is an interesting avenue for further research. Different drilling methods would have different contributions on the EUC. It might be necessary to analyze different data types in different uncertainty versus drillhole spacing curves. Moreover, exhaustive secondary data such as geophysics may reduce the drilling requirements.

Finally, in this thesis, the drillhole spacing required to achieve an expected uncertainty is given without considering costs. There is a need for some economical analysis when collecting data. The Value of Information (Bratvold et al., 2007; Warren, 1983) in addition to the Learning Curve could be used to support the decision of acquiring more information.

6.3 Recommendations

Drillhole spacing is used, in this thesis, as the measure of data availability by the author's choice. Drillhole density can replace the drillhole spacing, in all the uncertainty versus drillhole spacing plots, without changing the nature of the analysis. Drillhole density provides a direct measure of the cost of drilling, once drillhole density is given in terms of the number of drillholes within the considered area. In cases where drilling geometry and spatial oriental are complex, reliable drillhole spacing and density calculations are difficult. In such cases, the distance to the nearest drillhole could be considered. The interpretation of the uncertainty versus the geometric criteria curves must be done consistently and accompanied with professional expertise.

References

(1983). Optimum locations for exploratory drill holes. *International Journal of Mining Engineering*, 1(4):343--355.

Andricevic, R. (1990). Cost-effective network design for groundwater flow monitoring. *Stochastic Hydrology and Hydraulics*, 4(1):27--41.

Barnett, R. M. (2011). Tools for multivariate geostatistical modeling. *CCG Guidebook Series, Volume 13*.

Barnett, R. M. and Deutsch, C. V. (2013a). Assessing the uncertainty and value of ace transformations. *CCG Annual Report 15, Paper 402*.

Barnett, R. M. and Deutsch, C. V. (2013b). Tutorial and tools for ace regression and transformation. *CCG Annual Report 15, Paper 401*.

Barnett, R. M. and Deutsch, C. V. (2014). A compressed binary format for large geostatistical models. *CCG Anuual Report 16, Paper 413*.

Barnett, R. M., Manchuk, J. G., and Deutsch, C. V. (2014). Projection pursuit multivariate transform. *Mathematical Geosciences*, 46(3):337--359.

Bertoli, O., Paul, A., Casley, Z., and Dunn, D. (2013). Geostatistical drillhole spacing analysis for coal resource classification in the bowen basin, queensland. *International Journal of Coal Geology*, 112:107 -- 113.

Boucher, A., Dimitrakopoulos, R., and Vargas-Guzmán, J. (2004). Joint simulations, optimal drillhole spacing and the role of the stockpile. In Leuangthong, O. and Deutsch, C. V., editors, *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 35--44. Springer Netherlands.

Bratvold, R. B., Bickel, J. E., and Lohne, H. P. (2007). Value of information in the oil and gas industry: Past, present, and future. *SPE Annual Technical Conference and Exhibition,*

11-14 November, Anaheim, California, U.S.A.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580--598.

Bueso, M., Angulo, J., and Alonso, F. (1998). A state-space model approach to optimum spatial sampling design based on entropy. *Environmental and Ecological Statistics*, 5(1):29--44.

Bueso, M., Angulo, J., Cruz-Sanjulián, J., and García-Aróstegui, J. (1999). Optimal spatial sampling design in a multivariate framework. *Mathematical Geology*, 31(5):507--525.

Caers, P. (2011). *Modeling Uncertainty in the Earth Sciences*. John Wiley & Sons.

Carrera, J., Usunoff, E., and Szidarovszky, F. (1984). A method for optimal observation network design for groundwater management. *Journal of Hydrology*, 73(1):147--163.

Criminisi, A., Tucciarelli, T., and Karatzas, G. P. (1997). A methodology to determine optimal transmissivity measurement locations in groundwater quality management models with scarce field information. *Water resources research*, 33(6):1265--1274.

