Quantification of geological uncertainty and mine planning risk using metric spaces

Author:
Heidari, Seyed Mehrdad

Publication Date:
2015

DOI:
https://doi.org/10.26190/unsworks/18111

Quantification of Geological Uncertainty and Mine Planning Risk using Metric Spaces

BY

S.Mehrdad Heidari

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

School of Mining Engineering
The University of New South Wales
Sydney, Australia

March 2015

This thesis is dedicated

to my loves

Ryan & Rena

Acknowledgements

I am deeply grateful to Dr. Chris Daly my supervisor and postgraduate Coordinator (Research) of School of Mining Engineering for his guidance and technical support throughout the research, and his feedback and encouragement in the completion of this dissertation.

I would like to thank my supervisor Prof. Gary Froyland, from the School of Mathematics and Statistics for accepting me as his student and giving me the opportunity to flourish and learn different aspects of Mathematics, optimisation, data analysis, data mining and their applications in mining. His valuable suggestions have been a constant source of motivation for me during the time of this research. He provided me with the guidance and support to complete this thesis. He not only trusted me with my research, but also challenged me during this process. I am thankful for the fruitful discussions, continual support and advice throughout this work. The studies discussed in this thesis would not have been possible without the support of him.

I would like to acknowledge Prof. Serkan Sydam, The previous Postgraduate Coordinator (Research) of School of Mining Engineering for accepting me in the school of mining, supporting and giving me the opportunity for doing my research in the school of mining engineering.

Special thanks to Martin Thompson the computational research support manager of the School of Mathematics and Statistics for the useful suggestions and technical helps that improved the most of computing works which was done on the University of New South Wales‘s super computer.

I would like to express my sincerest gratitude to my friends Alireza Kaboli, Mojtaba Bahaaddini and Hossein Masoumi for their help during the period of my study.

Last but not least, I would like to thank my parents and my wife for supporting and believing in me. They have always encouraged me to become a successful person. Without their supports, I would have not accomplished this.

Abstract

Major sources of financial risk for mining projects include geological uncertainty and uncertainty in future commodity prices, costs, demand levels and interest rates. Geological uncertainty is difficult to model as there are complicated spatial considerations which are not present in the other sources of uncertainty. Stochastic simulations are now the common approach to assessing geological uncertainty, and one of the most common practical methods of producing realisations is conditional sequential simulation. Conditional sequential simulation algorithms can create multiple realisations that honour the original histogram and covariance matrix. One of the shortcomings of the conditional simulation algorithms is that there is no parameter that can provide further information about high order statistics for generated realisations. By visually comparing the colour realisation images (if they are 2D), we can easily see uncaptured spatial differences; therefore, any possible dissimilarity or similarity between the realisations cannot be captured by descriptive geostatistics.

Distance computation as a technique to measure dissimilarity or similarity between images, objects, and models has received attention in recent years. This thesis presents a formal measure of dissimilarity for generated realisations by adapting the Kantorovich metric to the geostatistics context. We propose a new methodology for mapping the space of uncertainty by a distance function that is based upon a physically meaningful notion of dissimilarity between pairs of realisations.

We are able to quantify the dissimilarity of different realisations. In this framework, the pairwise dissimilarities between realisations can be used to make a relation or a precise mathematical structure between them, which can describe the variability of parameters on interest (for example, grade) inside the space of uncertainty. This method provides a powerful tool to address how realisations are connected to each other and how this connection (structure) can answer some controversial questions in geostatistical simulations. Besides, we can use our methodology to optimally subsample a large collection of realizations, and quantify how well this high-quality subsample represents the overall uncertainty of the collection.

Moreover, such quantification of the space of uncertainty makes it possible to compare the impact of changing the geostatistical parameters or even simulation algorithms on the space of uncertainty. Our methodology has the potential to consistently compare the output of different geostatistical simulation algorithms, such as SGS, sequential indicator simulation (SIS) and turning bands (TBS) simulation.

Furthermore, if we place any deterministic geological reserve estimation, produced, for example, by Kriging, inside the space of uncertainty, the method can easily reveal how dissimilar other realisations are to the estimated model. In other words, measuring how close global accuracies (different realisations) are to the local accuracy (estimation) is now possible.

Finally, the mining processes such as mine optimisation, open pit design and long term scheduling are only able to handle relatively modest numbers of realisations. It is difficult to say how many realisations are required to achieve a prescribed level of accuracy based on a very large number of possible realisations. This method has the ability to construct a collection of schedules (coming from generated realisations) so that the overall uncertainty is captured in a way prescribed by the user. We argue that this small set of candidate schedules produce more robust outcomes than schedules selected by other existing risk-based approaches.

Table of Contents

Chapter 1

1 Introduction

Chapter 2

2 Literature review

Chapter 3

3 Geostatistical Simulations

Chapter 4

4 Data Preparation and Simulations

Chapter 5

5 Metric spaces

Chapter 6

6 Description of Methodology

Chapter 7

7 Numerical results on geostatistical simulations

Chapter 8

8 Stochastic mine design and risk analysis using metric space

Chapter 9

9 Conclusion and future work

Appendix A

List of Tables

List of Figures

Chapter 1

1 Introduction

1-1 Motivation

The best motivation for this study may be found in this quotation from Martinez (2009, p.1),

> "*A new era is coming for the mining industry. An era where mine planners, mining engineers and mine analysts, not only ask themselves the question, 'What if ...?' when evaluating their respective mine but also want to know what is the effect of these 'What if ...?'*".

This means that assessing the uncertainty, namely risk analysis, has to be an inseparable part of any mining project. The nature of mining projects is extremely complex due to the reason that they have to deal with a large number of parameters, some of them having been made millions of years ago. This shows that any future prediction about mining projects cannot be free of uncertainty.

Resource estimation or any geological modelling which involves spatial modelling can be a good example to describe uncertainty. Uncertainty exists in any geological model, namely a block model, which is normally created based on exploration data (boreholes data set) collected from a deposit. As the volume of taken samples (collected data) is considerably smaller than the size of the deposit, implicit or explicit assumption in this modelling is inevitable. Thus, the uncertainty would always be a part of modelling. As there is a clear linkage between the geological information and the subsequent studies of geosciences (for example, pit optimisation and mine design activities), this uncertainty propagate into that level.

An example of this impact was presented on Valle (2000), showing that 60% of the mines surveyed had an average rate of production that was less than 70% of the designed capacity in the early years. Other researchers (Rossi and Parker, 1994) reported shortfalls against predictions of mine production in later stages of production.

Moreover, new approaches to uncertainty not only do not consider it as a problem (that should be avoided), but also take advantage of it to create opportunities and value. Case studies show that applying the

uncertainty approaches in mine optimisation and mine planning activities can increase the value production schedules, as well as make a bigger optimal pit size with a better NPV (Dimitrakopoulos, 2011).

Dealing with such an important parameter has to be a mandatory part of any geosciences assessment. As uncertainty is not avoidable and any future prediction is uncertain, serious problems may occur during evaluating projects if uncertainty is ignored. Consequently, there is a good potential to make value if uncertainty is assessed.

1-2 Uncertainty and Risk

Uncertainty may have different meanings in many aspects of engineering and science or even in lay terms. Although several definitions of uncertainty can be found in textbooks, the following definition, which is considered causal, may be more effective. "*Uncertainty is caused by an incomplete understanding about what we like to quantify*" (Caers, 2011, p.39). The quantifying or measurement of any sort of incomplete knowledge has to deal with probability; therefore, the measurement of uncertainty has a probabilistic basis, which, in turn, is associated with the relevant degree of uncertainty. Risk, in this case, is a state of uncertainty, namely a degree of uncertainty where there is a possibility of loss or any other undesirable outcome. Risk in this definition has probability. Uncertainty - or risk - would be associated with major components of mining projects, such as geology, mining, processing plant, transporting network and markets; in addition, each of these components can be broken down into sub-components, which also contain uncertainty or risk.

Risk and uncertainty, which is addressed in this thesis, is mostly related to the modelling of uncertainty in geology and mine planning activities. This modelling or procedure of forecasting any interested parameters or even making any decision can be divided into three general classes: measurement and collecting data; making a model; and predicting the parameters from the model. Generally, each step of this procedure is subject to uncertainty, and thus would be at risk of not meeting financial and technical targets.
There are a few methods to deal with risk. However, it has to be mentioned here that any attempts to model or quantify the uncertainty (independent of what the methods or approaches are), would inevitably face the issue that it is difficult to ascertain out whether the modelled uncertainty is correct or not. This means that "*there is no true uncertainty, there is are only the models of uncertainty*" (Caers, 2011, p.6).

The following example may make it clearer. Assuming there is a geological block model, to which the indicator Kriging is applied; consequently, the probability of it being ore or waste is known for each block.

If the original probability of a block being ore is 80% and 20% of it being waste 20%, after digging, there are just two options for this block, that is, either it is or ore or waste. If it is ore it cannot be said that the probability of 80% of it being ore block was correct.

1-3 Sources of uncertainty in mining projects

The mining industry requires reasonable evaluation of the characteristics of deposits. For example, resource estimation, which is a fundamental study for any mining project, requires the information of grade, impurities, density, rock type, thickness and geological structures of deposit. However, these processes involve various sources of uncertainty, which are explained here.

The main source of uncertainty can be generally classified into two major groups (Caers, 2011): uncertainty due to process randomness; and uncertainty due to limited knowledge of the process.
Most of the uncertainty which is relevant to geosciences can be classified into the second group. It comprises the following sub-groups (which are relevant to the topic of the thesis) roughly explained below.

1. The first sub-group is geological uncertainty, as described here

    - The first one in this sub-group is the rock properties (attributes) or any other measurable parameters in deposits, such as grades, impurities and density.

    - The second one is the structural uncertainty (especially for coal mining), the bounding surface representing layers, deposit shapes, the shape of faults, folds or any other structural controller and how many of them are involved.

    Geological uncertainty is more difficult to model as there are complicated spatial considerations which are not present in other sources of uncertainty. This type of uncertainty plays an important role to quantify uncertainty in mining projects as these parameters are a base for all financial or technical assessment.

2. The second sub-group is financial uncertainty (risk). Major sources of financial risk include uncertainty in financial parameters, such as future commodity prices, costs, demand levels, interest rates and exchange rate, which may vary considerably over time. This uncertainty has a high impact

on the financial parts of the project. These risks are in some sense easier to quantitatively account for when assessing project risks, as financial parameters can be modelled as time series in a variety of ways. This type of uncertainty is generally addressed by the financial markets theory, which is beyond the scope of this thesis.

3. The third sub-group is uncertainty in any kind of geological interpretation by experts which may impact on the making of geological models. Having different geological interpretations even for a same geological feature is quite common in geosciences fields due to the insufficient and sometimes indirect data available.

4. The next sub-group is uncertainty in modelling, where modelling processes have a set of assumptions often made around simplicity, availability and utilisation, which are not essentially based on reality; therefore, modelling cannot typically consider all possible factors that may have an impact on predicting interested parameters, so the model(s) are unable to capture all features of reality, consequently making it uncertain.

5. The last sub-group is uncertainty in measurement. Any kind of measurement, for example a grade analysis, would have uncertainty regarding systematic error (classified as *validity issues*) and non-systematic error (classified as *reliability issues*, which is called the random error).

Most of this thesis addresses grade uncertainty in the first sub-group; the other sources of uncertainty are not discussed in this study.

1-4 Deterministic methods in geosciences and its limitations

Uncertainty in the majority of measurable properties can be expressed by the variability in value or probability. That is, random variables have to be taken into account instead of fixed parameters. Methods or solutions which do not consider any variability in measurable properties can be classified as deterministic approaches.

Deterministic methods have been established, developed and applied for modelling of the mineral deposits, optimisation, mine planning, valuation and decision-making for several decades.
These methods generally work based on a single estimation or model, be it in financial or technical terms. These estimations or models calculate the expected value of the interested parameters accurately, and,

therefore, ignore all possible variation that the parameters may have, namely variability of values. In many cases, this variability can be illustrated by the distribution of the values or variance. Figure 1-1 illustrates the classic example of the flaw of averages ignoring the variability of depth involved when somebody drowns in a river with 3 ft. average depth. That is, " *plans based on average assumptions are wrong on average*" (Savage and Danzi, 2009, p. 11).

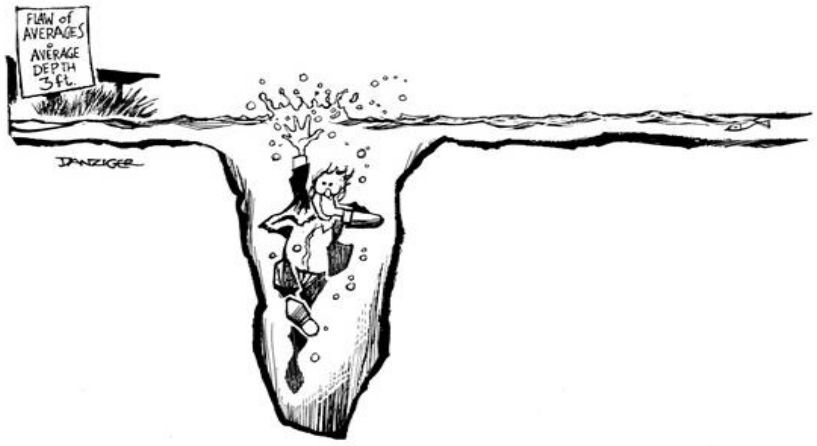


Figure 1-1: Ignoring the variability of depth and relying on its average may present a serious problem when crossing a river with 3 ft. average depth (Savage and Danzi, 2009)

The Kriging method could be the best example for this drawback of making use of an average in the mining industry. Kriging estimates the average grade of a block by weighted average of sample grades around the block. One consequence of considering only the average grade for an estimated block is that it kills the real variability, which is a part of block estimation; in addition, it creates a problem called the smoothing affect, which is explained in chapter 3.

The open pit optimisation can be another example of the smoothing affect. In general, pit optimisation is carried out on a Kriged block model to maximise the Net Present Value (NPV). Figure 1-2 shows the results (NPV) of this optimisation on the Kriging model against simulated models, namely realisations (as a non-deterministic method), which clearly illustrates the two following points.

First, simulations are able to cover a broad range of the project's NPVs in the form of probability of distribution, while only a single estimation derives from the conventional method. The second point is the NPV outcome for the conventional method, which is higher than the ninety-fifth quintile of the NPV distribution. This means that there is a tendency in the conventional method to overestimate the project's NPV.

These two examples are good answers to the question whether considering uncertainty and risk may or may not improve the decision making process. Consequently, not considering uncertainty approaches and relying on the use of an average value approach may underline the failure of many designs in terms of the ability to meet any sort of project performance indicators.

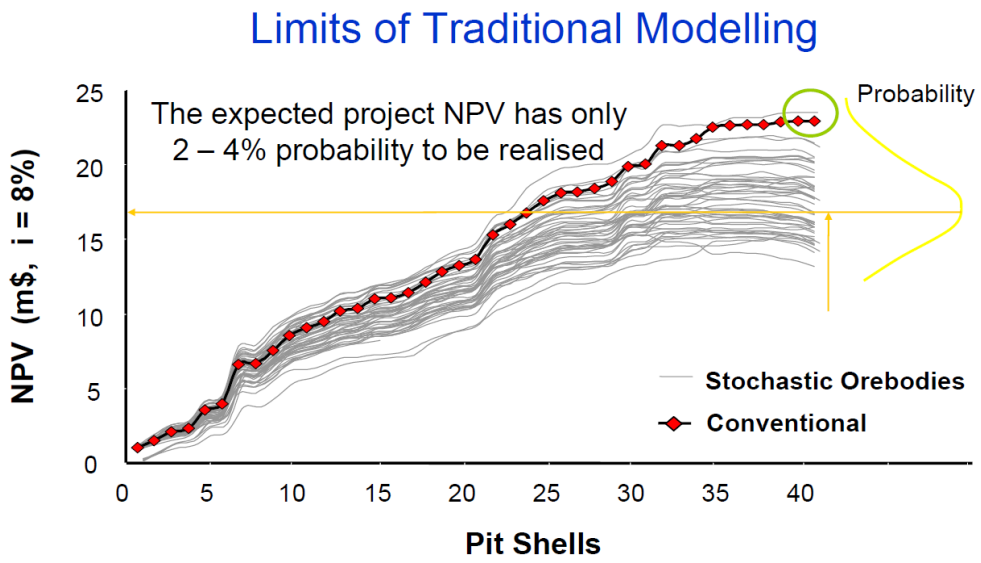


Figure 1-2: Optimisation of mine design in an open pit gold mine, NPV versus "pit shells" and risk profile of the conventionally optimal design (Dimitrakopoulos, 2011)

1-5 How to deal with uncertainty in geosciences

Stochastic conditional simulations (CS) (Deutsch and Journel, 1998), which are a part of Monte Carlo simulations, are normally used for modelling uncertainty in spatially distributed variables. The geostatistical simulation algorithms can generate equally probable realisations, which are able to mimic the in-situ orebody grade variability. By producing many conditional simulations (realisations) of a single geological reserve, a naive but instructive assessment of financial risk due to geological uncertainty would be to evaluate a putative mine exploitation plan against each conditional simulation, thus obtaining a spread of financial outcomes in terms of net present value, for example. The spreads of financial outcomes of several competing mine plans may be assessed in this way. Thus, the main advantage of conditional simulations over, say, a single deterministic geological reserve estimate, produced for example, by Kriging, is an estimate of the geological uncertainty and concomitant financial risk, which a single geological estimate cannot possibly provide.

Figure 1-3 illustrates open pit mine design stages from the geological model to the mine scheduling based on the conventional method and the uncertainty based approach, respectively. As all input variables in the conventional mine optimisation process and the mine planning activities are expected values, all project performance indicators would be represented by single values; while in the uncertainty based approach all project performance indicators are in the form of probability distribution function (pdf).

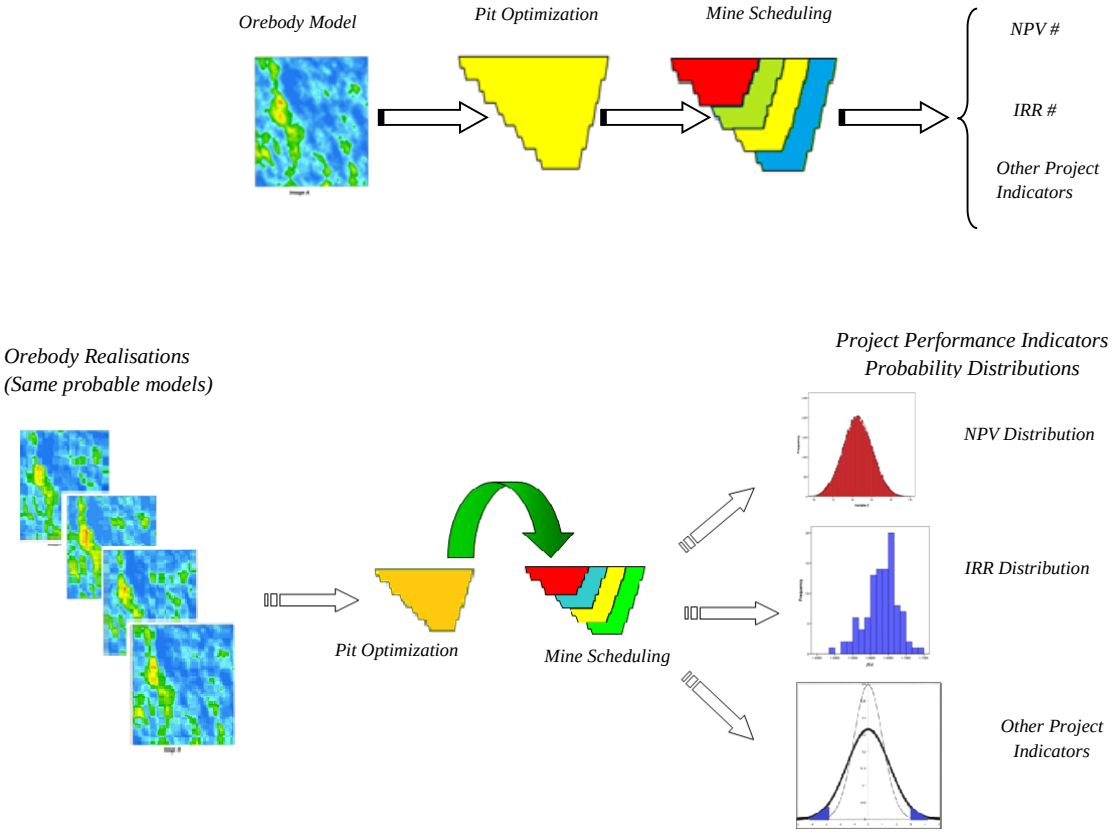


Figure1-3: Conventional (deterministic or single model) approach versus risk-based method for mine planning

1-6 Objective of the thesis

An open pit mine design and production scheduling is a complicated and difficult problem to solve regarding its scale and the several uncertain parameters involved. The objectives of mine planning activities are to maximise the NPV while production targets are satisfied. One of the most significant parameters affecting the optimisation and production scheduling is grade uncertainty. A set of simulated geological

models (generated realisations) provides a quantified description of the grade variability. These generated realisations can make a space, which is called the space of uncertainty. Quantifying the space of uncertainty by a notion of dissimilarity or “distance" between generated realisations is a key ingredient in this thesis.

Distance computation as a technique to measure dissimilarity and similarity between images, objects, attributes, observations and models has received attention in recent years. Many of the techniques used to measure dissimilarity try to find the best notion of distance for measuring dissimilarity between objects or models.

This study presents a formal measure of dissimilarity for generated realisations by adapting the Kantorovich metric (which is the physically meaningful notion of dissimilarity between pairs of realisations) to the geostatistics context. A new methodology is proposed to address the following different conceptual subjects in geostatistical simulations; and to find the solution for those, first, and then to apply this methodology on a practical application in mine design for doing risk assessment. Therefore, the focus is on the concepts and methods rather than on the detailed applications in items 1, 2 and 3; item 4 includes the application of the method.

The objectives of the thesis can be classified into four categories. It should be mentioned here that the proposed methodology may find different applications in other fields of geosciences and it is not limited to geostatistical simulations, mining planning or what are addressed as the objectives of this thesis below.

The four main objectives of the thesis are:

1. *Mapping the space of uncertainty* by a distance function that is based upon a physically meaningful notion of dissimilarity between pairs of realisations. We are able to quantify the dissimilarity of different realisations and to use this information for modelling and visualising the space of uncertainty. Consequently, such quantification of the space of uncertainty makes it possible to compare the impact of changing the geostatistical parameters on the space of uncertainty and to answer some controversial issues in geostatistical simulation, such as equal-probability of generated resolutions; likelihood; and how far the realisations are from local accuracy (Kriging model).

2. *Achieving realisation reduction* in order to select the sub-collection of realisations that best represent the possible outcome of stochastic simulation algorithms. The concept of Kantorovich distance has been used and a simple optimisation model has been developed to find the best

samples and to quantify how well this high-quality subsample represents the overall uncertainty of the collection. The optimisation model can be used as a general tool, which is able to select any subset of representative realisations against user defined criteria. For example, this methodology can determine the smallest number of conditional simulations that are required to cover 75% of the total geological uncertainty. Moreover, this approach identifies the corresponding conditional simulations.

3. *Evaluating of the stochastic simulation algorithms*, which is the set of realisations that can easily be generated by applying a variety of simulation algorithms that the user has to select, such as the sequential Gaussian simulation (SGS), the turning bands simulation (TBS) and the sequential indicator simulation (SIS). As the algorithms use different techniques or a random function (RF) model to generate realisations, in general it would be desirable to compare what they generate. Therefore, three common stochastic simulation algorithms are compared to evaluate their performance in the point of mapping the space of uncertainty without using any transfer functions.

4. *Assessing risk based mine design methods*. The common risk based methods usually generate different mine designs based on simulated realisations; consequently, these methods deal with several mine designs so that just one of them can be, in some way, selected. Finding the best criteria for this selection can be challenging for any approach. One of the drawbacks of the present approaches is the lack of a general classification method, namely, the clustering of the mine designs. In this section, by using the proposed methodology for mapping the space of uncertainty, it is possible to quantify the dissimilarity of different mine designs and to use this information to quantify how well the representatives or clusters represent the overall uncertainty to select the sub-collection of mine design that represents the best possible outcomes.

Chapter 2

2 Literature review

2-1 Introduction

The literature can be divided into two areas. The first is the concept and methodology of the approach. The other area of literature will focus on developing an application of the methodology that will be used for mine design activities. Due to this distinction, both parts of the literature review, which are detailed in this chapter are independent of each other to some extent.

First, quantifying the space of uncertainty is reviewed in the literature. Very few sources can be found on this issue, which may confirm that little attention has been paid to the definition of the space of uncertainty and related issues. The author is aware of no other pervious work, excluding the ones discussed here, that may tackle this problem. The largest part of the review is about geostatistical simulations, which address the space of uncertainty. The given references are classified into two groups, conventional and distance-based methods. Although our approach is classified into distance-based method as well, there are significant differences in the methodology between what has been addressed in the references so far and what is introduced in chapter 6.

Second, this thesis presents a formal measuring of dissimilarity between two extraction sequences of production schedules by adapting the distance-based method to the long term mine planning. Due to that, a brief review of the long term open pit mine production scheduling algorithms is presented here.

Fortunately, there are many mathematical programming models about optimising long term mine production scheduling. These mathematical models have been studied extensively in the literature since the 1960s. Algorithms reviewed here contain recent studies usually based on multiples realisations (risk-based methods) and also a few well-known methods that are based on a single geological model that is, deterministic or conventional methods.

2-2 A review of the state of the art of evaluating uncertainty in geosciences

Due to increasing computer performance, it is now possible to construct, for example, 1000 realisations for block models of size $N \approx 10^6$ in less than 48 hours on desktop PC using techniques such as Sequential Gaussian Simulation (SGS). This would be accomplished much faster if direct block simulation (Boucher and Dimitrakopoulos, 2009) is used. Due to this high speed computation, conditional simulation algorithms are commonly used by the mining industry and other geosciences fields to assess geological uncertainty.

However, almost all applications and methodology are applied on generated realisations to assess uncertainty without considering integral controversial issues in this type of simulations or mapping the space of uncertainty made by these type of geostatistical simulation algorithms. For example, in all of these applications, the multiple realisations are randomly sampled without knowledge about underlying probability space, commonly named as the space of uncertainty. In addition, there has been very little work, if any, on quantifying how well a given collection of conditional simulations represent the total, for instance, geological uncertainty.

Often, a number, say 30, is decided upon as the appropriate number of conditional simulations to produce and 30 conditional simulations are produced and used as a finite collection representing all possible outcomes. Several questions arise: Why was the number 30 chosen? Is 30 too few or too many? Have the first 30 conditional simulations produced the best 30 representatives? How well could 30 well-chosen conditional simulations represent all possible geological outcomes?

There are very few sources related to the impact of number of realisations, but just only on univariate and bivariate statistics, such as, mean, variance and covariance matrix (Deutsch and Journel, 1998). Their results, in turn, have confirmed nothing more than the expected ergodic fluctuations of realisations. Thus, all the above questions are still open. Other issues may be found in this quotation by Goovaerts (1999, p.163)

> "In *summary, contrasting with the increasing use of stochastic simulation in risk analysis, it appears that little attention has been paid to the definition of the space of uncertainty, and related issues such as the equiprobability of realizations, the equivalence of spaces of uncertainty generated by different algorithms, and the number of realizations required to sample this space. There is currently no theory that allows us to determine if the set of all possible outcomes is fairly sampled."*

Journel (1997) mentioned that for a simulation algorithm, equiprobability condition of generated realisations can be guaranteed if individual realisations are completely set off by one random seed drawn from a uniform distribution; however, each algorithm definitely samples a different subset.

Srivastava (1994) argued that generated realisations by one simulation algorithm cannot be re producible by the others; thus the simulation algorithm selection might be an important step for any study.

The controversial issues in these type of simulations are not limited to what has been mentioned here. Myers (1994) and Hu and Ravalec-Dupin (2004) are suitable sources where some conceptual issues of simulations are discussed in detail.

Mapping the space of uncertainty and its related issues had not received much attention in the literature before (Arpat and Caers, 2007; Suzuki and Caers, 2008). Previously, attempts at mapping the uncertainty space had been based on transfer functions responses (Gotway and Rutherford, 1993; Srivastava, 1996; Goovaerts, 1999; and Qureshi and Dimitrakopoulos, 2007). That is, these sources actually evaluated the space of uncertainty of response values, except for Srivastava, who did assess the actual outcomes of simulation algorithms.

These studies simply compared a series of responses created by transfer functions after applying the set of realisations which derive from different simulation algorithms. This comparison loses not only a significant amount of the underlying structure of the space of uncertainty, but also does not provide a general conclusion about them. This thesis briefly explains what could actually be found in the literature and divides the results into the two following categories, namely conventional and distance-based approaches.

2-2.1 Conventional approaches

Not much research has been found on the space of uncertainty and relative problems by conventional methods. Some of them, which address the comparison of stochastic simulation algorithm in mapping the space of uncertainty, are explained here.

Qureshi and Dimitrakopoulos (2007) and Gotway and Rutherford (1993)'s studies are similar to each other in relation to the stochastic simulation algorithms and transferred functions which were applied. The results of the former are brought here, as its findings are more recent.

In both these studies the performances of three different stochastic simulation algorithms, namely sequential Gaussian simulation (SGS), sequential indicator simulation (SIS), and probability field simulation (PFS) were studied in mapping the space of uncertainty. That was done by applying three following transfer functions with minimum cost path, threshold proposition and geometric mean on 100 conditional realisations with two different sample sets from exhaustive Walker Lake data set. The following results were obtained from Qureshi and Dimitrakopoulos' (2007) study: increasing the sample size improves the precision associated with the response distributions; uncertain distributions produced by SGS, SIS and JSGS are more precise than those based on PFS; sequential algorithms were found to perform well in mapping spaces of uncertainty; and that the ability for mapping the spaces of uncertainty depends on the complexity of the transfer function and that is not necessarily a well understood aspect of the modelling process (Qureshi, 2002).

Another study about comparing the spaces of uncertainty was conducted by Deutsch and Journel (1992). In their study, three unconditional simulation algorithms, a sequential indicator simulation (SIS), a sequential Gaussian simulation (SGS) and a simulated annealing (SA) were used with the same information and normal score semivariogram model. Regardless of the response variable, the results showed that all algorithms can generate comparable spaces of uncertainty but they may be different in the point of ergodic fluctuations in the histogram of their responses. For example, the response distribution of SA was slightly wider than SIS and SGS.

One of the best studies in comparing the space of uncertainty was carried out by Goovaerts (1999) bringing a new approach into the comparison of spaces of uncertainty by using the Principal Component Analysis (PCA). He displayed the set of realisations into the space defined by all characteristics together instead of looking at the space of uncertainty of each characteristic. His study, in the petroleum field (flow properties in a sandstone) was based on the comparison of four simulation algorithms: SGS, SIS, PFS and SA. He generated 100 simulations for each simulation algorithm and took a random subset from the set of generated realisations and responses variables by increasing the size of the subsets and then calculated univariate and bivariate statistics of each subset to check the impact number of realisations on them. He concluded by saying that "*the extent of the space of uncertainty increases with the number of realizations generated*", but the rate of that gets smaller after 20 realisations, in his case.

Results of this study, in that case, in the point of the applied response variable, showed that differences between the simulation algorithms are the most pronounced for long-term responses with SGS granting the most accurate prediction.

Perhaps the study of Srivastava (1996) is different from the others as transfer functions were not used in comparing the space of uncertainty. Although his example, as he described in his paper, is "*hopelessly small and simplistic*" (p.61), it has the advantage to generate all possible outcomes of the used simulation algorithms.

In his study, Srivastava (1996) made a fabricated case for simulations and then applied a few constraints to limit and reduce the number of possible realisations up to 276. That is, the space of uncertainty consists of 276 realisations that satisfied all of the constraints. Therefore, a procedure that equiprobably samples this space should generate any one of these realisations with a probability of $\frac{1}{276}$. Table 2-1 shows the least and the most frequently sample outcomes or realisations for each of the five simulation algorithms, namely SGS, SIS, Classic SA, Greedy SA and Fixed state SA for 10,000 realisations. As can be seen, each outcome, shown by a number, has different chances to occur. The author finally concluded that (p.65)

> "*If each of the 276 outcomes has an equal probability of being generated, then in a set of 10,000 realisations, each one should have a 99% chance of occurring at least at 18 times and not more than 54 times*".

Table 2-1: The least and the most frequent sample outcomes (realisations) for each of five simulation algorithms (Srivastava, 1996)

	Least common		Most common	
Case	Outcome Number	Frequency	Outcome Number	Frequency
Classical annealing	79	19	217	54
Annealing w/ fixed state	107	11	11	175
Greedy annealing	214	7	218	108
Sequential indicator	51	7	262	85
Sequential gaussian	40	20	184	58

The author clearly states that the equiprobable sampling of the space of uncertainty is of high importance and of critical concern for any study used to generate realisations for risk analysis.

As per Suzuki and Caers's study (2008), the conventional approaches in comparing the space of uncertainty generally followed these steps and, thus, generated many realisations by applying a transfer function and calculating response values followed by the evaluation of statistical studies on response values. The main disadvantage of these methods is that they cannot define any relation between the generated realisations; consequently, they are unable to find the structure of the space of uncertainty, a very important step toward

the interpretation of the outcome of simulation algorithms. This issue could be resolved by applying the distance-based method to measure dissimilarity between generated realisations.

2-2.2 Distance-based approaches

Distance computation as a technique to measure the similarity or dissimilarity between images, objects, databases and models has received attention in recent years. However, finding the right distance function that may be suitable for measuring the similarity in the field of interest would be challenging.

Distance can not only quantify how different the images, objects and models are from each other, but it can easily construe a structure between objects in a metric space. For example, distance may be a proper way to measure how similar or dissimilar the outcomes of simulation algorithms (generated realisations) are in a metric space and to reveal the relations between them, namely the structure, by calculating all pair distances between all generated realisations. That is, this approach can map the space of uncertainty in a metric space by a meaningful notion of dissimilarity between pairs of realisations by using this information for modelling and visualising the space of uncertainty. As a consequence, questions about sampling the space of uncertainty; clustering; evaluating the simulation of algorithms; relative locations between the random realisations; and the distance between the realisations from local accuracy (Kriging model) can be answered by this approach.

The concept of dissimilarity between multi-point geostatistical realisations has been exploited in the context of evaluating oil or gas reservoirs by Suzuki and Caers (2008) and by Scheidt and Caers (2009), opening a new era of uncertainty modelling in metric space by the distance-based method. The topic of selecting subsets of simulations originates from petroleum reservoir engineering, where generated realisations of a reservoir are used as input to reservoir flow simulators that forecast reservoir production performance. A petroleum reservoir simulator is essentially the numerical implementation of the multiphase flow process of water, oil and gas as per fluids laws. This method has also been used for oil reservoirs to model the uncertainty of channel facies or patterns modelled by them. In this method, the distance function, which has usually been used to measure the similarity between generated realisations of a reservoir, is Hausdorff distance (Dubuisson and Jain, 1994).

Figure 2-1 illustrates this method in a few steps. As can be seen, after calculating the pair distances between the set of S realisations (step A), the square symmetric distance matrix $S \times S$ (dissimilarity distance matrix) with zero diagonal elements is constructed (step B), and then the distance matrix is embedded into the

Euclidean space by applying the Multi-dimensional scaling method (MDS) (Cox and Cox, 2000). This method, including the embedding points (step C), plays a critical role for the next steps. The next step is to apply MDS, which involves the application of the Kernel method (Sarma, 2006) on the given Euclidean space in order to transfer the points into a high dimensional space. This technique enables the configuration of a better linear separation for the points in this space, therefore providing a better clustering result (step D). Consequently, the high dimensional space is again mapped into its original space to decrease the dimensionality (step E), which is called the pre-image problem. After that, a small collection of realisations which are able to represent all generated realisations is chosen as the final selection (step F). The application of the transfer function on a small subset of realisations allows uncertainty quantification (e.g., P10, P50, P90 quantiles) of the response variable.

Below are four major differences between the oil and gas reservoirs modelling and what is usually used in mining industries to explain why this methodology is not applicable in mining applications.

1. In mining, the model realisations come from variogram-based geostatistical simulation algorithms such as SGS and SIS, while in new oil and gas reservoirs these realisations come from training images and they are usually generated by multiple-point geostatistical algorithms, such as Single Normal Equation Simulation (SNESI) (Strebelle and Journel, 2001) and Filter-Based Simulation (FBSI) (Zhang et al., 2006).

2. In mining, the models or realisations are normally based on exact measurements, which are called hard data, such as drilling data or any sampled points; however, in oil/gas reservoirs these realisations are based on hard data and the data with uncertainty, which is called soft data such as, geophysical survey and seismic response.

3. In oil and gas reservoirs modelling, for example in these studies by Suzuki and Caers (2008), Arpat and Caers (2007) and Honarkahah and Caers (2010) the prior probability (determined from historical observation or all possible prior models) and posterior probability (after applying all data) are usually used to gather the Bayesian approach to assess uncertainty in posterior models, while in mining, it does not have any application. That is, determining the prior probability in mining is very difficult and sometimes impossible.

4. Oil/gas reservoirs modelling is distinctly different from optimising scheduling and forecasting a mine's production over its life. Firstly, this is not a physical process, as per the flow chart below.

Secondly, it is based on some of the methods available in the field of mathematical programming that is fundamentally different in both technical and numerical aspects than in finite elements and related methods needed for implementing the flow simulation.

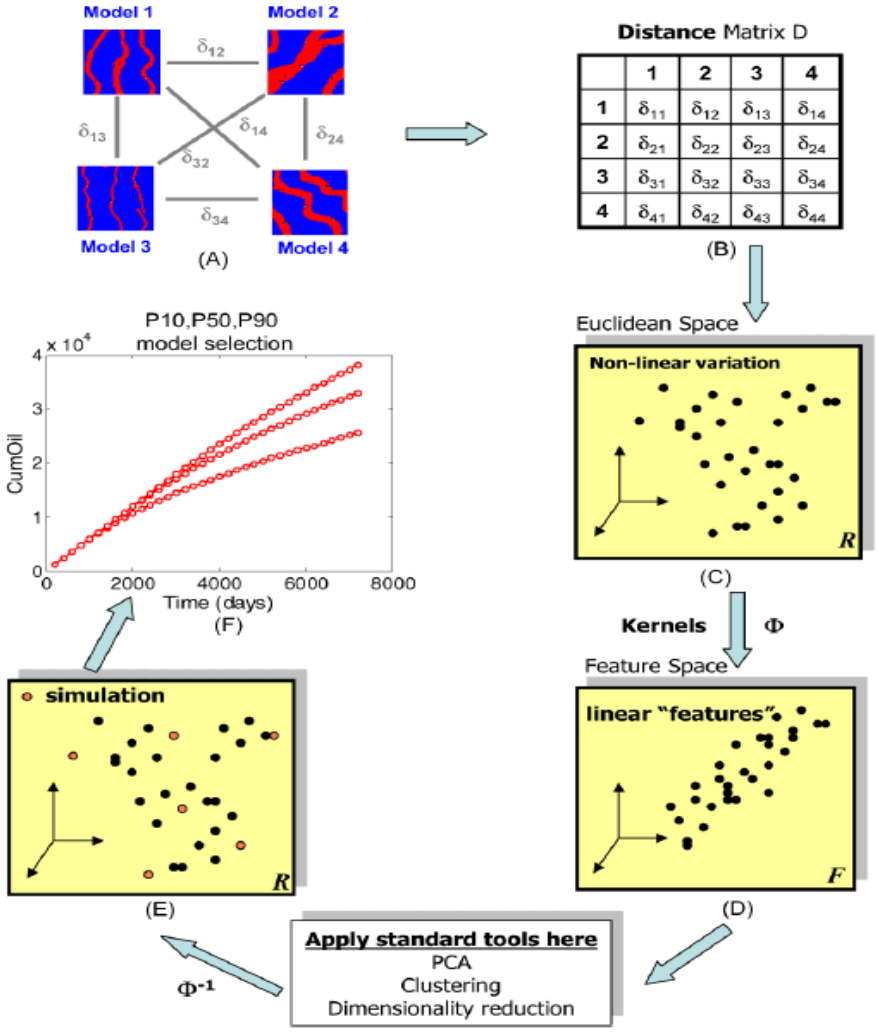


Figure 2-1: Scheidt and Caers' proposed workflow for uncertainty quantification-(A) distance between two models, (B) distance matrix *D*, (C) models mapped in Euclidean space, (D) feature space, (E) pre-image construction, (F) P10, P50 and P90 quantile estimations (Scheidt and Caers, 2009)

In this thesis, the Kantorovich distance has been introduced and applied with a different flowchart to assess the uncertainty, which is substantially different to what has been introduced for oil/gas reservoirs.