CRIRSCO (2013). International reporting template for the public reporting of exploration results, mineral resources and mineral reserves. http://www.crirsco.com/templates/crirsco_international_reporting_template_2013.pdf.

Daniels, E. B. (2015). Prediction of local uncertainty for resource evaluation. Master's thesis, University of Alberta, Edmonton, Alberta, Canada.

Darcy, R. and Aigner, H. (1980). The uses of entropy in the multivariate analysis of categorical variables. *American Journal of Political Science*, 24(1):155--174.

Deutsch, C. (2005). A sequential indicator simulation program for categorical variables with point and block data: Blocksis. *CCG Annual Report 07, Paper 402*.

Deutsch, C. and Beardow, A. (1999). Optimal drillhole spacing for oil sands delineation. *CIM Annual Meeting, Calgary, Alberta*.

Deutsch, C. V. and Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide*. Oxford University Press New York.

Deutsch, J. L. and Deutsch, C. V. (2011). Plotting and checking the bivariate distributions of multiple gaussian data. *Computers & Geosciences*, 37(10):1677 -- 1684.

Diehl, P. and David, M. (1982). Classification of ore reserves/resources based on geostatistical methods. *CIM Bulletin*, 75(838):127--136.

Draper, N. R. and Smith, H. (2014). *Applied regression analysis*. John Wiley & Sons.

Englund, E. J. and Heravi, N. (1992). Conditional simulation: practical application for sampling design optimization. In *Geostatistics Tróia'92*, pages 613--624. Springer.

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

Froidevaux, R. (1982). Geostatistics and ore reserve classification. *CIM Bulletin*, 75(843):77--83.

Gershon, M. (1987). Comparisons of geostatistical approaches for drill hole site selection. In *Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries (APCOM)*, pages 93--100. Kluwer Academic Publishers.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford university press.

Gotway, C. and Rutherford, B. (1994). Stochastic simulation for imaging spatial uncertainty: Comparison and evaluation of available algorithms. In Armstrong, M. and Dowd, P., editors, *Geostatistical Simulations*, volume 7 of *Quantitative Geology and*

Geostatistics, pages 1--21. Springer Netherlands.

Isaaks, E. H. (1990). *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data*. PhD thesis, Stanford University, Stanford, United States.

Isaaks, E. H. and Srivastava, R. M. (1989). *Applied geostatistics*, volume 2. Oxford University Press New York.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th Edition)*. Pearson Education.

JORC (2012). Australasian joint ore reserves committee. JORC code, 2012 edition. http://http://www.jorc.org.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics*. Blackburn Press.

Journel, A. G. (1974). Geostatistics for conditional simulation of ore bodies. *Economic Geology*, 69(5):673--687.

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

Keogh, E. and Mueen, A. (2010). Curse of dimensionality. In Sammut, C. and Webb, G. I., editors, *Encyclopedia of Machine Learning*, volume 1 of *Quantitative Geology and Geostatistics*, pages 257--258. Springer US.

Koppe, V. C., Costa, J. F. C. L., de Lemos Peroni, R., and Koppe, J. C. (2011). Choosing between two kind of sampling patterns using geostatistical simulation: Regularly spaced or at high uncertainty locations? *Natural Resources Research*, 20(2):131--142.

Leuangthong, O., Deutsch, C. V., , and Ortiz, J. M. (2006). A case for geometric criteria in resources and reserves classification. *CCG Annual Report 15, Paper 312*.

Leuangthong, O. and Deutsch, C. V. (2003). Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, 35(2):155--173.

Loaiciga, H. A. (1989). An optimization approach for groundwater quality monitoring network design. *Water Resources Research*, 25(8):1771--1782.

Manchuk, J. G. and Deutsch, C. V. (2015). Matrix transformations and their effect on the variogram. *CCG Annual Report 17, Paper 132*.

Matheron, G. (1973). The intrinsic random functions and their applications. *Advances in Applied Probability*, 5(3):439--468.