2-3 Long-term mine planning: A literature review

The open pit mine planning activities begin with this three-dimensional 3D geologic block model, which presents the structure of orebody (see Figure 2-2). An accurate prediction of the shape, location, size and specifications of a mineral deposit is required to present a reliable mineral reserve, as well as the technical and financial planning for exploitation. To achieve these goals, it is necessary to prepare a 3D model based on the shape, distribution and mineral species for any given mineral deposit (Glacken and Snowden, 2001).

These blocks are assigned different numbers over open pit mine planning activities such as, Rock Properties (Grade), by applying any resource estimation method, for example Kriging, and the economic value ($), which can be calculated by the following formula for individual blocks, and the sequence (#) or time (*t*), by using any pen pit mine production scheduling algorithm:

Block Value (BV) = Revenues- Costs

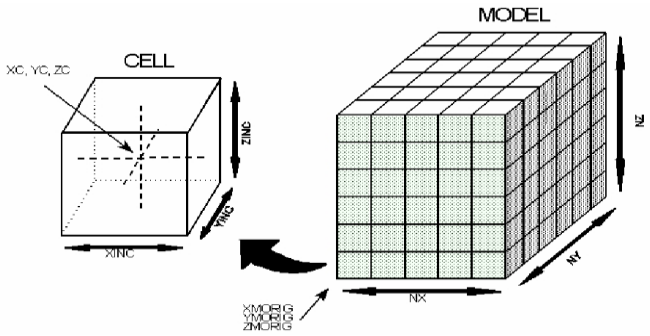


Figure 2-2: 3D geologic block model which presents the structure of orebody

Long-term open pit mine planning activities is a general term for calling all mathematical solutions, namely optimisation problems, which might be used to find the best block extraction sequence that is able to get the maximum Net Present Value (NPV) (Whittle 1989). This optimisation process has to consider a variety of practical and financial constraints of the project. The results of this optimisation problem, whatever it would be, have a high impact on the feasibility and profitability of mining projects.

As the mining industries current go deeper with lower grade ores, mine planning is becoming a key factor that can result in ceasing operations or continuing the project (Osanloo et al., 2008).

There are several mathematical solutions to optimise open pit mines as the nature of an open pit mine optimisation problem can easily be modelled by mathematical models However, for any method that might be used the following questions should be answered (in the following order):

- Which block has to be mined (including ore and waste)
- When it has to be mined
- Where it has to be sent (process lines or waste dump)

Answering the above questions can define the mining strategy and determine how much income a mine would have during its life. It can be seen that there are several different approaches that directly depend on how the decision is made for each of the blocks. Nevertheless, most approaches can be divided into two groups, namely, the deterministic approach which assumes all inputs have fixed known values; and the uncertainly-based approach which accounts for variability in some data, for example, ore grade, future product demand and future product price.

A few well-known deterministic and uncertainty-based algorithms have been reviewed to shed light on the location of the proposed methodology to deal with this problem. The advantages and disadvantages of these algorithms are briefly explained.

2-3.1 Deterministic approaches

Traditional methods of mine planning involve the creation of a schedule using information from a single deterministic block model. Often the schedule is optimised to maximise net present values (NPV). Plenty of research has been conducted on the mine planning problem since 1965 and several types of mathematical models have been considered for solving the problem. These mathematical models can be divided into the following classes, each class containing sub-classes:

- Linear programming (*LP*)
- Mixed integer programming *(MIP)*
- Integer programming *(IP)*
- Dynamic programming *(DP)*

2-3.1.1 Linear programming (LP)

Johnson (1969) introduced the following formulation to optimise mine scheduling using an LP model. The mathematical form of target function, which is subject to mining capacity, processing capacity and maximum and minimum of acceptable grade for the plant, can be represented as a maximisation problem (the constraints are not mentioned here).

$$Max.\ Z = \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{i=1}^{N} C_i^{tm} . TB_i . x_i^{tm} \qquad (2.1)$$

T = the number of scheduling periods

N = the total number of blocks in pit

i= block index $(1, 2, \ldots, N)$

C_i^{tm}=NPV of block

x_i^{tm} =the proportion of block i to be mined in period

TB_i=the total tonnes of block

This LP model was solved by decomposing the mine life production planning model into small subsets through Dantzig–Wolf decomposition principles (Tebboth and Richard, 2001); each subset is subsequently solved as a single problem. After solving all subsets, solving the main problem would be relatively simple.

Although Johnson's (1969) approach can provide the optimum solution by considering the time value, different processing lines and the cutoff grade strategy, it is unable to solve the problem in its full capacity. Furthermore, although this model allows for the mining of the portion of a block, it cannot remove the overlying blocks before that (Osanloo et al., 2008), which causes impractical mine planning (see Figure 2-3).

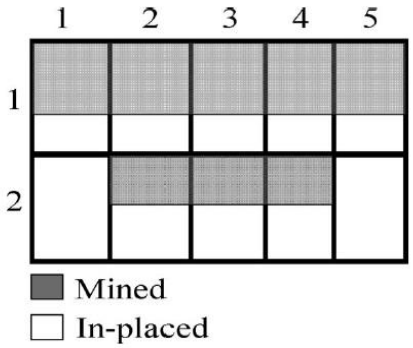


Figure 2-3: Problem of partial block mining of model (Johnson, 1969)

2-3.1.2 Mixed integer programming (MIP)

To improve Johnson's (1969) model (LP), Gershon (1983) introduced an MIP model by adding additional decision variables to Johnson's (1969) LP model. Through this change, partial blocks can be mined if overlaying blocks have been completely removed. This model can provide better practical mine planning than LP. However, Gershon's (1983) model is unable to deal with a large block model because too many binary variables need to be applied to solve the MIP.

2-3.1.3 Integer programming (IP)

The general IP mathematical formulation can be given in the following formula.

$$Max.\ Z = C_1X_1 + C_1X_2 + \cdots + C_nX_n \qquad (2.2)$$

X_n = A column vector of N variables x_i^n while $\sum_{n=1}^{n} x_i^n = 1$

C_n = A row vector of N objective function coefficients containing c_i^n elements that represents the NPV resulting from mining block i in period n.

Which is subject to the constraints such as, mining capacity, processing capacity and maximum and minimum of acceptable grade, and slope angle.

Dealing with a large number of zero–one variables that binary IP formulation has to accomplish to solve the problem usually constitutes serious limitations for this approach. Several researchers have tried to solve this problem by reducing the binary variables with different approaches, as follows:

- Lagrangian relaxation (Dagdelen and Johnson, 1986)
- 4D-network relaxation method (Akaike and Dagdelen, 1999)
- Clustering (Ramazan et al., 2005)
- Branch and cut (Caccetta and Hill, 2003)
- Disaggregation (Boland, Dumitrescu, Froyland and Gleixner, 2008)

2-3.1.4 Dynamic programming (DP)

Dynamic programming methods are the common methods that commercial optimisation software uses to optimise the open pit mine planning Therefore, it will be more thoroughly explained than the other methods.

The theory of Dynamic programming (DP) was formulated by Bellman (1957) based on searching and selecting all possible options. Roman (1974)) applied Dynamic programming in open pit mine planning for the first time, following the algorithm steps below :

- Determine the location of the starting block of sequencing process
- Check all possible sequence blocks above the given block are checked (satisfying the slope constraint)
- Select the sequence giving the highest NPV
- Place outside the pit limit the blocks not giving a positive NPV
- Continue this procedure continues until no block needs to be put outside the pit limit

Dowd and Onur (1992), Onur and Dowd (1993) and Tolwinski and Underwood (1992) proposed different methods to solve the problem. The proposed method by Tolwinski (1998) and Tolwinski and Golosinski (1995) is now used in a mining commercial optimisation software called *NPV scheduler* and can be implemented on large deposits.

As this software is used with all types of mine optimisation and mine planning in this thesis, the procedure, from optimisation to mine scheduling, which is progressively developed into practical designs by applying different constraints, or even stockpiling, is explained below. The other commercial mine optimisation/planning software is similar to Whittle 4D and partially follows the same procedure.

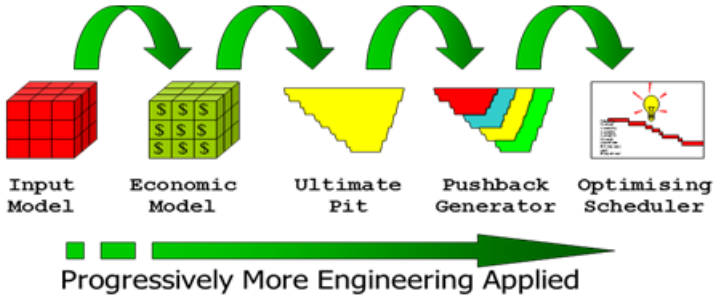


Figure 2-4: The procedure from optimisation to mine scheduling is progressively developed into practical designs by applying different constraints

This approach has the following steps, which see can be seen in Figure 2-4:

1. *Determine the ultimate pit limit.* Lerch-Grossmann (1965) algorithm is used to define the optimum ultimate pit limit. Using this algorithm yields the boundary with the highest cash flow and pushing back beyond this boundary decreases the profit.

2. *Create Lerch- Grossmann phases (Nested pits).* The starting point for this method is to generate *LG* Phases by applying the Lerch-Grossmann pit optimisation algorithm. By stepping (for example m steps) the economic parameters, such as the revenue factor and the commodity price factor in percentage increments, a series of nested pits is generated and all blocks are divided into maximum m sequential optimal pits $(1,2,\dots,m)$.

3. *Create Practical pushback selection.* The philosophy of the design of each pushback is to realise the greatest amount of metal content as early as possible with the minimum waste stripping. If the number of pushbacks is theoretically increasing, then the mining scenario will gain significant benefit. However, practically, there will not be sufficient access space between two pushbacks or minimum mining operating space in mines; therefore, the opportunity to increase the number of pushbacks is limited. Since the pushback selection is very important, several combinations of the nested pits have to be investigated to find the best NPV.

4. *Create Block extraction sequences.* Optimal Extraction Sequence of the blocks (OES) are the sequential numbers which are usually generated by the scheduling algorithms (or even optimisation and pushback design algorithms) to maximise, for example, the NPV.

5. *Create Block extraction time.* This is possibly the last step of this approach design procedure and it can generally classify the blocks into yearly scheduling by considering the practical mining constraints; as a consequence, the blocks would be given numbers $\{1,2,\dots,\boldsymbol{N}\}$ while $\boldsymbol{N}$ is the mine life.

The main advantage of this method is that it is able to generate a practical mine schedule with considering mining constraints and it is also applicable on large deposits.

2-3.1.5 The common limitations of the mine planning deterministic approaches

The common limitations of the mine planning deterministic approaches are the same as what has been discussed in Section 1-3. Nevertheless, these approaches are unable to quantify the risk of not meeting mine scheduling targets or the other project indicators due to the fact that they deal with a single geological average model which cannot take into account in situ grade variability.

2-3.2 Uncertainty-based approaches

The uncertainty based approaches and how they generally deal with mine planning activities have been explained in chapter 1, section 1-5. A great deal of research has been conducted on the mine planning uncertainty based approaches since the last decade. Still, contrary to deterministic methods, there is not any clear classification about uncertainty-based methods. The majority of them have used a combination of conditional simulation with deterministic approaches. Some of them are explained below.

2-3.2.1 Dynamic programming (DP)

Dowd (1994) introduced the following flow chart (see Figure 2-5) for risk assessment in open pit mining. In this method grade uncertainty can be assessed with other financial uncertainty (see section 1-3), such as commodity price, mining costs and processing cost. After generating the n realisations and selecting random combination of input parameters, the revenue block model of each realisation can be made. By applying pit optimisation and scheduling (using dynamic programming) on individual models, the probability distribution of each project performance indicator such as NPV and IRR may be calculated.
Although this method can handle both grade and financial uncertainty, it is unable to produce an optimal schedule and is a time consuming procedure.

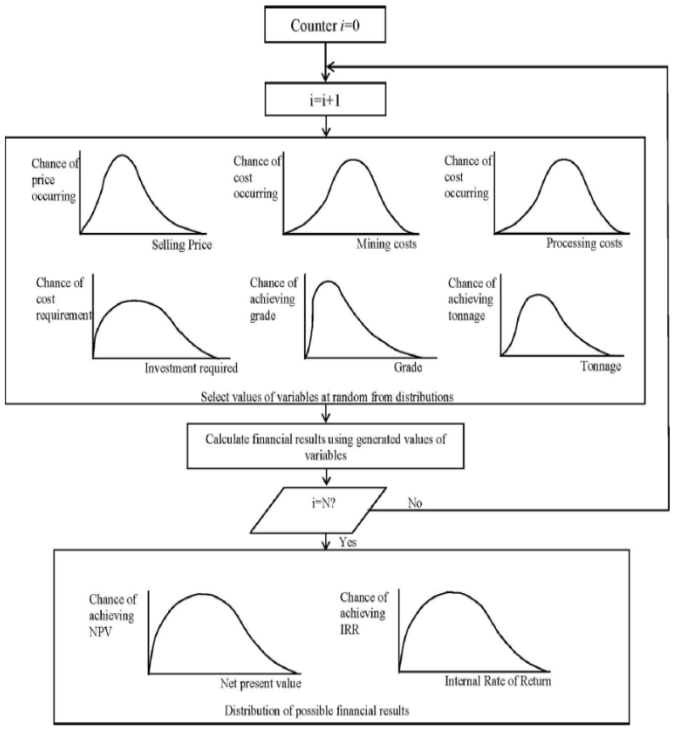


Figure 2-5: Procedure of risk assessment in open pit mine based on Dowd (1994)

2-3.2.2 Linear Programming (LP)

Dimitrakopoulos and Ramazan (2004) proposed a mathematical formulation based on linear programming, which has been able to consider grade uncertainty and equipment access and mobility, including other typical operational requirements. This formulation is based on expected block grades and the probabilities of different element grades being above required cutoffs, both derived from generated realisations. In this formulation, due to the consideration that needs to be given to equipment access and mobility, probabilities given to each block is related to the desirability of that block being excavated in a given period. To consider fleet access to each block, two concentric windows around target block *i* are defined (see Figure 2-6).

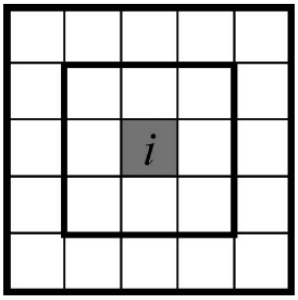


Figure 2-6: Inner and outer windows around the target block ***i*** (Dimitrakopoulos and Ramazan, 2004)

The first target is to mine that block i with the inner window. If the inner window block cannot be mined, the percent of the tonnage of the block that cannot be mined is called a 'deviation' (Y_{2i}^t), which is associated with costs (C_2) for the objective function below. Hence, the outer window blocks would be mined and again each percent of deviation (Y_{3i}^t) would be assigned a cost (C_3). That is, for this approach it is more desirable to mine block i with the blocks in the inner window than the blocks in the outer window. Yet, it is even better for the smoothness of mining schedules to mine the farther blocks with block i, if possible.

The objective function is:

$$Max. \sum_{t=1}^{T} \left[C_1^t \times Y_1^t + \left(\sum_{i=1}^{N} C_2 \times Y_{2i}^t + C_3 \times Y_{3i}^t \right) \right] \qquad (2.3)$$

T =The total number of time periods for scheduling

N =The total number of blocks in the model

Y_1^t =The deviation percent from 100% probability that the material will be mined in period t would have the desired properties

C_1^t =The cost coefficient for the probability deviation in period t

This method can guarantee producing a practical mine scheduling and minimising movement of large mining equipment inside the pit. Hence, it can reduce the risk during the first stages of mine production.

However, as NPV is not maximised in the objective function, this method is unable to generate maximum NPV in the presence of grade uncertainty.

2-3.2.3 Mix Integer Programming (MIP)

Ramazan and Dimitrakopoulos (2004) suggested a mixed integer programming model formulation that is able to consider grade uncertainty in mine planning. In this method, after generating some random realisations, scheduling patterns on each realisation are generated by applying a traditional MIP formulation on each generated realisation (with maximising NPV). Subsequently, by these given schedules patterns, the probability of each block in a given time period can be calculated. A zero probability means that the block would not be mined on that period, while one means the block will be mined. The probabilities between zero and one are considered in a new optimisation model with the following objective function:

$$Max. \sum_{t=1}^{T} \left[\sum_{n=1}^{N} \left((v_n^t \times p_n^t) \times x_n^t \right) - \sum_{m}^{M} w \times d_m^t \right] \qquad (2.4)$$

T =The maximum number of scheduling periods

N =The total number of blocks to be scheduled

v_n^t =The NPV to be generated by mining block n in period t

p_n^t =The probability of block i to be scheduled in period t

x_n^t =A binary variable, equal to 1 if the block i is to be mined in period t and 0 otherwise

w =The cost of unit deviation associated with generating a smooth scheduling pattern

d_m^t =The deviation from a smoothed production pattern when mining block m

M =The total number of blocks with smoothness constraints

The first part of the objective function tries to maximise the probability of the blocks being scheduled in the period given by realisations. The second part provides the blocks to be accessed by equipment to minimise the mobility of the mine's fleet equipment (same as LP method). As it can be seen from the objective function, this method, contrary to the previous approach, is able maximise NPV with the consideration of equipment movements and block access in such a way that it is able to produce a schedule pattern that is less risky than the traditional methods.

There are a few other stochastic MIP methods which were addressed by using these approaches. One of the most recent of these methods is by Boland, Dumitrescu and Froyland (2009)., in which the need to take into account geological uncertainty in open-pit mine production scheduling to produce schedules that adapt over time in response to the information acquired through mining was addressed. They used multiple geological estimates in a mixed integer multistage stochastic programming approach, in which decisions made in later time periods can depend on observations of the geological properties of the material mined in earlier periods. Since the material mined in earlier periods is determined by their decisions, the information received about uncertain properties as well as the time frame when that information is available is decision-dependent. Thus, the difficult case of stochastic programming with endogenous uncertainty was attempted; in addition, a successful mixed integer programming formulation of the open pit mine scheduling was extended. The problem of this stochastic case is that it is unable to show that can non-anticipatively be modelled with linear constraints involving variables already present in the model. This observation was extended to the general class of endogenous stochastic programs, and the special structure of this model was exploited to show that in some cases a significant proportion of these constraints may be omitted.

2-3.2.4 Maximum upside and minimum downside approach

The maximum upside/minimum downside (Leite and Dimitrakopoulos, 2007) in assessing grade variability/risk in different realisations assumes that there is a probability that a given mine design may perform better than expected; thus, there is an upside potential relating to the orebody considered, similarly to a downside risk where the project performance indicator's prediction is not fulfilled (see Figure 2-7).

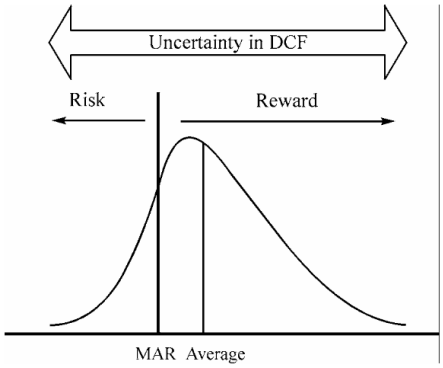


Figure 2-7: Uncertainty in a distribution of a key project performance indicator (DCF), reward or upside potential and downside risk with respect to a point of reference such as the minimum acceptable return (MAR) (Leite and Dimitrakopoulos, 2007)

This method first applies traditional optimisation and mine scheduling on each random generated realisation to design a pit. Second, it generates the distributions of any other project indicator by applying a given mine plan on each realisation. Third, it discards the designs that may not meet the user defined criteria and selects one single pit design that can capture the maximum upside reward and minimum downside risk (see Figure 2-8). The design selection in the approach outlined above is based on type I of the response variable.

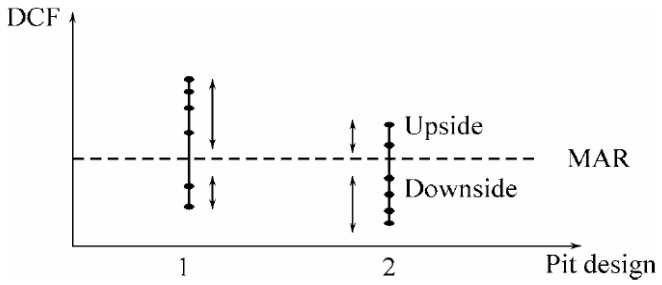


Figure 2-8: Upside potential and downside risk for two pit designs for the same orebody based on Discounted cash flow ($) (Dimitrakopoulos et al., 2007)

2-3.2.5 Limitation of maximum upside and minimum downside approach

The maximum upside and minimum downside approaches and any other risk-based mine design method (Leite and Dimitrakopoulos, 2007) which is based on generating different mine designs by the stochastically simulated orebody have two drawbacks. The first one is that the distribution of the project indicators (type I) cannot give any information about different optimised pits, pit designs or mine scheduling and it is likely that different pit designs have approximately the same project indicators. Therefore, this sort of response variable (type I) is unable to reflect the dissimilarities between different optimised pits and pit designs which come from different realisations. Generally, type I highlights the impact of grade uncertainty in mine planning activities on the assessment of the financial or technical indicators of the blocks but is not able to indicate this impact on mine planning; moreover, it assesses uncertainty in sequences or the time extracting the block's uncertainty in place (which block) and time (when to mine).

The second drawback is that the chosen design(s) would no longer be equiprobable. This means that some designs are highly likely to occur, but others would be less able to represent the actual mine design. To quote from Leite and Dimitrakopoulos, (2007). "*Although the simulated orebody models are equally probable, the corresponding designs are not*" (p.76). The common risk-based methods assume that chosen design(s) are all equiprobable, and this assumption may cause misleading results in inaccurate design selection.

Chapter 3

3 Geostatistical simulations

3-1 Introduction

The modelling grade or any parameter in geological deposits where too many factors control the mineralisation process constitute a complex procedure and cannot be described by simple models. Recourse estimation methods are based on little amounts of data which are sampled from a deposit; however, regardless of how big the deposits are, the collected data would be too small with respect to the volume of the deposits. The cost and time and sometimes even the technical difficulty are the most important factors which can limit access to more data from the subsurface. Although this complexity and the lack of access to all needed data make modelling difficult some techniques are able to handle these models to estimate the geological parameters. These techniques are classified into the major groups in the following paragraphs.

Similar to Inverse Power Distance (IPD), Nearest Neighbour (NN), and Polygonal, non-geostatistical estimation methods usually calculate the geological parameters based on different ways of determining the average of sampled points or grades situated around the point or block that is supposed to be estimated. For this purpose, the methods need to apply a weighting function to sample points and further establish how many samples have to contribute to the estimation. That is, it is necessary to find a radius of influence. Therefore, even though finding the best weights for sample points and determining the radius of influence is not always easy, there is no fixed method to calculate these geological parameters.

Physical processes do not behave randomly in space and possess spatial continuity in their properties. For example, a river-bed that is shaped by water flow is governed by a complex physical phenomenon that makes continuity on the path of the river-bed; otherwise, mineral deposits tend to concentrate in certain spatial locations. Thus, for describing these variables their regionalised aspect should be considered. Geostatistics is based on the notion of regionalised variables. A regionalised variable is distributed in space and has a certain spatial structure which is not limited to the Euclidean space. In addition, the regionalised property of variables does not allow them to be randomly embedded in the space where they are and, thus make spatial continuity.

Prior to applying any geostatistical estimation methods, such as the Kriging method, the structure of spatial continuity of the grade should be modelled. This structure can be defined in the form of spatial trend,

isotropy or anisotropy and so forth in mineralisation, for example. This spatial continuity model is called variogram. A variogram can give a mathematical solution to find the best weights for sample points and to determine the radius of influence for the estimation process not only for Kriging, but also for the other non-geostatistical estimation methods.

In geostatistics, simulation techniques are the methods or random functions which are able to generate many models – or realisations - containing the same statistical parameters as the sample data. Contrary to geostatistical estimation methods, which are only able to provide a single model for a geological phenomenon and ignore all possible variation for the variables, through simulation techniques all possible features of a phenomenon, for example possible grades of each block in a block model, can theoretically be modelled. These realisations are usually used to quantify spatial uncertainty of variables under study.

Different geostatistical simulation algorithms rely on their models of spatial continuity and spatial uncertainty; an overview of geostatistical simulations and their algorithms will be provided in this chapter.

3-2 Geostatistical estimation (Kriging)

Georges Matheron (1965) formalised the concepts of the theory behind Kriging and for first time "*geostatistics*" was introduced as a new terminology in earth science. One of the best definition for geostatistics states that it is *"the study of phenomena that fluctuated in space"* (Deutsch and Journel, 1998). By the early 1970s, the Kriging method had proved to be very useful in the mining industry. Although most of the geostatistical theory was established in geology and mining engineering in order to map the spatial distribution of grade and grade estimation, it now comprises several applications in other fields of science and other areas of engineering.

Kriging now is known by the acronym BLUE, which stands for 'The Best Linear Unbiased Estimator' That is, it is a linear combination of weighted sampled points around an unsampled point, which is supposed to be estimated (see Figure 3-1) and which can be presented as formula (3.1). These weights λ_i are proportional to the distance between sampled points and unsampled points and are calculated so that the variance of estimation becomes minimum. Thus, Kriging can guarantee an acceptable correlation between the estimated point and the neighbour data, which is called *Local accuracy* in geostatistics.

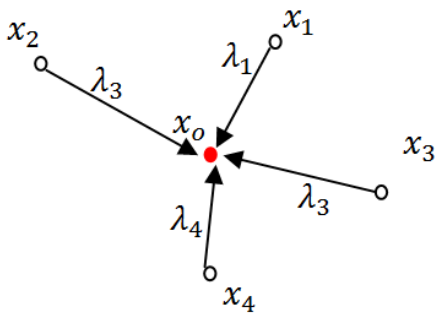


Figure 3-1: Sample Kriging weights for four sampled points located around the point $\boldsymbol{x_o}$ where estimation supposedly occurs

$$Z(x_0)_{Kriging} = \sum_{i=1}^{n} \lambda_i z(x_i) \qquad (3.1)$$

$Z(x_i)$ = Regionalized variable and show the value at location x_i

The spatial variability of a measurable geological parameter can be modelled by what is called variogram, which shows the linear correlation between any pair sample in space. Thus, if there are a group of sample pairs, a variogram can be drawn by calculating their average half squared difference. That is, a variogram presents the variance of increment $[Z(u) - Z(u+h)]$ between pairs at distance h (Deutsch and Journel, 1998), when a variogram is defined.

$$2\gamma(h) = var\{Z(u) - Z(u+h)\} \qquad (3.2)$$

$$\gamma(h) = C(0) - C(h), \forall u \qquad (3.3)$$

$C(h)$: The covariance presents spatial relationship between points as function of distance h.
Covariance can be presented in the form of a matrix; therefore it is called a covariance matrix. Figure 3-2 illustrates the relation between a Variogram and the covariance function.

Estimation methods based on weighted average, such as inverse distance or Kriging, suffer from a well-known problem in geostatistics called the smoothing effect (Isaaks and Srivastava, 1989). Smoothing relates to the reduction of variability in the estimated parameter; that is, low grade samples are usually overestimated while large values are underestimated (Goovaerts, 1997).

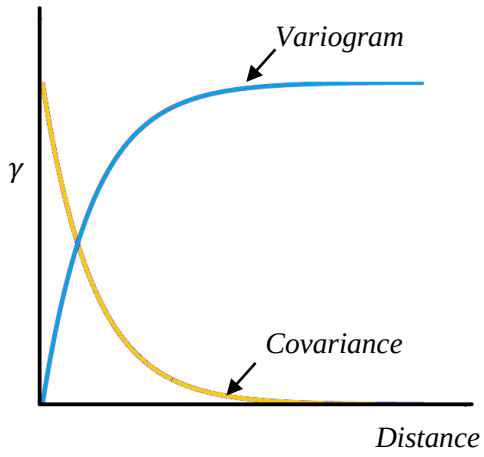


Figure 3-2: The relation between Variogram and covariance function

Figure 3-3 illustrates the smoothing effect on the distribution of the original sample point after estimation. As it can been seen, extreme samples in tails of distribution disappear after the estimation and the results shows a new distribution with fewer extremes. Thus, Kriging estimation is not able to reproduce the samples histogram and, consequently, its covariance matrix. This problem causes a big flaw in *global accuracy*, which is essential for an estimation method such as Kriging.

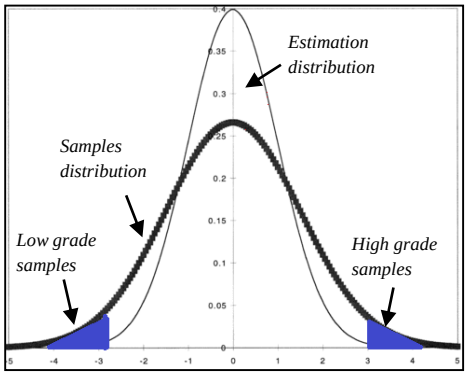


Figure 3-3: Estimation distribution pdf (thin line) shows the smoothing effect on the distribution of the original sample point (thick line)

Different approaches have been proposed by estimation and stochastic geostatistical simulation to correct the issue of the smoothing affect, which are beyond the scope of this thesis. For the interested reader, further background on how to correct the smoothing effect may be found in these references Journel et al. (2000), Yamamoto (2005) and Yamamoto (2008).

Stochastic geostatistical simulations, which will be explained in detail in the next section, have been used as an alternative to the Kriging method because they can easily provide multiple equiprobable images of the interest properties, for example grade, in such a way that both the samples histogram and the covariance matrix can be reproducible. That means that the stochastic geostatistical simulations can guarantee global accuracy while the Kriging method cannot (Yamamoto, 2008). However, stochastic geostatistical simulations suffer from lack of local accuracy.

3-3 Stochastic simulations in general

3-3.1 Monte Carlo Simulations

The Monte Carlo Simulation (Malvin and Whitlock, 1986) is a method to iteratively evaluate models by drawing random numbers from a probability distribution function. As a technique to deal with complex models, nonlinear models, or those models involving uncertain parameters, the Monte Carlo Simulation has received attention from different branches of engineering and science and has recently become a powerful tool for uncertainty analysis. This method is explained here, as the Monte Carlo technique is the base of all geostatistical simulations to evaluate grade uncertainty.

Figure 3-4 illustrates how this technique works assuming the distribution of x as known and $f(x)$ as a deterministic function. This technique, generally using a computer program, can draw several random numbers from distribution x (random input $x_1, x_2 \cdots x_m$), and then calculate the deterministic function $f(x)$ for each individual x_i.

The series of the output numbers $(f(x_1), f(x_2), \cdots f(x_m))$ which is a probability distribution of possible outcomes (pdf), can be analysed by statistics methods to achieve the results, for instance, by drawing the histogram and calculating the confidence interval.

Ultimately, the Monte Carlo method deals with random numbers and probability statistics to solve problems; thus, it is classified as a stochastic technique possessing a major advantage over the deterministic method to handle uncertainty and risk.

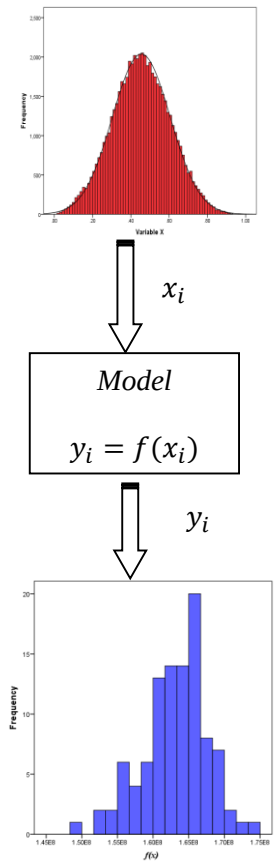


Figure 3-4: The Monte Carlo technique is able to draw many random numbers from distribution $\boldsymbol{x_i}$, and then calculate the deterministic function $\boldsymbol{f(x)}$ for each individual $\boldsymbol{x_i}$

3-4 Stochastic geostatistical simulations

Stochastic geostatistical simulation, a class of the Monte Carlo method, is the process of generating alternative, equally probable and high-resolution models of a spatial distribution. The main advantage of conditional simulations over, for instance, a single deterministic geological reserve estimate, produced, for example, by Kriging, is an estimate of the geological uncertainty and concomitant risk which a single geological estimate cannot possibly provide.

Unlike the Kriging, which suffers from smoothing problems, geostatistical simulation can easily reproduce variance, covariance and probability distribution of data set. The Stochastic geostatistical simulations are divided here into two following different groups as both are used in this thesis.

3-4.1 Unconditional simulations

Unconditional simulation is very similar to conditional simulation, where the same principle can be applied to generate the realisations but with no reference to the actual sampled points. However, it has to be faithful to mean, variance of the data set and needs to be able to reproduce the covariance matrix or variogram. This type of simulation is not common in geostatistical simulation as it cannot predict or mimic the reality of the deposit.

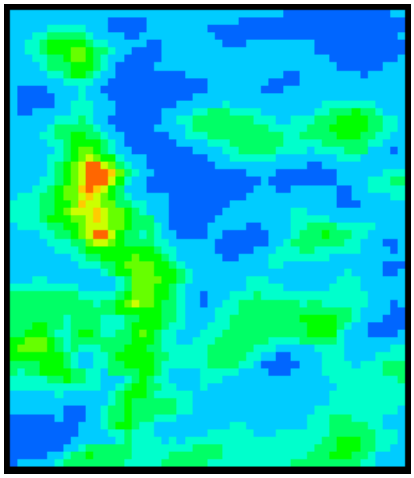
3-4.2 Conditional simulations

The simulation is called conditional if generated realisations are faithful to the sampled points or reference points at their locations. These realisations try to be honoured distributions of the sampled points and the spatial correlation of the properties without attempting to ascertain whether any one of the realisations can be real or not. The main advantage of using these algorithms over deterministic approaches is to correct the smoothing effect, produced, for example, by Kriging, that may cause less spatial variability than what it already contains (Deutsch and Journel, 1998).

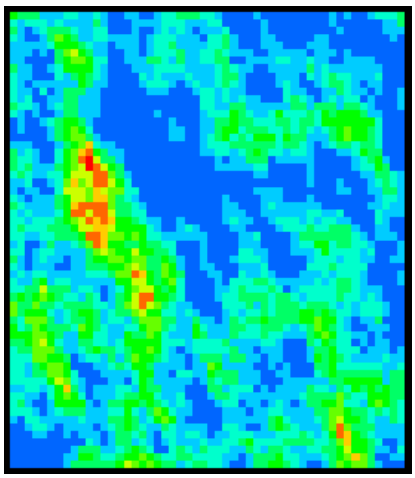
Conditional simulations are widely used for assessing uncertainty and risk analysis in verities of fields, such as resource and reserve estimation (Dowd and David, 1976); feasibility study and risk assessment (Dowd, 1994); open pit design and production scheduling (Dimitrakopoulos, 1998); hydrogeology (Yoram, 2003); environment (Webster and Oliver, 2007); and petroleum (Caers, 2011).

The set of realisations can be easily generated by applying a variety of simulation algorithms such as sequential Gaussian simulation (SGS) ; sequential indicator simulation (SIS); probability field simulation (PFS); turning bands simulation (TBS); sequential indicator simulation (SIS); LU decomposition and simulated annealing (SA). In addition, each algorithm uses different techniques or random function (RF) models to generate realisations. In the next section the three common algorithms used in this study are explained.

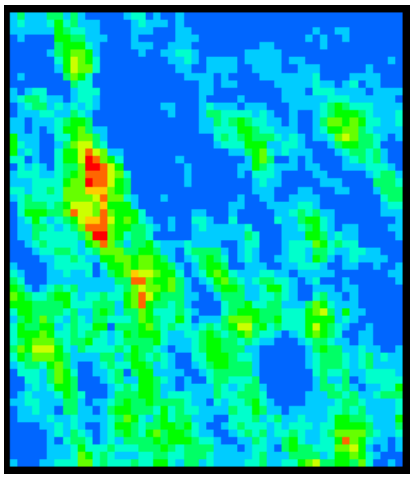
Figure 3-5 shows images of Kriging versus conditional simulation for the same geological deposit. All models are conditioned to the 470 sampled points including the same variogram model.



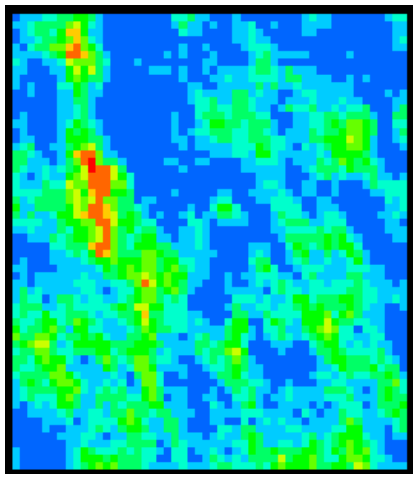
Kriging Model (A)



Realisation No.1 (B)



Realisation No.2 (C)



Realisation No.3 (D)

Figure 3-5: The Kriging result against three different realisations (B, C and D) showing reduction of variability (smoothing problem) in the estimated parameter in the Kriging model (A)

According to Figure 3-5, the grade fluctuations in the simulated images (B, C and D) are much higher than the single Kriging model (A). Although Kriging can only produce a single model, the number of realisations are fairly extensive.

3-5 Stochastic conditional simulation algorithms

We briefly review the stochastic conditional simulation algorithms that are used in this study and refer the interested reader to the relevant chapters of Goovaerts (1997) and Deutsch and Journel (1998) for more details.

Let $z(u)$ denote a spatial distribution where z is a random (RV) variable. It can be shown that the N point multivariate distribution can be decomposed into N- one point conditional cdf (Goovaerts, 1997).

$$F\left(u_{1,\dots,}u_n; z_{1,\dots}z_n\middle|(n)\right) = F\left(u_N; z_N(n+N-1)\right).F\left(u_{N-1}; z_{N-1}(n+N-2)\right)\cdots F\left(u_2; z_2|(n+1)F(u_1; z_1|(n)\right) \quad (3.5)$$

n is number of sampled data or size of original conditioning data set, and $F\left(u_N; z_N(n+N-1)\right)$ be the cumulative conditional distribution (cdf) modelling the uncertainty about $Z(u_N)$. As mentioned above, instead of picking up a single estimated value $z^*(u)$ from cdf, a series of L simulated values $z^l(u), l = 1,2,\dots,L$ can be driven. Thus, $z^l(u)$ presents a realisation of (RV) of random function $Z(u)$. This simulation is called conditional if they are faithful to the sampled points (known data) at their locations: $\mathrm{z}^{\mathrm{l}}(\mathrm{u}_\alpha) = \mathrm{z}(\mathrm{u}_\alpha), \forall l$

3-5.1 Sequential Gaussian Simulation (SGS)

Sequential Gaussian simulation (SGS) is the most common practical method of producing realisations. The conditional simulation of a variable z modelled by a Gaussian random function $Z(u)$, consists of the following steps shown in Figure 3-6.

After normalising all z data into y data with a standard normal cdf, a node on a simulation grid is randomly selected; subsequently, the Kriging estimate is applied by using neighbour sample data in order to calculate the mean and the standard deviations. Then we assume that the grade distribution of the estimated node is a normal distribution with the calculated mean and standard deviations. Then, a value $y^l(u)$ is randomly picked from the Gaussian distribution (cdf) and assigned to the node. Next, the algorithm repeats the procedure for another node, including already simulated nodes, until all nodes are simulated. Finally, after assigning a value to all nodes, back transforming the normal values $y^l(u)$ into $z^l(u)$ is applied.