McBratney, A., Webster, R., and Burgess, T. (1981a). The design of optimal sampling schemes for local estimation and mapping of of regionalized variables—I. *Computers & Geosciences*, 7(4):331 -- 334.

McBratney, A., Webster, R., and Burgess, T. (1981b). The design of optimal sampling schemes for local estimation and mapping of of regionalized variables—II: Program and examples. *Computers & Geosciences*, 7(4):331 -- 334.

Meyer, P. D. and Brill, E. D. (1988). A method for locating wells in a groundwater monitoring network under conditions of uncertainty. *Water Resources Research*, 24(8):1277--1282.

Mohammadi, S. S., Hezarkhani, A., and Erhan Tercan, A. (2012). Optimally locating additional drill holes in three dimensions using grade and simulated annealing. *Journal of the Geological Society of India*, 80(5):700--706.

Montgomery, D. C. and Runger, G. C. (2002). *Applied Statistics and Probability for Engineers, 3rd Edition*. Wiley.

Mory, D. F. M. and Deutsch, C. V. (2006). A program for robust calculation of drillhole spacing in three dimensions. *CCG Annual Report 8, Paper 309*.

Naus, T. (2008). Unbiased lidar data measurement (draft).

Pilger, G., Costa, J., and Koppe, J. (2001). Additional samples: Where they should be located. *Natural Resources Research*, 10(3):197--207.

Pinto, F. A. C. (2015). Guide to data spacing and uncertainty. *CCG Guidebook Series, Volume 19*.

Pinto, F. A. C. and Deutsch, C. V. (2014). Calculating the proportional effect based on the univariate distribution. *CCG Annual Report 16, Paper 407*.

Postle, J., Haystead, B., Clow, G., Hora, D., Vallée, M., and Jensen, M. (2000). CIM standards on mineral resources and reserves definitions and guidelines.

Pyrcz, M. J. and Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford university press.

Rojas, O. and Cáceres, A. (2011). The use of conditional simulation for drill hole spacing evaluation and decision-making in Telégrafo Project, Northern Chile, in Proceedings Eighth International Mining Geology Conference 2011. *The Australasian Institute of Mining and Metallurgy: Melbourne*, pages 443--452.

Ross, S. M. (2010). *A First Course in Probability*. Pearson Prentice Hall.

Rossi, M. E. and Deutsch, C. V. (2014). *Mineral Resource Estimation*. Springer Netherlands.

Rouhani, S. and Hall, T. J. (1988). Geostatistical schemes for groundwater sampling. *Journal of Hydrology*, 103(1):85--102.

Silva, D. and Boisvert, J. (2014a). Mineral resource classification: a comparison of new and existing techniques. *Journal of the Southern African Institute of Mining and Metallurgy*, 114(3):265--273.

Silva, D. S. F. (2015). Mineral resource classification and drill hole optimization using

novel geostatistical algorithms with a comparison to traditional techniques. Master's thesis, University of Alberta, Edmonton, Alberta, Canada.

Silva, D. S. F. and Boisvert, J. B. (2014b). Two new tools: Directional survey to gslib xyz format and drill hole spacing. *CCG Annual Report 16, Paper 404*.

Silva, D. S. F. and Deutsch, C. V. (2015). Multivariate data imputation using gaussian mixture models. *CCG Annual Report 17, Paper 104*.

Soltani, S. and Hezarkhani, A. (2013). Proposed algorithm for optimization of directional additional exploratory drill holes and computer coding. *Arabian Journal of Geosciences*, 6(2):455--462.

Soltani, S., Hezarkhani, A., Erhan Tercan, A., and Karimi, B. (2011). Use of genetic algorithm in optimally locating additional drill holes. *Journal of Mining Science*, 47(1):62--72.

Storck, P., Eheart, J. W., and Valocchi, A. J. (1997). A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogenous aquifers. *Water Resources Research*, 33(9):2081--2088.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. E. (2012). *Probability and Statistics for Engineers and Scientists*. Pearson Education.

Wang, D. and Murphy, M. (2004). Estimating optimal transformations for multiple regression using the ace algorithm. *Journal of Data Science*, 2(4):329--346.

Warren, J. E. (1983). The development decision: Value of information. *SPE Hydrocarbon Economics and Evaluation Symposium, 3-4 March, Dallas, Texas*.

Webster, R. and Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons.

Wilde, B. J. (2010). Data spacing and uncertainty. Master's thesis, University of Alberta,

Edmonton, Alberta, Canada.

Wilde, B. J. and Deutsch, C. V. (2009). Automatic determination of uncertainty versus data density. *CCG Annual Report 11, Paper 116*.

Zagayevskiy, Y. V. and Deutsch, C. V. (2011). Updated code for sensitivity analysis based on regression. *CCG Anuual Report 13, Paper 401*.

Zhang, Y., Pinder, G. F., and Herrera, G. S. (2005). Least cost design of groundwater quality monitoring networks. *Water Resources Research*, 41(8).

Appendix : Software

The software were developed during the case study and proposed methodology for data spacing and uncertainty, discussed in this thesis. The GSLIB codes cover ore and waste SMU management, uncertainty assessment in different production scales and sampling from a grid to generate new sets of drill holes.

The new codes support modellers to better understand uncertainty in their data spacing studies, giving more options when incorporating economic analysis. These tools go beyond the simple management of realizations, they provide users the possibility to quantify uncertainty in the economic variables based on any production scale, for a fixed volume or not. The parameter files are explained below and a brief explanation of the calculations is given.

Ore and Waste Indicator (*Owind*) program assigns to all SMUs an indicator for ore (1) and waste (0) based on the equivalent grade and cut-off. The number of variables to be used in the equivalent grade calculation do not need to be the same in all files, although the size (number of SMUs) must be the same. The equivalent grade is given by the summation of the variables given a multiplicative factor, power (exponent) and a constant. The use of an exponent is just for flexibility when calculating the equivalent grade.

As example, if the equivalent grade is given in Equation A.1 below, the parameters are: factor 1 = 0.25, power 1 = 2; factor 2 = -1.5, power 2 = 1; factor 3 = 1, power 3 = 1; and constant = 0.01.

$$EquivalentGrade = 0.25 \times Var1^2 - 1.5 \times Var2 + Var3 + 0.01 \qquad \text{(A.1)}$$

If a SMU has equivalent grade greater or equal than the cut-off it receives the indicator of 1 (ore), otherwise it is considered waste and gets 0 as indicator. There is no grid definition,

the program reads line by line from the input files, thus single or multiple realizations are read the same way.

The number of files are defined in line 4, for each file a different trimming limits can be applied, lines 5-9. Variable factors and powers are defined as they are declared, thus variable 1 is the first variable from the first file and so on. The constant is given in line 14, if no constant is applied to the equivalent grade, set its value to zero. The cut-off and output file are in lines 15 and 16 respectively.

The output file from *Owind* is used as one of the input files in the other two codes, *PRSR* and *SMU_unc*.

```
1   Parameters for Owind
2   ********************
3  START OF PARAMETERS:
4  2                            - number of files to manipulate
5  ./inputs/au_smu.out          - file with first variable
6  -998   1.0e21                - trimming limits
7  1  1                          - number of variables and columns
8  ./inputs/cu_ni_smu.out       - file with second variable
9  -998   1.0e21                - trimming limits
10 2  1  2                       - number of variables and columns
11 1   3                        - Var_1 (factor1; power1)
12 2   2                        - Var_2 (factor2; power2)
13 -1  2                        - Var_3 (factor3; power3)
14 0.1                          - constant
15 0.25                         - cut-off
16 ./outputs/owindicator.out    - output with ore/waste indicator
```

Production Scale Resources (*PRSR*) calculates the resources for every production volume. This code can be better understood as similar to block average in which some statistical calculations are done. The ore/waste indicator file from *Owind* is used as input, since this code needs the variable values in ore SMUs and also calculates some statistics for waste.