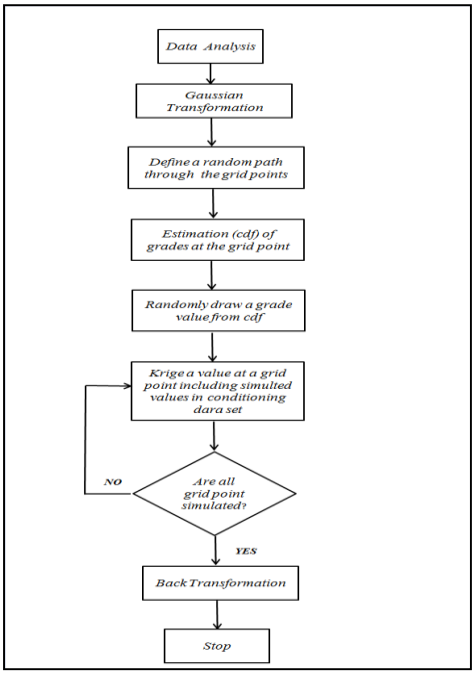


Figure 3-6: The Gaussian conditional simulation flow chart to generate the realisation

3-5.2 Turning Band Simulation (TBS)

The Turning Band simulation (TBS) is an unconditional Gaussian simulation algorithm which is designed to reduce the dimensional of simulation from 3D to one-dimensional by projecting each node in 3D on 1D lines (L). First, a set of n lines with different random directions partitions the 3D space is generated. Then, each node in 3D is projected on each of $the\ n$ lines. After that, the value for the each node is estimated as the sum of the simulated values, which come from the n lines. After completing the unconditional simulation procedures, the model is finally conditioned by sampled data. It is important to note that the number of lines should be large enough to provide smooth 3D partitioning and also allow for the quality of the simulations.

However, this simulation algorithm may have an error caused by approximations used in the TBS, such as the finite number of lines (L).

Figure 3-7 illustrates the grid value for the X point $z_s(X)$ in the first quadrant that will be estimated as the summation of the simulated values obtained from the projections of this point onto the simulated values from the eight 1D lines.

$$z_s(X) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} z(x_i) \qquad (3.6)$$

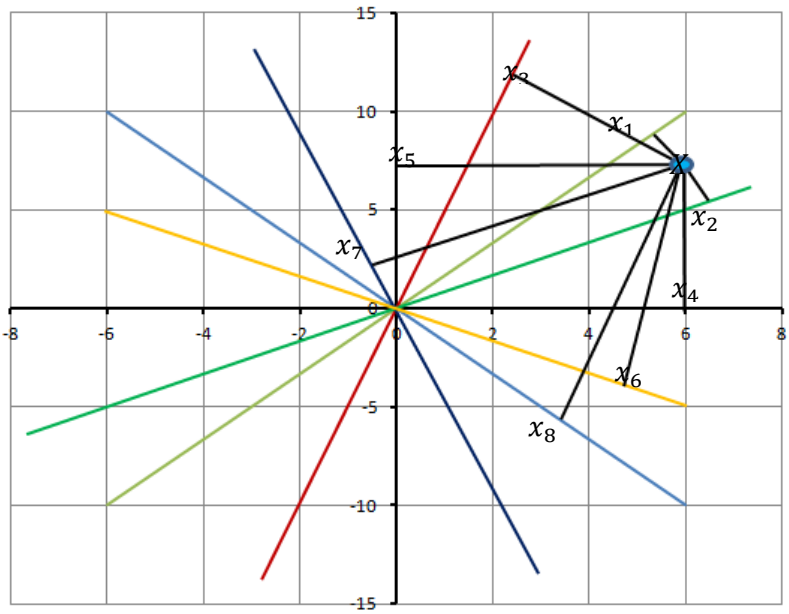


Figure 3-7: Simulation value at the point X using the Turning Band simulation (TBS)

3-5.3 Sequential Indicator Simulation (SIS)

If we want to avoid the assumption of multivariate normality, which is the common assumption for both mentioned algorithms, the Sequential Indicator Simulation (SIS) is the alternative. The algorithm is the same as SGS; however, in this method the indicator Kriging (IK) at various cut-offs estimate the conditional distribution function at each node and simulated values are randomly picked from these local distributions.

IK needs a series of threshold values between the smallest and largest values in the data set. These thresholds are usually call IK cut-offs and are used for drawing the variograms and generate the cdf of the estimation point. Traditionally, the nine deciles of the overall distribution of de-clustered data are selected (Journel, 1982). To illustrate, the first decile could be 10% of samples with a grade below and 90% with the grade above; the second 20% of samples below and 80% above. But there is a simplified form of IK called multiple indicator Kriging (MIK). MIK has just one variogram model, which comes from the indicator variogram at a cut-off of the median of the grade distribution. We use this type of Kriging to generate SIS in this study.

3-6 Relation between stochastic simulation algorithm and the space of uncertainty

The concept of space of uncertainty and its related issues, such as of equi-probability and independency of realisations, reproduction of covariance matrix and sampling from random function, have been explained in this section. There are generally two ideas about the concept of the space of uncertainty (Goovaerts, 1999). The first group believes the space of uncertainty has to be theoretically defined outside any simulation algorithm. The other groups define the space of uncertainty through the simulation algorithm including all possible realisations that the simulation algorithm can generate.

In this study we follow the second idea and show the properties of the space of uncertainty by producing many conditional simulations of a single geological model.

3-7 Transfer function

For practical problems, transfer functions apply on the space of uncertainty to yield a distribution of the output value, which is called 'response'. Mathematically, this means that the transfer functions are applied on all generated realisations and not only produce a distribution of possible project indicators, but also make create this new space, which is commonly named 'response'.

Transforming a variable using any mathematical function, for example $y_i = f(x_i)$, can be divided into two following transformations: linear transformation and nonlinear transformation. A linear transformation can keep linear relationships between variables and parameters. Thus, the correlation between x and y would be the same after a linear transformation. In addition, the nonlinear transformation depends on what kind of function is used; increases or decreases linear relationships between variables and parameters; and, thus, the correlation between x_i and y_i would be completely changed after the transformation.

A transfer function is a term used to describe a generally non-linear function, mathematical model or algorithm used to describe a process and predict its behaviour, namely responses. A transfer function may be an algorithm used to optimise the design of an open pit mine (for example, LG) requiring as input the spatially varying properties of an orebody together with other parameters. Alternatively, it could require a three-phase reservoir flow simulator as the spatially varying rock properties of the reservoir in addition to flow characteristics and engineering specifications or, similarly, a simulator of contaminant flow. The predictions from transfer functions may be evaluated over each realisation of input parameters generated

with stochastic simulations so as to obtain an uncertain distribution of the response parameters that reflect the spatial variability and uncertainty in the parameters of interest. Examples of parameters of interest may be the production schedules in a mine over the life of the mine, or the production curved on a petroleum reservoir or the cash flows from oil or mineral production, or the parts of a contaminated site which needs to be remediated.

The following sources contain a few transfer functions which have been applied on generated realisations in order to make ‘response’: Qureshi and Dimitrakopoulos (2007) , Gotway and Rutherford (1993), Qureshi (2002) and Goovaerts (1999).

Chapter 4

4 Data preparation, geostatistical analysis and simulations

4-1 Introduction

In this chapter, we briefly explain the data (as raw material) and methods that are used to generate realisations (as the final stage of this chapter). This Chapter contains the following two different datasets that were selected to illustrate four different purposes in this thesis such as mapping the space of uncertainty, realisation reduction, evaluation of stochastic simulation algorithms and risk based mine design method. We briefly explain the two dimensional (2D) and the three dimensional (3) data sets.

This two dimensional (2D) data set was first used by Isaaks and Srivastava (1989)to illustrate geostatistical concepts throughout their book called *'An Introduction to Applied Geostatistics'*. This open source data set, known as the Walker Lake data set, is derived from a digital elevation model in the Walker Lake area near the California-Nevada border, in the USA, and is a well-known data set which is used to illustrate different geostatistical techniques. This data set is actually considered in two forms: an exhaustive data set, including all data points (real case) and a smaller collected sample data (from the exhaustive data set). Both contain three different continuous variables named U, V and T. The actual meaning of these variables is not given but they are viewed as concentrations (in ppm) throughout the book.

As the majority of the statistical and geostatistical parameters of the Walker Lake were known, the calculation which is required prior to generating the realisations, such as data normalising and Kriging estimations, is not mentioned here.

As the Walker Lake data set is two dimensional, it cannot be used for illustrating the methodology in the risk based mine design. Thus, a three-dimensional data set from a typical disseminated copper porphyry deposit is used to illustrate that. The data set is based on a total of 14,500 m of drilling in 100 diamond drill holes (see Figure 4-1). As there is no statistical or geostatistical information about this deposit, all the steps required for simulation, including data preparation, database setup, statistics analysis, compositing, variography, modelling, Kriging and normalising the data were devised for this case alone, which included time consuming tasks.

4-2 Walker Lake data set

The Walker Lake data set consists of the following three measurements: V, U and T, each of 78,000 points on a $1 \times 1\, m$ grid (in this study, only the measurement V is used). From this extremely dense data set a subset of 470 sample points has been selected to represent a typical sample data set.

The complete set of all information for the 78,000 points is called the exhaustive data set which can be assumed as a population of interest or reality; subsequently, the smaller subset of 470 points is called the sample data set, which is supposed to be a representative of the reality.

Figure 4-1 shows the sampled points layout. As it can be seen, some of the samples were collected on a regular grid; however, on the west side samples were taken in clusters because of the greater interest in sampling in the high grade area. Figure 4-2 shows the distribution of V in the exhaustive data set and sample data set, respectively.

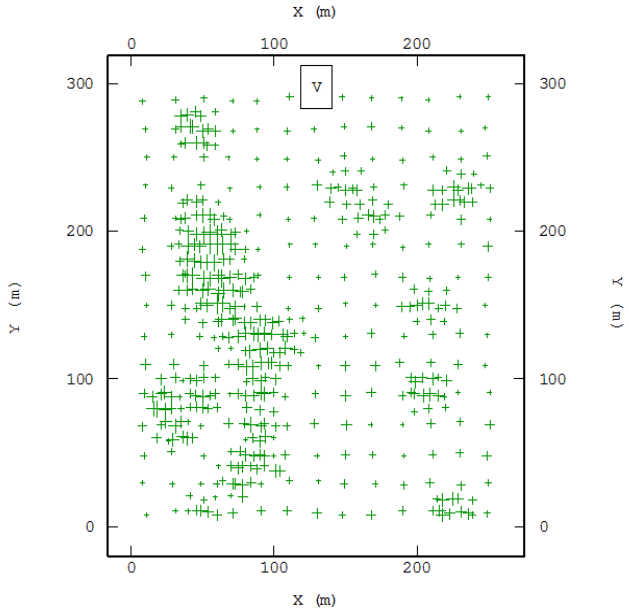


Figure 4-1: The sampled points layout of Walker Lake data set, high grade area has been sampled in more than the other area causing a clustered of sampling

4-2.1 De-clustered statistics

As it can be seen in Figure 4-2, the mean value of the sample grades is completely different from the true mean value of the exhaustive data set. The main reason of this issue can be found in Figure 4-1 as the

clustered nature of sampling from points in the Walker Lake data set causes this significant difference (those samples are not on a regular grid).

This introduces local sampling biases. For this reason, the mean value of the sample grades may be un-representative. However, it is possible to estimate the "de-clustered" mean (there are a few methods for this estimation).

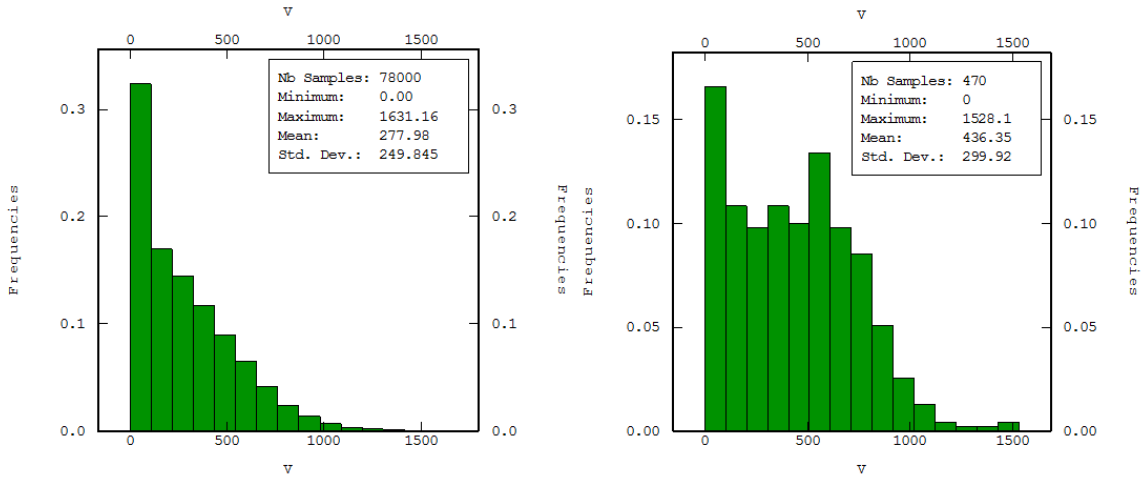


Figure 4-2: The histogram of exhaustive data (left side) set vs. the histogram of sample data set (right side) including significant difference between their average grades (V)

This is an estimate of the mean of the samples with elimination of bias due to sample clustering. It is an approximation of the mean of the deposit volume. Figure 4-3 shows the declustered histogram of the sample data which approximately has the same mean of the exhaustive data set.

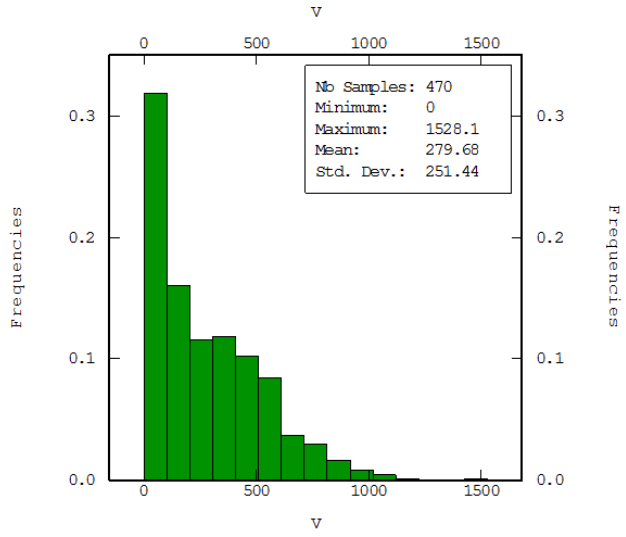


Figure 4-3: Declustered histogram of the sample data, which approximately has the same average grade of the exhaustive data set

4-2.2 Variogram models

As mentioned earlier, we do not bring any calculation required prior to generating realisations. However, regarding the importance of variograms, the experimental and model of normal score variograms and indicator variogram are mentioned here. The experimental variogram was calculated for measurement V for both the total V grades and normal scores. Figure 4-4 shows normal score experimental variograms and their models. The variogram sets were calculated in the X, Y orthogonal directions, while the two-stage spherical models were used to model the experimental variograms. This is equivalent to the variogram model parameters listed in Table 4-1. This variogram is used for generating SGS and TBS realisations.

For generating SIS realisations, indicator variograms have to be calculated. Figure 4-5 shows the experimental indicator variograms and their models calculated for measurement V. The variogram sets were calculated in the X, Y orthogonal directions and one-stage spherical models was used to model the experimental variograms. This is equivalent to the variogram model parameters listed in Table 4-1.

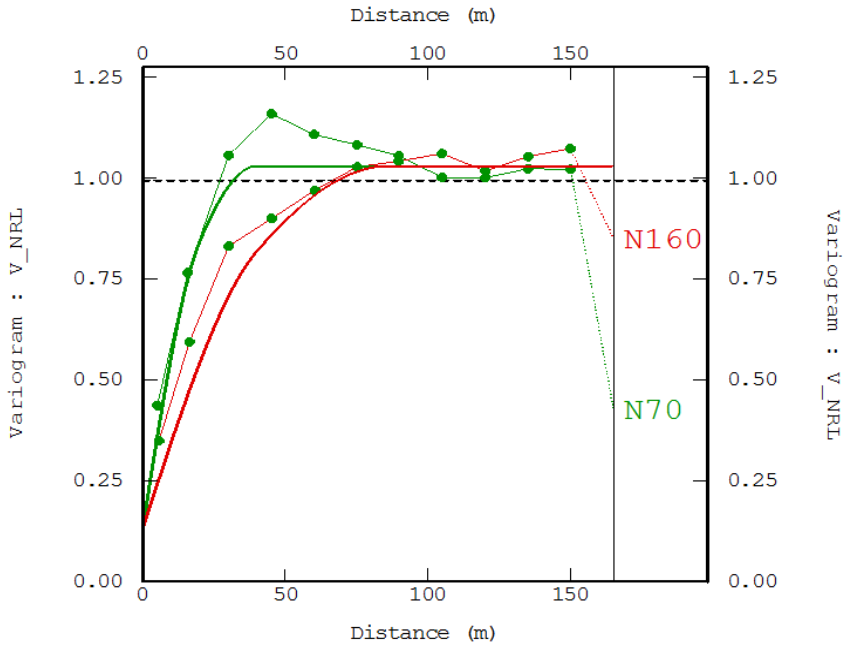


Figure 4-4: Normal score experimental variograms and their models in two major directions for the Walker Lake sample data set

Note that instead of calculating and modelling the indicator variograms at each cut-off a simplified form of indicator variogram (just one variogram model) derived at a cut-off corresponding to the median of the V grade distribution is calculated. This variogram is used for generating SIS realisations.

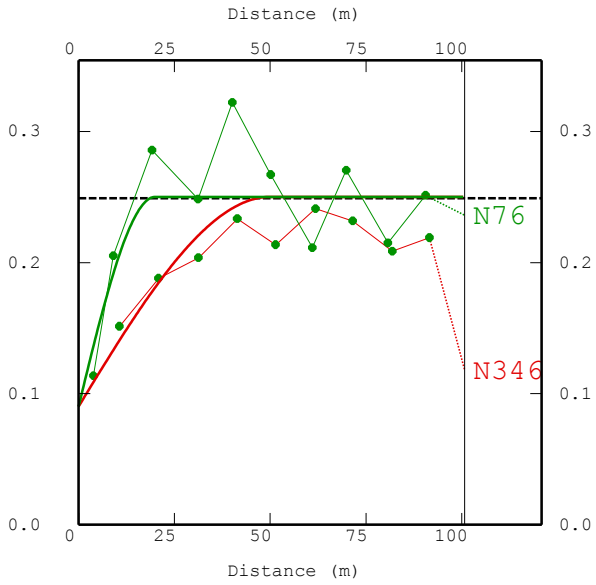


Figure 4-5: Indicator experimental variograms and their models in two major directions for Walker Lake sample data set

Table 4-1: Parameters for the variogram models for the Walker Lake case

				First structure		Second Structure	
Type	Model	Azimuth	Nugget	Sill	Range	Sill	Range
Normal Score Variogram	Two-stage spherical models	70	0.12	0.3	20	0.6	40
Normal Score Variogram	Two-stage spherical models	160	0.12	0.3	40	0.6	85
Indicator Variogram	One-stage spherical models	76	0.09	0.25	20	-	-
Indicator Variogram	One-stage spherical models	346	0.09	0.25	50	-	-

4-2.3 Generated realisations and models

Three different conditional simulation algorithms are applied on the normalised sample data set on the blocks size $5 \times 5m$. Two of the algorithms are based on Gaussian distribution hypothesis, namely Sequential Gaussian Simulation (SGS) and Turning Bands simulation (TBS). The other algorithm, Sequential indicator simulation (SIS), is based on non-Gaussian distribution. Table 4-2 shows the number of generated realisations using the mentioned algorithms in two different steps or two different seeds.

Table 4-2: Number of generated realisations using three simulation algorithms in two different steps

Simulation algorithm	Step 1	Step 2	Total
Sequential Gaussian Simulation (SGS)	450	600	1050
Turning Bands Simulation (TBS)	450	600	1050
Sequential indicator Simulation (SIS)	650	-	650

Further to the generated realisations, three different types of block models are estimated by using a block size $5 \times 5m$. The first type is estimated by Kriging; the second type by averaging the total number of each simulation algorithm, which is normally called E-type (three different E-type models); and the third model is based on Exhaustive data set.

For example, Figures 4-6, 4-7, 4-8, 4-9 and 4-10 present images of the Kriging Model, the Exhaustive Model and three generated realisations selected from the total generated realisations of SGS, TBS, SIS with their histograms, respectively.