The grids must match, there is no calculation of fraction of blocks. All input files must be in binary format, Barnett and Deutsch (2014), the outputs are in ASCII format.

The production scales cannot overlapped, all SMUs falling inside the production volume are used in the calculations. Besides the resources, as expected value of tonnes of ore, this code calculates the expected grade of all variables, the expected quantity of metal and their standard deviation and probability to be around a percentage of the mean. The expected tonnes of ore are also calculated by lithology if the categorical model is given as input. A production volume that has no ore will have the standard value of -999 for expected grade of a variable, although the quantity of metal is set up to zero. Moreover, the expected grade of a variable is weighted by the tonnes of ore.

The ore and waste file is given in line 4, the number of variables is defined in line 5. In line 6, besides the ore/waste variable, the columns of the economic variables must be given, the equivalent grade is not an input for this code.

The input grid refers to the SMU grid, lines 8 to 10. The options for a file with the SMUs densities are in lines 11 and 12, this file must have the same size of the other inputs, hence the same number of realizations. If this file is not provided, then a standard density will be used instead, in line 13.

Users have the option to average SMUs that are not adjacent (regular grid), given a file with the different production scales indicator for all SMUs, line 14. All indicators (their integer IDs) are read from the first realization if multiple realizations are used, there is no need to specify in the parameter file the many production volumes IDs, it is expected that every SMU have an indicator value to identify the production volume it belongs to. The option for multiple realizations or single model is given in line 16. If multiple realizations, then this file must have the same size of the other inputs.

If the file with categorical realizations is given in line 17 then the number of categories and their integers must be provided in line 19. This file must have the same size of the other inputs, in the same grid scale defined for the input (in lines 8 to 10).

The statistical parameters are given in lines 20 and 21. If the number of quantiles is set to zero, then no probability is calculated and written in the output file. Cumulative distribution functions that cannot be calculated receive a value of zero. Start and finish realizations are set in line 22, if input files have only one realization, then the summary file is not written.

The output grid refers to the regular production scale grid if no irregular volume production was input in line 14. The summary output file contains the average (expected) calculations over all realizations, whereas the detailed output file keeps the information of all realizations, lines 26 and 27.

The block index option, in line 28, will add additional three columns to the output files with the indices i, j and k in GSLIB standard format for all production volumes, with the block with index i = j = k = 1 as the first block on the bottom left corner of the grid.

```
1        Parameters for Production Scale Resources
2        ****************************************
3  START OF PARAMETERS:
4  ./inputs/owind.gsb   - input file
5  4                     - number of variables
6  1 3 4 5               - columns for ow ind and variables
7  -998   1.0e21        - trimming limits
8  4   2.0   4.0        - Input size:  nx,xmn,xsiz
9  4   2.0   4.0        -              ny,ymn,ysiz
10 4   2.0   4.0        -              nz,zmn,zsiz
11 density.gsb          - density file
12 1                     -  column for density variable
13 3.00                  -  density (if no density file is given)
14 nofile                - irregular production indicator file
15 1                     -  column for volume production indicator
16 1                     -  1=multiple realizations, 0=single model
17 ./inputs/sis.gsb      - categorical variable realizations file
18 1                     -  column for categorical variable
19 4 200 300 400 500     -  number of categories, categories
```

```
15.0                    - probability to be within % of mean
2 0.1 0.9               - number of quantiles, probabilities
1  100                  - start and finish realizations
2   4.0   8.0           - output size: nx,xmn,xsiz
2   4.0   8.0           -              ny,ymn,ysiz
2   4.0   8.0           -              nz,zmn,zsiz
./outputs/summ.out      - output with summary volume production
./outputs/detai.out     - output with detailed volume production
1                       - output block index (0=no, 1=yes)
```

SMU Uncertainty (*SMU_unc*) has options similar to *PRSR*. Whereas *PRSR* will average all SMUs inside production scales that do not overlap, *SMU_unc* will calculate the uncertainty in the production scale at a SMU resolution. The uncertainty in a SMU is calculated considering all blocks inside the production scale volume (the window), centred in that SMU. The uncertainty calculates in each SMU is then based on the surrounding blocks, and not on the whole fixed production volume from *PRSR*. All input files must be in binary format, although the outputs are in ASCII format.

The parameter file for *SMU* is the same of *PRSR*, however there is no option for irregular volume productions. The production scale size is not given by number of blocks and block sizes (the regular grid format), users have to input the volume size by x, y and z length, in lines 20 to 22. If the SMU size is 12 m x 12 m x 12 m and the production volume is composed by 9 SMUs in a cube (3x3x3 SMUs), then the production size to be input in the parameter file is xsizo = ysizo = zsixo = 36 m.

The block index option for this code can output the standard format i, j and k, and also the SMU number, following the GSLIB standard format, with the SMU with index i = j = k = 1 as the first block on the bottom left corner of the grid, in lines 25 and 26.

```
       Parameters for SMU Uncertainty
       ******************************
START OF PARAMETERS:
```

```
4  ./inputs/owind.gsb    - input file
5  4                      - number of variables
6  1 3 4 5                - columns for ow ind and variables
7  -998   1.0e21          - trimming limits
8  6   1.0   2.0         - SMU grid: nx,xmn,xsiz
9  6   1.0   2.0         -           ny,ymn,ysiz
10 1   0.5   1.0         -           nz,zmn,zsiz
11 ./inputs/density.gsb - density file
12 1                       -  column for density variable
13 3.00                   - density (if no density file is given)
14 ./inputs/sis.gsb       - categorical variable realizations file
15 1                       -  column for categorical variable
16 4 200 300 400 500       -  number of categories, categories
17 15.0                   - probability to be within % of mean
18 2 0.1 0.9              - number of quantiles, probabilities
19 1  100                 - start and finish realizations
20 6.0                    - Production size: xsizo
21 6.0                    -                  ysizo
22 1.0                    -                  zsizo
23 ./outputs/summ.out     - output with summary volume production
24 ./outputs/detai.out    - output with detailed volume production
25 1                      - output block index (0=no, 1=yes)
26 1                      -  If 1, block index (0) or number (1)
```

Sampling from grid programs can be easily written in languages other than Fortran. The advantage of using Grid Sample, *Gsample*, is the fast alternative and better handle of large files that Fortran provides. Users have the option to sample any realization from a file and also sample from a specific region inside the grid. Azimuth and dip were also added as option, regular and irregular sampling are available as well.

There is no interpolation in this code, if a sample falls anywhere inside a block, then its value is given by the block value.

The sampling grid will define the volume to be sampled, it is not given by number of blocks and block size, instead it is given by minimum and maximum values of coordinates. Users have then the option to have the collar of a drill hole starting from the top of a block or from inside of it. As example, in GSLIB the bottom and top of the grid given in Z by nz = 10, zmn = 2.5 and zsiz = 5 are respectively 0 and 50. The user that wants a new drill hole with collar equal to 50 should then set the zmax option to 50. If the user wants the collar to be in the first block of the top of the grid, but not at the top of that specific block, one should set up zmax to a value less than 50 and greater than 40. The collar will coincide with the center of the block if zmax is set up to 45.

To clarify, a sample that falls at the top of a block belongs to the block below. In the example above, a sample that falls in z=40 belongs to the second block from the top and not to the first one. This is also valid when sampling blocks from left to right, a sample that falls between two blocks belongs to the right side block.

Users can sample as many files as they want, in line 4. The files to sample and the variables options are given in lines 5 to 8. The realization to sample is in line 9, choose one if the input files are from a single model.