Note that all generated realisations contain the same variograms, the same conditional points and approximately the same histograms; however, by visually comparing the colour realisation, the differences in the images are revealed (classified as dissimilarity). These differences cannot be described by geostatistical parameters.

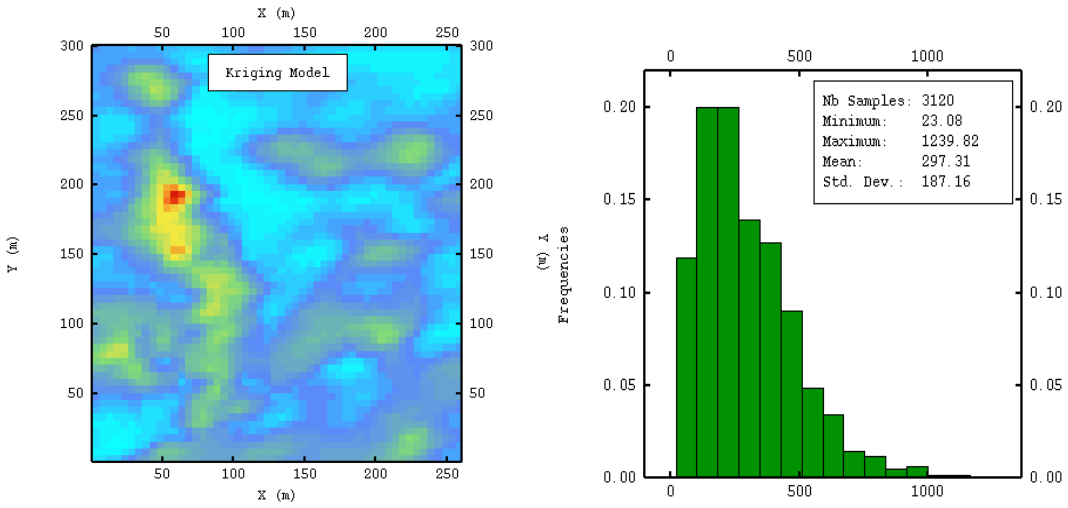


Figure 4-6: The Kriging model and its histogram of measurement V including the smoothing effect which results from the Kriging estimation which can be easily seen. The histogram has higher minimum and lower maximum of V rather than exhaustive data set and the other generated realisations and thus contains the lowest variance

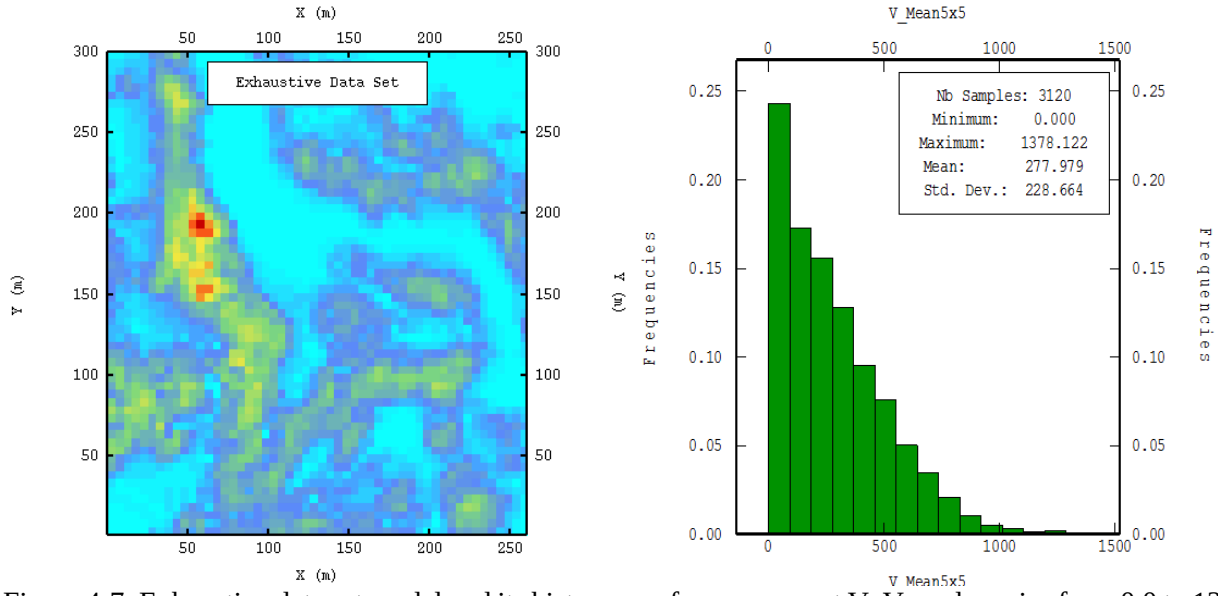


Figure 4-7: Exhaustive data set model and its histogram of measurement V. V grade varies from 0.0 to 1378 with standard deviation 228.6; the shape of grade distribution is different from the Kriging model

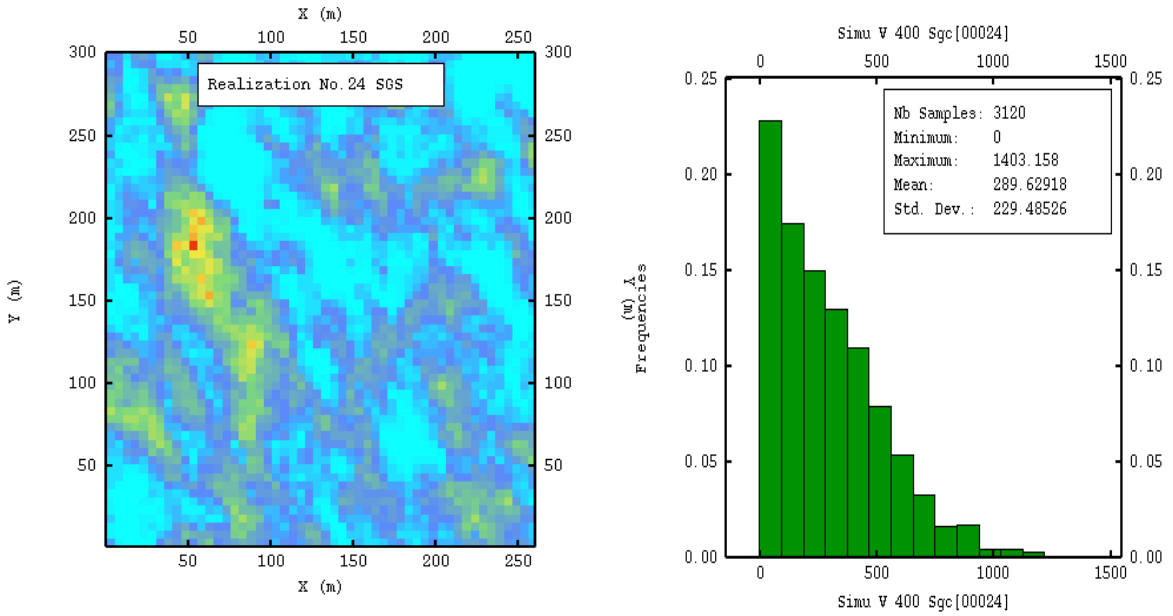


Figure 4-8: SG realisation and its histogram of measurement V. V grade varies between 0.0 to1403 with standard deviation 229.4 (no smoothing effect); the shape of grade distribution is the same as the Exhaustive data set

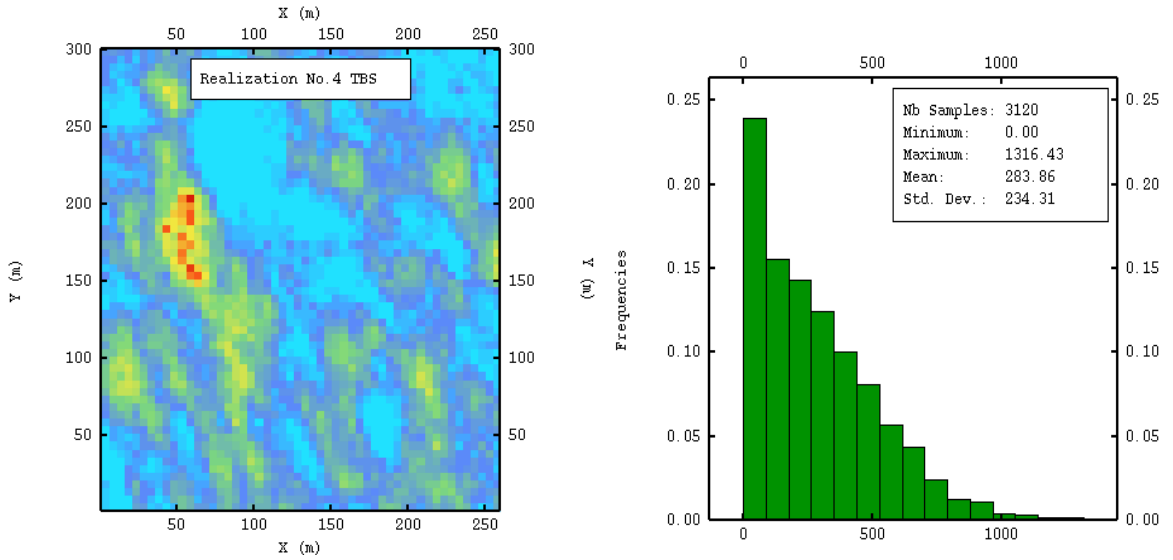


Figure 4-9: TB realisation and its histogram of measurement V. V grade varies between 0.0 to1316.4 with standard deviation 234.3 (no smoothing effect); the shape of grade distribution is the same as the Exhaustive data set

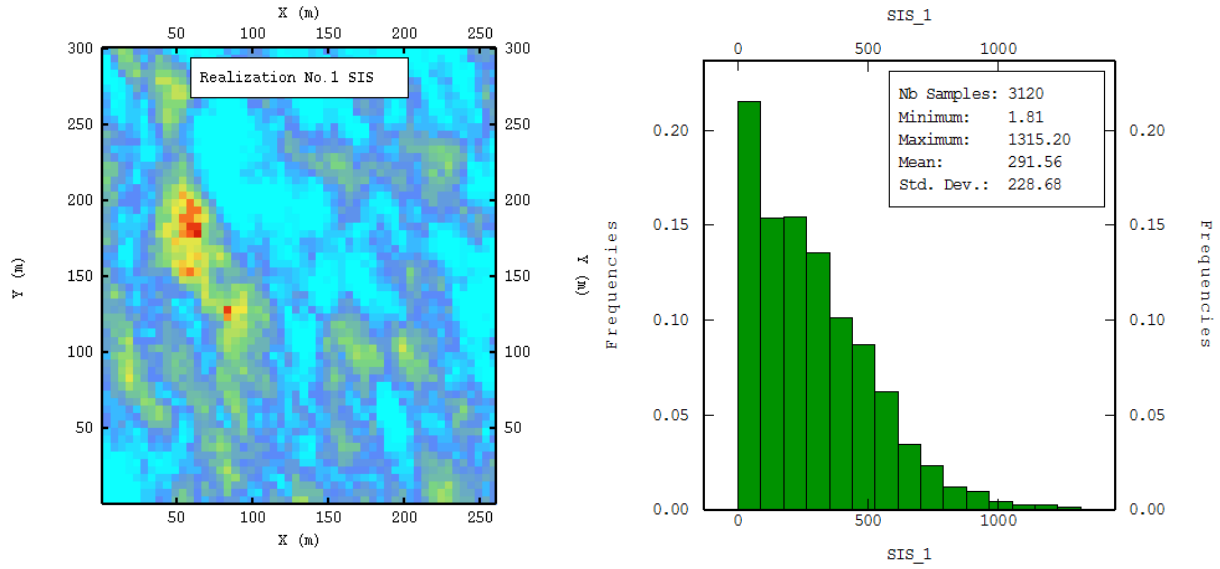


Figure 4-10: SI realisation and its histogram of measurement V. V grade varies from 0.0 to 1315.2 with standard deviation 228.7 (no smoothing effect); the shape of grade distribution is the same as the Exhaustive data set

4-3 Copper porphyry data set

A typical disseminated copper porphyry deposit with following geological feature is used in this study to illustrate the methodology not only for quantifying the space of uncertainty, but also for an application of this approach to the risk based mine design method.

The mineralisation is hosted in quartz-monzonite porphyry (QMP), which has undergone several phases of hydrothermal alteration common in porphyry systems. The economic mineralisation appears as small veins and disseminated grains, primarily in the QMP. No preferred orientation in this mineralisation has been observed although it is thought to parallel the dykes and main faults. Mineralised zones (domains) have been classified by the degree of Leaching and Supergene enrichment of the original hypogene sulphide. Primary mineralisation and high grade economic mineralisation occurred within the Supergene zone. This conforms to a classical porphyry copper style model, but on a small scale. We focus on a single estimation domain, called supergene, containing 48 Mt of measured and indicated geological resources. Within the supergene domain, the copper grades are determined for a set of 413 six meter composites from the drill holes. The mean copper grade and the standard deviation are 0.81% and 0.35, respectively. Figure 4-11 shows the histogram of copper grade distribution in the supergene zone. The block model contains 2805 blocks of $25 \times 25 \times 12.5$ (x, y and z direction) size.

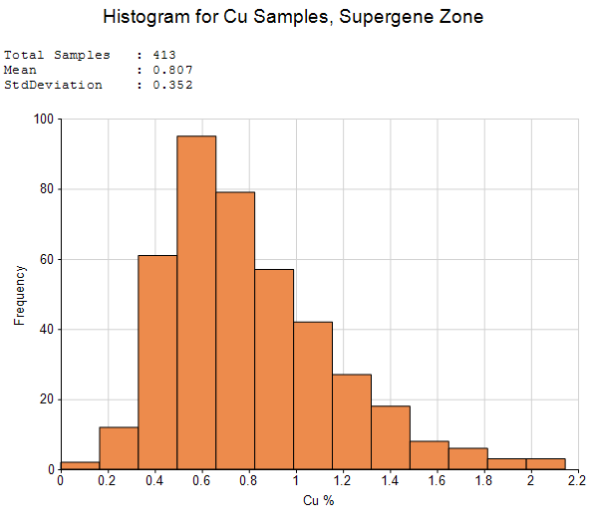


Figure 4-11: Histogram of copper grade distribution in supergene zone

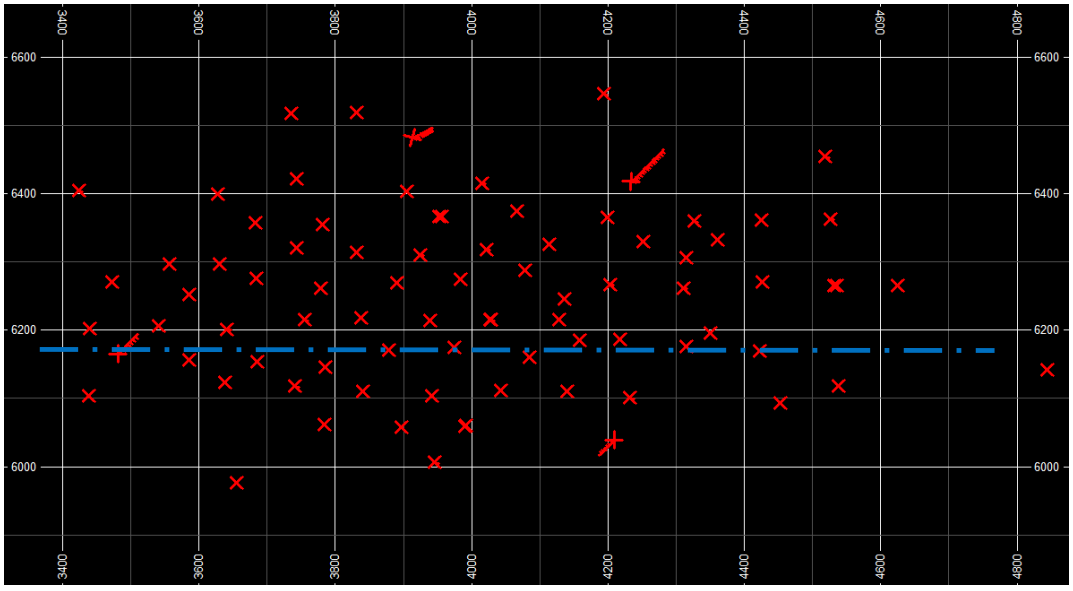


Figure 4-12: The exploration boreholes layout of the copper porphyry deposit

After verification of the exploration database a geological, mineralogical and structural model, the final geological model was developed for deposit. All models were made using Datamine Studio software. Figure 4-13 shows the North-south (6180 N) section of the orebody with three different domains, namely Leached & Oxide (blue), Supergene (green), Hypogene (red) and closely located boreholes.

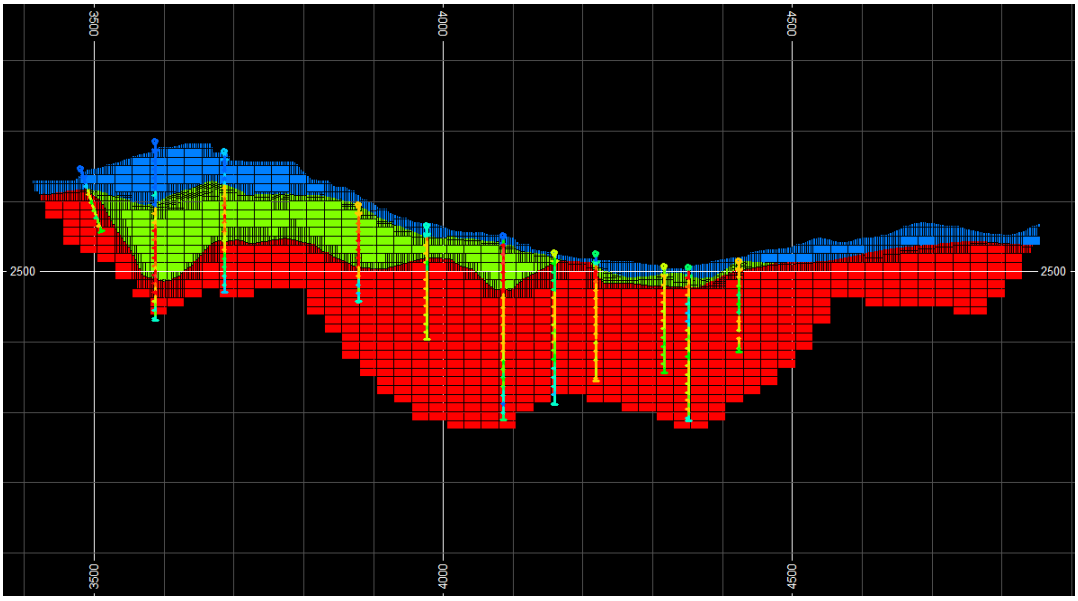


Figure 4-13: North-south section showing domains (blue is Leach zone, green is Supergene zone and red is hypogene zone) and closest boreholes from the copper porphyry deposit

4-3.1 Sample Compositing

The sample interval length for the drilling is generally about 2.0 m. Composites are used to reduce the impact of using different sample lengths on the geostatistical support. A composite length of 6.0 m was chosen for this resource estimate as it is half of 12.0 m, the bench height, and probably close to a Selective Mining Unit (SMU). During compositing there is some loss of data, which is probably acceptable considering that the introduced bias in the Supergene is minor, at around 0.1%, which will be well within the margin of accuracy for this estimation.

4-3.2 Variograms

Experimental variogram maps were calculated using 6.0 m composite samples for total copper grades and its normal scores; the variograms map sets were calculated in the x, y and z orthogonal directions. One stage spherical models were used to model the experimental variograms with parameters shown in Table 4-3. The variogram results revealed that the model has a high nugget effect, which confirms high small scale variations over the supergene zone. Based on that, we expect to have a high grade variation in the realisations. Figure 4-14 illustrates the normal score variograms and their models used to generate SGS realisations.