The input grid is given in GSLIB format, lines 10 to 12. The option for regular or random sampling is in line 13, if random sampling is chosen then an extra option is provided to assure a minimum spacing between the new drill holes created, otherwise they are created completely random and more than one drill hole can sample the same block. If regular grid, users have to specify the spacing in X and Y directions and also fix the collar of the first drill hole. All other drill holes will be created respecting the spacing in X and Y. The first drill hole must be inside the sampling region, lines 14 and 15. If irregular sampling, the maximum number of new drill holes must be given, line 16.

Azimuth and dip are the same for all new drill holes, regardless of the option for regular or random spacing, in line 17. Sample spacing downhole, or the distance to sample in direction Z, is defined in line 18. New drill holes can have their IDs starting from the number defined in line 19. The drill hole coordinates will have the precision given in line 20, as the input

grid informations are read in double precision, this will avoid coordinates to have lots of decimal places.

The sampling grid defines the region inside the input grid to sample, lines 21 to 23. If users want to sample the whole input grid, then set them minimum and maximum values in order to cover exactly the input grid dimensions. To sample only the center of the blocks, in the example given above, set up the zmax = 45 and sample spacing downhole to 5. In other words, to sample only the center of the blocks, choose the sampling distance equal to the block size and set up the minimum and maximum coordinates to coincide with the center of the blocks from the input grid.

Output file is defined in line 24. There are two options for a keyout file, lines 25-27. This keyout file has the same size of the input grid, the same number of blocks. The option 0 (input grid inside sampling grid) will assign the indicator of zero for all blocks of the input grid that are outside the sampling grid and the indicator of one for blocks inside the sampling region. The option 1 (input blocks with assays) will output the indicator of zero for all blocks that were not sampled and the indicator of 1 otherwise.

The extra option defines the minimum and maximum spacing in X and Y (collar) directions when a random sampling is used. This option can be used to avoid sampling a block more than once.

```
1   |       Parameters for gsample V2.000
2   |       ****************************
3   | START OF PARAMETERS:
4   | 2                   - Number of files to sample
5   | ./inputs/smu1.out   - File with first grid
6   | 1  1                 - Number of variables and columns
7   | ./inputs/smu2.out   - File with second grid
8   | 1  1                 - Number of variables and columns
9   | 1                   - Realization to sample
10  | 100   5.0   10.0    - Input grid: nx,xmn,xsiz
11  | 100   5.0   10.0    -             ny,ymn,ysiz
```

```
50    5.0   10.0   -              nz,zmn,zsiz
0                  - Sampling spacing (0=regular, 1=random)
25  25              -  If 0, spacing in X, Y
5.0  5.0         -  If 0, X and Y of the first sample
10  69069           -  If 1, number of DHs and seed number
0  90              - Azimuth and dip
5                  - Sample spacing downhole
120                - Starting drill hole ID for new drills
2                  - Decimal places (precision, up to 6)
5   95             - Sampling grid: xmin, xmax
5   95             -                ymin, ymax
5   45             -                zmin, zmax
./outputs/new.out  - Output file with new drill holes
0                  - Output keyout file? (0=no, 1=yes)
0                   - If keyout (0 or 1, see notes)
./outputs/key.out   - Keyout file
-----------------------------------------------
KEYOUT:
0=input grid inside sampling grid, 1=input blocks with assays
OPTIONAL Extra Options for random drill sampling
         (sampling option 1)
NOTE: if extra option is used sampling will only occur
      in middle of blocks
0       - Use Extra Options (0 = NO, 1 = YES)
15  15  - Minimum spacing in x and y direction (dont use 0)
-----------------------------------------------
```

The following code transforms a 3D grid to a plan, for a single model or multiple realizations. The final number of blocks is then reduced to the number of blocks in X and Y directions, *nx* and *ny* in GSLIB format. All calculations are done vertically in the model. This code calculates the stripping ratio, economic value, predominant category, expected

drill hole spacing, and the expected grade of variables. The predominant rock type is the mode of all indicators. The files with categorical model and drill hole spacing are optional.