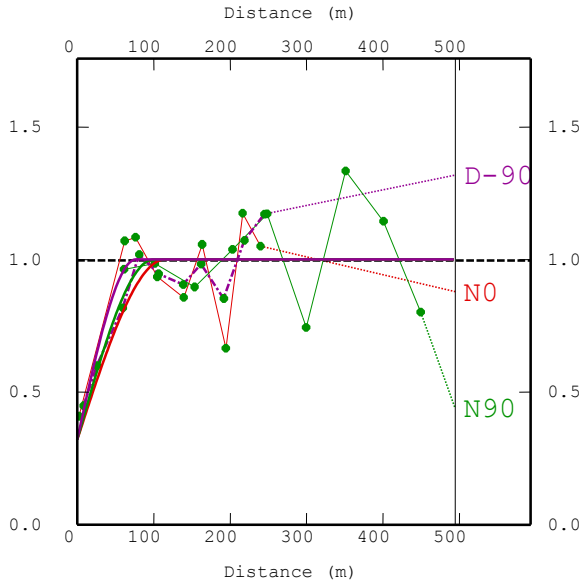


Figure 4-14: Normal score experimental variograms and their models in three major directions

Table 4-3: Parameters for the variogram models for Cu and its normal score for copper deposit case

Models	Nugget	Sill	Range (E-W)	Range(N-S)	Range(Z)
Variogram	*0.035*	*0.119*	*129.5*	*140.2*	*65.3*
Normal Score Variogram	*0.32*	*0.68*	*100*	*115*	*77*

4-3.3 Generated realisations and models

In order to better present the effects of spatial continuities (variogram ranges) and variability (variogram sills) on the space of uncertainty unconditional sequential Gaussian realisations, two types of simulation, namely unconditional and conditional, are used.

The unconditional simulation is applied, first, to generate three series of 3D-realisations with parameters shown in Table 4-4. For each set of realisations, the variograms are isotropic spherical, the search neighbourhood is spherical and there are no nugget effects. Unconditional simulation algorithm applies on normal score data, using the same grid (the copper porphyry block model) for generating unconditional realisations.

The next series (400 realisations) is generated using conditional sequential Gaussian simulation (SGS) with the variogram parameters shown in Table 4-4. Two images of these generated realisations and are shown in Figures 4-16, 4-17 (with plans of the deposit).

As it is seen, like in the Walker Lake case, the realisations contain the same variograms, the same conditional points and approximately the same histograms, but they are different from each other.

Further to the generated realisations, two types of block models are estimated. The first type is estimated by Kriging (see Figure 4-15) and the second type by averaging 400 realisations of conditional sequential Gaussian simulations (E-type).

Table 4-4: Parameters for the variogram models for unconditional realisation case

Models	Number of Realisations	Nugget	Sill	Range (E-W)	Range (N-S)	Range (Z)
Unconditional simulation						
Series 1	100	0	1	500	500	500
Series 2	200	0	1	250	250	250
series 3	100	0	1	100	100	100

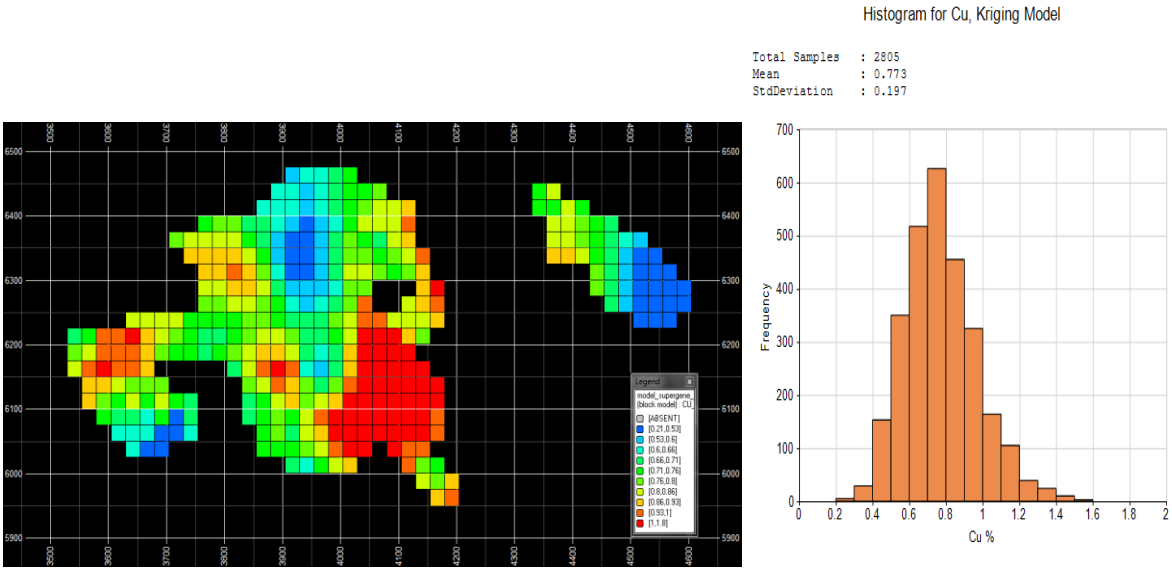


Figure 4-15: A plan view of the Kriging model and its histogram of copper grade including the smoothing effect, which is the result of the Kriging estimation and can be easily seen; the histogram has higher minimum (0.2%) and lower maximum (1.6%) of Cu rather than generated realisations, and thus has the lowest variance

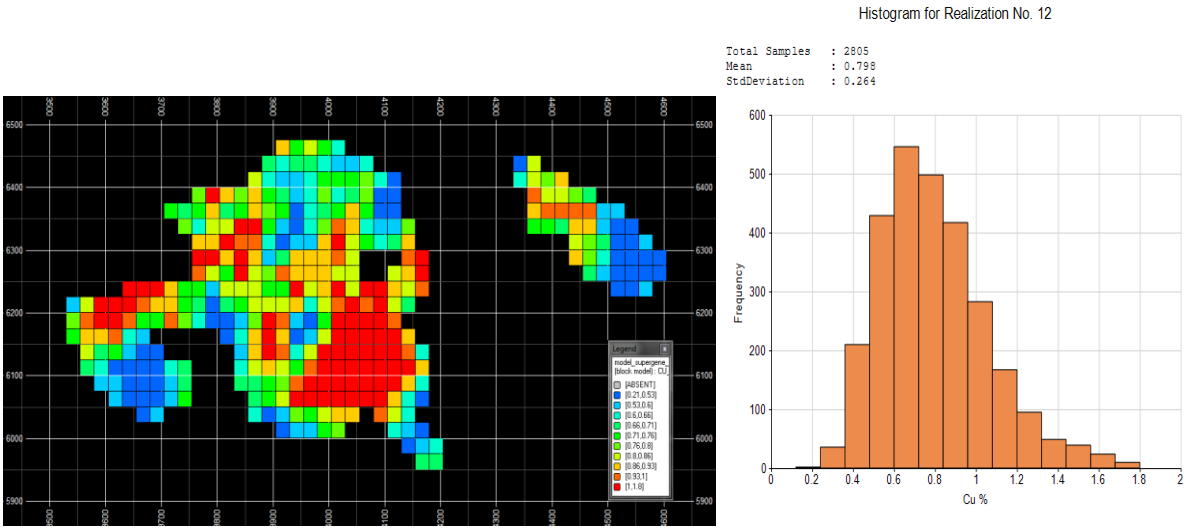


Figure 4-16: A plan view of realisation No.12 and its histogram of copper grade, Cu grade varying from 0.1% to 1.8% with standard deviation 0.264 (no smoothing effect); the shape of the grade distribution is the same as the sample data set

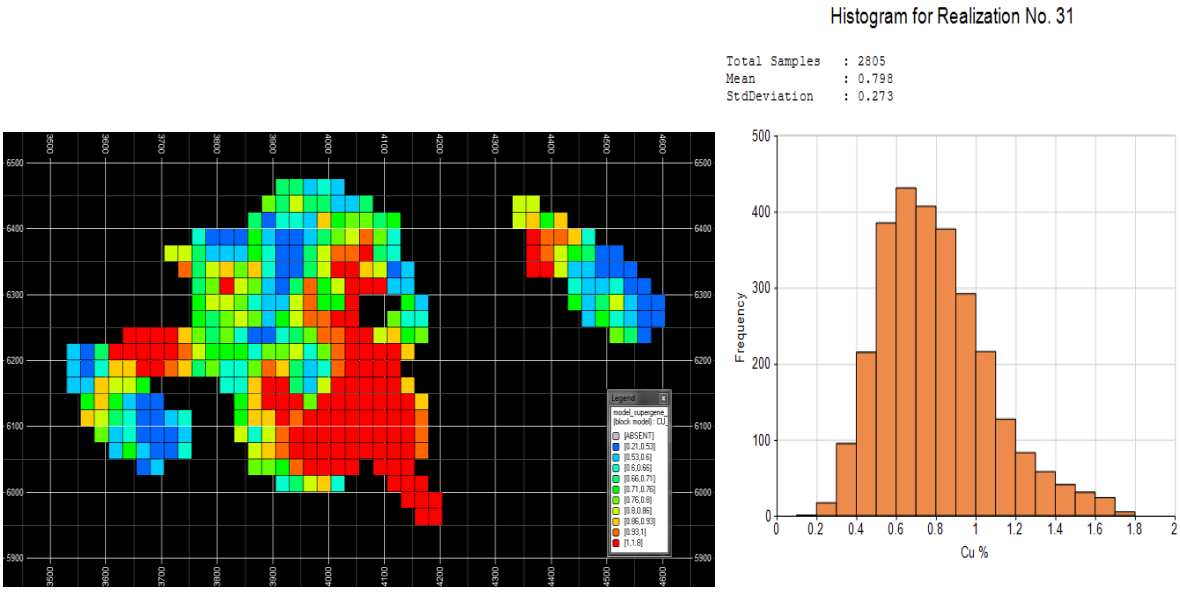


Figure 4-17: A plan view of realisation No.31 and its histogram of copper grade, Cu grade varying from 0.1% to1.8% with standard deviation 0.273 (no smoothing effect); the shape of the grade distribution is the same as the sample data set

4-4 Conclusion

As we mentioned in this chapter, the data preparation (for both cases) involved many different tasks such as statistical analysis and geostatistical interpretations, which have been accomplished in its totality. However, some of these tasks were not fully explained here as many of these data preparation activities are routine and it is not the main aim of this thesis to describe them. It is obvious that the methodologies, which will be discussed later, are generally applicable to any kind of data set, those data sets being just two examples.

Based on the two mentioned data sets (2D and 3D) applying the geostatistical study, estimation and simulation, the sets of different realisations were generated by three geostatistical simulation algorithms (SGS, TBS and SIS). As was explained in chapter 3, each simulation algorithm uses different algorithms to generate the realisation; therefore, the space of uncertainty, where created by these algorithms, should be different. Although all conditional realisations are faithful to hard data, histogram and spatial correlation (Variogram), there is no parameter that can provide further information about high order statistics for realisations. In other words, the space of uncertainty consisting of all generated realisations cannot be defined by low order statistics, while the shown images about the few generated retaliations (2D and 3D) in this chapter are different from each other.

In order to extend the space of uncertainty and having enough spatial variability of parameter V in the Walker Lake case and Cu in copper porphyry case, the number of generated realisations in this study is much higher than what is normally generated for assessing the variability and uncertainties of parameters in geosciences.

These series of generated realisations are provided to verify the performance of the methodology in assessing and mapping the space of uncertainty, the impact of changing the geostatistical parameters on the space of uncertainty, as well as comparing the output of three different simulation algorithms.

Chapter 5

5 Metric space

5-1 Introduction

In the entirety of this thesis, the word ‘space’ is used several times, but what does ‘space’ really mean? In ancient mathematics and also in the meaning used in everyday life, ‘space’ denotes a geometric abstraction of the three-dimensional space observed. This fundamental role or method used since Euclid, the Greek mathematician, construed a logically coherent framework for the idea of ‘space’. Thus, it is called Euclidean space. The new concept of ‘space’ in Mathematics constitutes a set with objects containing a structure. For example, in this thesis, a set of generated realisations (objects) can create a space if a structure can be defined between them.

Figure 5-1 shows a hierarchy of mathematical spaces with the inner product space inducing a norm space. The norm induces a metric pace. The metric space induces a topology.

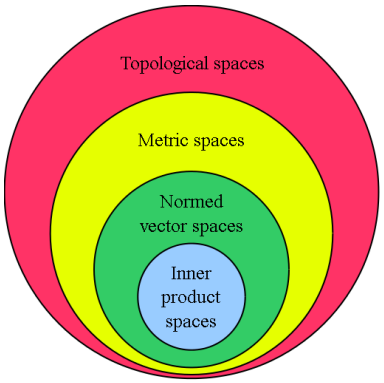


Figure 5-1: Hierarchy of mathematical spaces

The metric space is used in this thesis as the proposed methodology used for calculation and measurement of dissimilarity between the realisations is a distance-based method, which satisfies the properties of this space.

5-2 Metric space

This section briefly explains some basic and essential concepts, terminology and definitions, which are important for modelling uncertainty in a metric space and are used in this thesis. For the interested reader, further background on what is covered here may be found in the following references: Mukherjee (2005), Shirali and Vasudeva (2006) and O'Searcoid (2006).

Metric space is a set which consists of a pair (X, d), where X is a set, and d is called distance function, $d : X \times X \rightarrow R$ can be any notion of distance between the set elements. Distance functions satisfy all of the following properties expected of a notion of distance:

For all $x, y \in X$:

1. Non-negativity: The distance between two points is a nonnegative real number
$$d_{xy} \geq 0 \qquad (5.1)$$
2. Zero implies identity: The distance between two distinct points is strictly positive
$$d_{xy} = 0 \; if \; x = y \qquad (5.2)$$
3. Symmetry: The distance from point x to point y is the same as the distance from point y to point x
$$d_{xy} = d_{yx} \qquad (5.3)$$
4. Triangle inequality: Given a triangle xyz, the length of any side is less than or equal to the sum of the lengths of the remaining sides
$$d_{xy} \leq d_{xz} + d_{zy} \; \; for \; all \; z \qquad (5.4)$$

There are many distance functions which can satisfy these requirements (see Table 5-1). Among these distance metrics, the Euclidean distance is the most commonly used because of the simplicity of the distance function In Euclidean norm R^n (a small sub set of metric space) the notion of distance $d_{xy}, \quad \forall \, x, y \in R^n$ is defined as below:

$$d_{xy} = \|x - y\| \qquad (5.5)$$

$$\|x\| = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}} \qquad (5.6)$$

$$\|y\| = (y_1^2 + \cdots + y_n^2)^{\frac{1}{2}} \qquad (5.7)$$

Table 5-1 shows some of the common distance functions that may be used for dissimilarity measure. A notion distance between sets is a key ingredient in quantifying dissimilarities. This notion of distance

depends on the application; must be meaningful; and must be able to explain what causes the dissimilarity between sets.

As explained above, a metric space is represented by distances between points (which are called interpoint distances). To describe the vicinity between the points in the metric space, neighbourhood radius (as mentioned below) is generally used. A neighbourhood of a point is a set containing the point where you can slightly move it without leaving the set.

In a metric space (X, d), neighbourhood of a point C is a set of the points $\{x_1,..., x_n\}$ which are located inside an open ball with centre x_i and radius r, in such a way that $d(x_i, C) < r$. Radius r is called neighbourhood radius.

Table 5-1: Some of the common distance functions that may be used for dissimilarity measure (Webb and Copsey, 2011)

Dissimilarity measure	Mathematical form
Euclidean distance	$d_e = \left\{ \sum_{i=1}^{p} (x_i - y_i)^2 \right\}^{\frac{1}{2}}$
City-block distance	$d_{cb} = \sum_{i=1}^{p} \lvert x_i - y_i \rvert$
Chebyshev distance	$d_{ch} = \max_i \lvert x_i - y_i \rvert$
Minkowski distance of order m	$d_M = \left\{ \sum_{i=1}^{p} (x_i - y_i)^m \right\}^{\frac{1}{m}}$
Quadratic distance	$d_q = \sum_{i=1}^{p} \sum_{j=1}^{p} (x_i - y_i) Q_{ij} (x_j - y_j)$, Q positive definite
Canberra distance	$d_{ca} = \sum_{i=1}^{p} \frac{\lvert x_i - y_i \rvert}{x_i + y_i}$
Nonlinear distance	$d_n = \begin{cases} H & d_e > D \\ 0 & d_e \le D \end{cases}$
Angular separation	$\frac{\sum_{i=1}^{p} x_i y_i}{\left[\sum_{i=1}^{p} x_i^2 \sum_{i=1}^{p} y_i^2\right]^{1/2}}$

It should be mentioned here that the common definitions of the Euclidean spaces such as coordinates, axis, direction, quadrants, geometric shapes and plane are quite meaningless in the metric spaces. However, there are a few techniques to embed a metric space into the Euclidean space (R^n) by assigning coordinates to the points, which is explained later.

5-3 Kantorovich Metric

The transportation problem has had an important role in mathematical linear programming due to its general formulation and methods of solution. The original transportation problem, formulated by the French mathematician Monge in 1781(Vershik, 2005), consists of finding an optimal way (minimal cost) of transporting different piles of sand with total volume of V into holes of the same total volume; meanwhile, all piles are connected to the holes and, therefore, the piles or even a few of these piles may be transported between holes. In the 1940s, the Russian mathematician and economist Kantorovich introduced a relaxed formulation of the problem and proposed a variational principle for solving the problem (Deng and Du, 2009).

Distance can be calculated and measured between patterns, objects, subsets, or sometimes between groups of objects, or even probability density functions (PDFs). Kantorovich distance belongs to the latest group and is generally used to compute the distance between probability density functions or histograms (in case of discrete data). That requires a notion of distance between the basic features that are aggregated into the distributions, which is often called ground distance (Deng and Du, 2009). Before going through the formal Kantorovich metric, it is essential to explain the two following terms in the context of metric space:

Definition 1: A metric space (X, d), is called compact if every sequence in X has a convergent subsequence. A set Y of X is compact if every sequence in Y has a subsequence converging to a point in Y.

Definition 2: A function μ from metric space X to the extended real number line is called a measure if it satisfies the following properties:

$$\mu(u) \geq 0\,, \forall u \subset X \qquad (5.8)$$

$$\mu(\emptyset) = 0 \qquad (5.9)$$

$$\mu\left(\bigcup_i u_i\right) = \sum_i \mu(u_i) \qquad (5.10)$$

And a measure is called a probability measure if its total measure is equal to one($\mu(X) = 1$).

The Kantorovich metric provides a way of measuring the distance between two distributions.
Let (X, d) a compact metric space and two μ^r, μ^s are probability measures on X and the Kantorovich distance between μ^r, μ^s is defined by the following formula:

$$D(\mu^r, \mu^s) = sup\left\{\left|\int f d(\mu^r) - \int f d(\mu^s)\right| : \| f \| \leq 1\right\} \qquad (5.11)$$

$$\|f\| = sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)} \qquad f: X \to \mathbb{R} \qquad (5.12)$$

Kantorovich metric can be defined in the following format as well. This alternative explanation is closer to what it is used in the thesis.

Let (X, d) a compact metric space on X and $\mathfrak{S} = \{\mu^{s1}, \cdots, \mu^{sS}\}$ set of probability measure on X, $\forall\, z \,\epsilon\, X$ if $\mu^{s1} \in \mathfrak{S}$ then $\int_X d(x, z)\, d\mu^{s1}(x) < \infty$.
Let $M(\mu^{s1}, \mu^{s2})$ set of all probability measure on space $X \times X$ with marginal measures μ^{s1} and μ^{s2}.
If $\mu \in M$ then $\int_{y \in X} d\mu(x, y) = \mu^{s1}(x)$ and $\int_{x \in X} d\mu(x, y) = \mu^{s2}(x)$.
Regarding what was mentioned above, for $\mu^{s1}, \mu^{s2} \in \mathfrak{S}$ Kantorovich can be defined as below.

$$D(\mu^{s1}, \mu^{s2}) = inf\left\{\int d(x, y) d\mu(x, y) : \mu \in M(\mu^{s1}, \mu^{s2})\right\} \quad (5.13)$$

In some contexts, Kantorovich distance may be known as the Earth Mover's Distance (EMD). The EMD between two distributions, namely histograms, is the minimum required work for changing one histogram into the other or the minimal work required to transform the histogram r into the histogram s.