Stripping ratio, economic value and expected grade are calculated from the ore/waste file from *Owind*. The expected grade is not weighted by any other variable such as tonnes of ore; all SMUs are assumed to have the same density. Regarding the economic value calculation, there is one single value given to all waste SMUs and a weight given to all ore SMUs. The economic value, Equation A.2, is given by the summation of all waste and ore SMUs with their weights. All waste SMUs receives the same value, whereas the economic values of the ore SMUs depend on their equivalent grades.

Consider the simple example with only two blocks, one ore and one waste, with the equivalent grade of the ore equal to the cut-off of 0.25. Using the value of -1 to waste and the weight of 4 to ore, the economic value calculated for these two blocks would be zero (0), -1 plus 4 times 0.25. The economic value can be standardized with the maximum value throughout the grid.

$$Value = WasteValue \times \sum^{n} SMUs_{waste} + W_{eq.Grade} \times \sum^{n} EqGrade_{ore} \quad \forall \quad 0 \leq n \leq nz \tag{A.2}$$

Ore and waste file is given in line 4, variable definitions are in lines 5 and 6. The economic value for waste and the weight of equivalent grade are defined in line 7, option for standardize the output in line 8. File with categorical model and their integer codes are in lines 9 and 10, this file can have one or multiples realizations, defined in line 12. It is not expected multiple realizations of drill hole spacing, in line 13 and 14. Only the number of blocks from the input grid is required, line 16. The realizations for calculations and output file are given in lines 17 and 18.

```
1         Parameters for 3d2plan
2         **********************
3  START OF PARAMETERS:
```

```
./in/owind.out    - File with ore/waste indicator
1  2               - Ore/waste and equivalent grade columns
3  3  4  5         - Number of variables and columns
-1  4              - Value for waste and equivalent grade weight
1                  - Standardize economic value? (1=yes, 0=no)
./in/sis_smu.out - Categorical model file (optional)
3  100  200  300   - Number of categories and their integer codes
1                  - Variable column
1                  -  1=multiple realizations, 0=single model
./inputs/dhs.out - Drill hole spacing file (optional)
1                  - Variable column
-998   1.0e21     - trimming limits
100  100  10      - Input grid: nx,ny,nz
1  100            - start and finish realizations
./out/3d2plan.out - Output file
```

Grid Moving Window program, *Gmwind*, calculates moving window statistics such as the mean and standard deviation. Fortran is a fast alternative to other software as Matlab and Python. It was written to calculate local statistics similar to those from *SMU_unc* code, the local mean in a SMU given a production volume (the window size). A block is considered inside the window only if at least its center is inside of it.

The number of files, number of variables and their columns are defined in lines 4 to 8, these files must have the same size, although the number of variables used in calculations can be different in them. The input grid, in GSLIB standard format, is given in lines 10-12. The realization for calculation, if multiple realizations files are given, is set up in line 13, use 1 if the input files are single models. This code does not perform calculations through all realizations, users have to choose the realization they want.

The window size is defined in terms of total length of X, Y and Z directions, line 14. Output file in line 15.

```
       Parameters for gmwind
       *********************
START OF PARAMETERS:
2                           - Number of files
./inputs/smu1.out           - File with first grid
1  1                         - Number of variables and columns
./inputs/smu2.out           - File with second grid
1  1                         - Number of variables and columns
-998   1.0e21               - trimming limits
100   5.0   10.0            - Input grid: nx,xmn,xsiz
100   5.0   10.0            -             ny,ymn,ysiz
50    5.0   10.0            -             nz,zmn,zsiz
1                           - Realization for calculations
30  30  30                  - Window size: xsizo,ysizo,zsizo
./outputs/mov_wind.out      - Output file
-----------------------------------------------
```