5-4 Similarity in Metric space

The concept of similarity and dissimilarity as a method has been used in many technical fields, such as data mining; image processing; pattern recognition; machine learning; and classification and categorisation between the realisations (objects or patterns). The notion that " similarity-based methods (SBM) are a generalisation of the minimal distance (MD) methods" (Duch and Grudzinski, 1998) may be able to convey

the idea contained in the method of similarity and dissimilarity. That is, similarity and dissimilarity can be measured in a metric space by minimising an applied distance function on realisations.

Considering the distance function D is applied on a given set of $\boldsymbol{S}$ generated realisations, the distance function indicates how far pairs of realisations are from each other in a metric space, similar to Euclidean distances between two points in R^n. Thus, dissimilarity between all the pairs of realisations, namely interpoint distances, can be calculated for the $\boldsymbol{S}(\boldsymbol{S}-1)/2$ distinct pairs of realisations; consequently, the square symmetric distance matrix $\boldsymbol{S} \times \boldsymbol{S}$, which is usually called dissimilarity matrix, is constructed. The dissimilarity matrix has zero diagonal elements $D_{ii} = 0$, as there is no any distance or dissimilarly between a point and itself.

Calculating the dissimilarity matrix creates a structure between points (realisations) in a metric space, which can be used for any further study such as, classification, clustering and sampling inside the metric space, namely the space of uncertainty.

5-5 Multi-dimensional scaling (MDS)

As described before, although the data that we deal with (in this thesis) is in the form of pairwise similarities or dissimilarities between generated realisations, it can be presented in the form of points; however, these points in the metric space do not possess any coordinates. Therefore, the data cannot be visualised to get a sense of how near or far points are from each other. To solve these problems Multi-dimensional Scaling (MDS) techniques are applied on the dissimilarity distance matrix.

MDS techniques can embed the points from a metric space into the Euclidean space (R^n) in such a way that the inter-point distances after embedding are close to what they were (Cox and Cox, 2000). That is, for a given dissimilarity matrix $\mathrm{S} \times \mathrm{S}$ of distances D_{rs}, the MDS techniques attempt to find $\boldsymbol{S}$ points $\{\mathrm{y}_1, \cdots, \mathrm{y}_\mathrm{S}\}$ in the Euclidean space R^n for some specified n, such that the distance array $\widetilde{D}_{rs} := \|y_r - y_s\|^2$ approximates the array D_{rs}. For visualisation purposes, the chosen n has to be 2 or 3.

Although there are many different MDS techniques for embedding, all of them use approximation to find the best points $\{y_1, \cdots, y_S\}$. The following factor, which is called "Stress" in the Kruskal MDS method (Kruskal, 1964), can describe the accuracy of this embedding. The Stress factor has to be minimised to get a better embedding result. The Stress factor would be zero, if the MDS can perfectly reproduce the original distances.

$$Kruskal\ Stress = \left[\frac{\sum_{i,j}(\widetilde{\mathrm{D}}_{i,j} - \mathrm{D}_{i,j})}{\sum_{i,j} \mathrm{D}_{i,j}{}^{2}}\right]^{\frac{1}{2}} \qquad (5.14)$$

By increasing the dimension from $n = 2$ or 3 to larger values the inter-point distance approximation, namely D_{rs} would usually become better, and consequently the stress factor would decrease; however, in most of cases the decreasing trend stops and finally becomes flat. That means that adding more dimension into R^n could not improve the stress factor. Figure 5-2 schematically shows the number of dimension versus Kruskal stress.

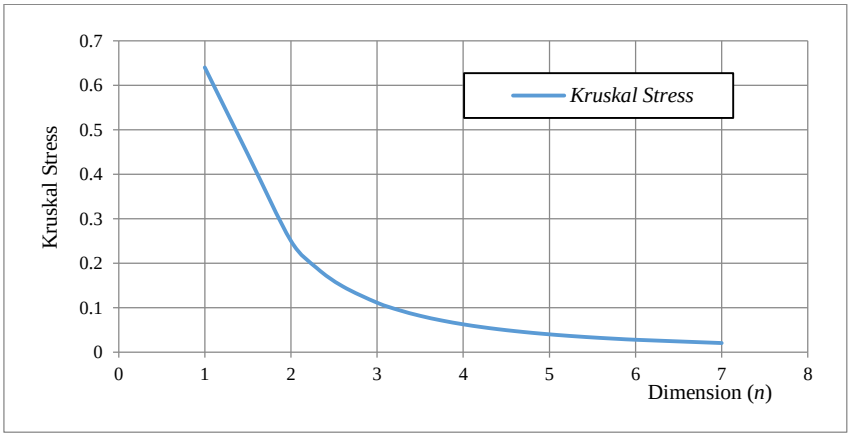


Figure 5-2: Schematically illustration of the relation between dimensions of embedding space against Kruskal Stress

Nevertheless, in this study, we advocate MDS for visualisation purposes only. This means that all further calculation on the dissimilarity matrix would be based on original distances and not on approximation. The reasons are the following: first, one finds optimal subsamples based on the original distances D_{rs} and not approximations. In addition, the objective minimised has a physical meaning in terms of 'work' done. Next, it contains the flexibility of optimising over the weights assigned to the subsamples. Finally, optimality (or the solution quality) is guaranteed and the process is reproducible.

5-6 Clustering in metric space

For a given set of data or objects, the classification process is about accurately assigning labels to data by a defined rule, which is generally called supervised classification.

Clustering, unlike the classification, is about the class label of the data or objects which are unknown.

That means that the clustering process does not rely on predefined classes therefore that is called unsupervised classification. Objects or data within a cluster are more similar to each other and are also very dissimilar to objects in others clusters. The clustering technique or unsupervised classification has been widely used in different fields and that is a very powerful tool in data mining.

There are many clustering methods; however, they can be mainly divided into the following three groups, namely distance-based method; density based method; and hierarchal clustering method (Webb and Copsey, 2011).

The clustering method used in the thesis is distance-based. Generally, in this study, we refer to the clustering process as an optimisation problem, which tries to classify objects, data and patterns into individual sub-collections or clusters using Kantorovich distance (as criteria for the dissimilarity). We will explain our methodology (clustering algorithm) to optimally subsample a large collection of realisations using Kantorovich distance in chapter 6.

The data clustering can be split into two following steps: selecting of the clustering algorithm and decision of the number clusters in data set. We briefly explain these steps.

A very rich literature on clustering algorithm has been developed over the past decades; however, the most cited and popular method for clustering, is the k-means algorithm (Han et al., 2011). It starts with an initial solution, which is iteratively improved using two optimality criteria in turn until a local minimum is reached (Franti and Kivijarvi, 2000). The algorithm steps are easy to implement and give reasonable results in most cases; however, the clustering results are highly dependent on the initialisation points, that is, different initialisation points may give different clustering results. There are several methods of clustering, such as Hierarchical clustering, Genetic Algorithms and the Fuzzy method, which are beyond the scope of this thesis (Webb and Copsey, 2011).

One of the main issues in any type of clustering method is to make a decision about the number of clusters which have to be made prior to the execution of the algorithm. This needs to include the user in the second subclass, called *the number of clusters in data set.* A few techniques can be used here, the elbow method being the most commonly method to find an appropriate number of clusters. Assuming that the k-means algorithm is chosen, the elbow method selects number n as the optimum number of clusters if $n + 1$ cluster does not add sufficient information or give much better modelling of the data. More precisely, if you graph the percentage of variance explained by the clusters against the number of clusters, the percentage of

variance initially increases dramatically and later starts to flatten out as the optimal number of clusters increases. Thus, the elbow of that graph angle in the graph) would correspond to the optimal number of clusters (see Figure 5-3).

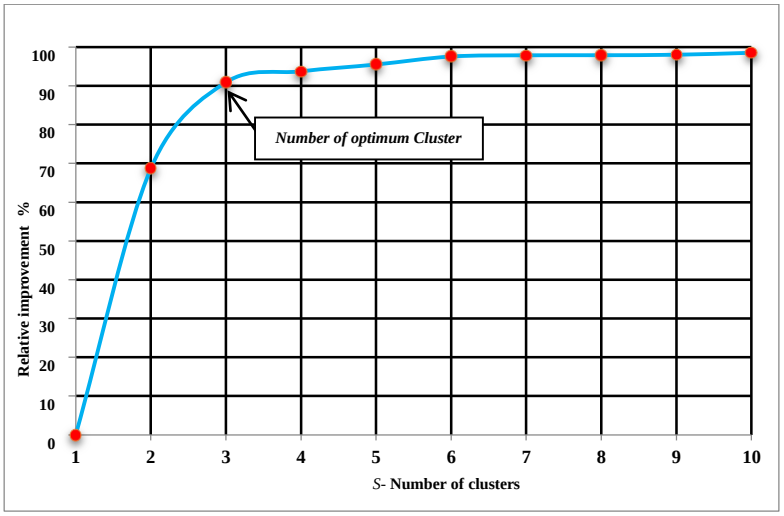


Figure 5-3: The elbow method for indicating the optimal number of clusters. The optimum number of clusters is 3

Chapter 6

6 Description of methodology

6-1 Introduction

It was explained in chapter 4 that all conditional realisations are faithful to hard data, histogram and spatial correlation (Variogram). In addition, there is no parameter that can provide further information about high order statistics for realisations and, therefore it is obvious that two realisations can be significantly different in ways that cannot be captured by descriptive geostatistics. Furthermore, by visually comparing the colour realisation images (if they are 2D), we can easily see uncaptured spatial differences, which may affect the results of transferring functions. Although there are several different realisations, some of them are visually more similar to each other than others. For example, Figures 6-1 and 6-2 present images of four realisations selected from 450 generated realisations (SGS) of the Walker Lake data set. By visually comparing these images, we conclude that the images of Figure 6-1 (images A and B) are closer to each other than the images of Figure 6-2 (images C and D).

In this framework, the pairwise dissimilarities between realisations can be used to make a relation or a precise mathematical structure between them, which can describe the variability of parameters of interest (for example, grade) inside the space of uncertainty. This method provides a powerful tool to address how realisations are connected to each other and how this connection (structure) can answer some controversial questions in geostatistical simulations.

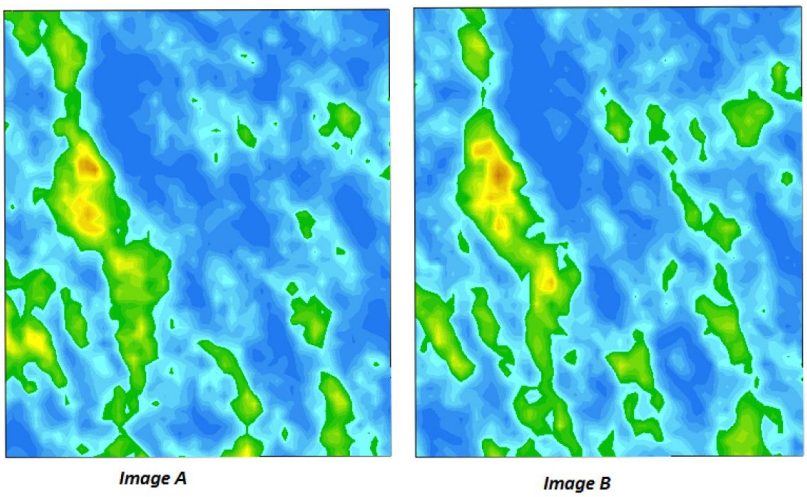


Figure 6-1: Selected realisations (SGS) of Walker Lake data set that are very close to each other

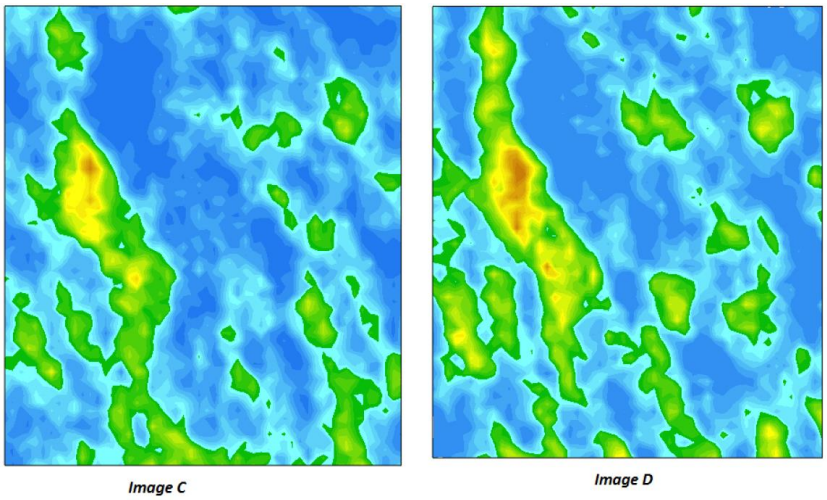


Figure 6-2: Selected realisations (SGS) of Walker Lake data set that are far from each other

Moreover, quantifying this spatial dissimilarity can be a powerful technique to assess and map the space of uncertainty. Such quantification of the space of uncertainty makes it possible to compare the impact of changing the geostatistical parameters or even simulation algorithms on the space of uncertainty. Our methodology has the potential to consistently compare the output of different geostatistical simulation algorithms, such as SGS, sequential indicator simulation (SIS) and turning bands (TBS) simulation.

Furthermore, if we place any deterministic geological reserve estimation, produced, for example, by Kriging, inside the space of uncertainty, the method can easily reveal how dissimilar other realisations are to the estimated model. In other words, measuring how close global accuracies (different realisations) are to the local accuracy (estimation) is now possible.

Moreover, the mining processes, such as mine optimisation, open pit design and long term scheduling, are only able to handle relatively modest numbers of realisations. It is difficult to say how many realisations are required to achieve a prescribed level of accuracy based on a very large number of possible realisations. This method has the ability to construct a collection of realisations so that the overall uncertainty is captured in a way prescribed by the user. For example, in open pit design, because of the smoothing effects, the NPV outcome from an ordinary Kriging model (estimation) is usually higher than the mean of the distribution of NPVs of conditionally simulated realisations (Dimitrakopoulos et al. 2002). This means that selecting realisations that are not close to the Kriging model is an effective way to indicate the range of possible NPV estimates; while selecting realisations that are far from one another indicates the range of geological uncertainty affecting the open pit mine design. To quantify this spatial dissimilarity, we present the Kantorovich distance as an important and intuitive measure of dissimilarity. This distance may be used to detect and identify the structural relationships between realisations and has obvious applications to clustering, selecting representative realisations and visualising uncertainty. In this chapter, we focus on the concept and methodology of our approach rather than the applications. In the next chapter, we present applications based on the concepts introduced here.

6-2 Definition of dissimilarity by Kantorovich distance

A notion of 'distance' between generated realisations is a key ingredient in quantifying the space of uncertainty and relative matters. This notion of distance must be geologically meaningful. We now begin to describe the necessary background constructions and the Kantorovich metric.

Throughout this chapter, we assume that our block model contains N blocks; we denote the three-dimensional coordinates of the centre of block i by c_i , $i = 1, \cdots, N$.

Let $X = \{1, \dots, N\}$ and we define a metric (or distance function) $d : X \times X \rightarrow R^+$ as

$$d(i,j) = \left\| c_i - c_j \right\|_2 \quad 1 \leq i,\ j \geq N, \qquad (6.1)$$

Where $\|.\|_2$ is the standard Euclidean norm on R^3 (between the centres of blocks i and j).

Let us consider a particular pair of conditional simulations r and s. In the simplest setting, we assume that there is only one material of interest (for example, copper). For each block $i = 1, \cdots, N$, simulation r will assign a mass of material, which we denote by m_i^r . Similarly, from simulation s we have $m_i^s,\ i = 1, \cdots, N$.

We assume that $\sum_{i=1}^{N} m_i^r = \sum_{i=1}^{N} m_i^s =: M.$ this is not unreasonable in view of the law of large numbers and the fact that N is typically very large. Also, all generated realisations have to have approximately the same average grade (metal content); and otherwise, the realisation would have a bias and should be rejected from the set of realisations.

We define a probability measure μ^r on X by $\mu^r(\{i\}) = m_i^r/M$; similarity we define $as\ \mu^s$. The probability measures $\mu^r$and μ^s describe the normalised distribution of mass amongst the N blocks in simulations r and s, respectively. We denote the space of all probability measures on X by $M(X)$.

We are now applying the Kantorovich metric $D: M(X) \times M(X) \to R^+$ in our setting. For further background on the Kantorovich metric, please refer to chapter 5 and the references therein.

$$D(\mu^r, \mu^s) = \min_{\tilde{\mu} \in M(X) \times M(X)} \left\{ \sum_{i,j=1}^{N} d(i,j)\, \tilde{\mu}(i,j)) : \tilde{\mu}(i, X) = \mu^r(i), \tilde{\mu}(X, j) = \mu^s(i), 1 \leq i, j \geq N, \right\} \tag{6.2}$$

We illustrate why this metric is a good notion of distance for conditional simulations with an example. Consider the 'extreme' pair of realisations r and s, where all of the valuable material in realisation r and s are contained in blocks i and j , respectively. Then $\mu^r(\{i\}) = 1, and\ \mu^r(\{k\}) = 0, k \neq i$ and $\mu^s(\{i\}) = 1, and\ \mu^s(\{k\}) = 0, k \neq i$. The minimising $\tilde{\mu}$ in equation (6.2) is the probability measure satisfying $\tilde{\mu}\big((i,j)\big) = 1$ and $\tilde{\mu}\big((k,l)\big) = 0,\ (k,l) \neq (i,j)$, thus $D(\mu^r, \mu^s) = d(i,j) = \|c_i - c_j\|_2$ is the three-dimensional Euclidean distance between the centres of block i and block j.

Thus, if these special blocks i and j are spatially close, the distance between simulations r and s is small; the opposite is true if i and j are spatially distant. The metric d is, thus, consistent with a reasonable geological notion of 'distance' between simulations. For more realistic simulations than those in example 2, where the mass is distributed amongst the N blocks, the metric d provides a sum of Euclidean distances, weighted by how much mass needs to be transferred over these distances. Exactly how much mass needs to be transferred is determined by the minimisation in equation (6.2). In the next section we describe how to set up a simple optimisation problem to calculate D.

6-3 Kantorovich distance computation

In this section we describe how to set up a simple optimisation problem to calculate distances between all pairs of realisations (inter-point distances) to construct the distance matrix or the space of uncertainty.

Our notion of dissimilarity for realisations r and s will be based upon the 'work' required to 'transform' realisation r into realisation s. Both realisations contain approximately the same total mass M, but the spatial distribution of mass is different. To 'transform' realisation r into realisation s, mass (the metal content) must be moved between individual blocks. The total 'work' for moving mass will be proportional to the mass moved and the *distance moved*.

$$\min_{f^{rs}} \quad \sum_{i,j=1}^{N} d_{ij} f_{ij}^{rs} \qquad (6.3)$$

$$\text{Subject to} \quad \sum_{j=1}^{N} f_{ij}^{rs} = m_i^r, \qquad 1 \le i \le N \qquad (6.4)$$

$$\sum_{i=1}^{N} f_{ij}^{rs} = m_j^s, \qquad 1 \le j \le N \qquad (6.5)$$

$$f_{ij}^{rs} \ge 0, \qquad 1 \le i, j \le N \qquad (6.6)$$

One way to think of this transportation problem is as follows. The mass in realisation r needs to be moved to the mass configuration in realisation s. The value f_{ij} represents the total mass in the ith block in simulation r that is moved to the jth block in realisation s.

If blocks i and j are spatially distant, the corresponding distance penalty per unit of mass moved, d_{ij}, will be large and such a move will be discouraged. However, if blocks i and j are spatially close, moving mass from i in realisation s to j in realisation r is comparatively attractive. The linear program (6.3)-(6.4) finds the minimal amount of work (mass moved multiplied by distance moved) to turn realisation s into realisation r. This minimum amount of work defines the distance between realisations r and s.

This problem is a transportation problem and is easily solved using the simplex method or other specialised methods.

6-3.1 Transportation problem

The transportation problem (TP) is a well-known mathematical programming problem classified as a special type of linear program (LP), which tries to find the minimum cost to transport (supply) mass, goods or any commodity from a set of sources or suppliers $i = 1, \cdots, M$ to a set of destinations or demanders $j = 1, \cdots, N$. For example, assume there are M iron ore mines and N smelters where the iron ore that the mines produce is consumed. Mine i has a supply of s_i units, and smelter j has a demand of d_j units. The cost per iron ore unit (for example, tonne) transported from mine i to smelter j is denoted by d_{ij} , and the number of iron ore units transported is denoted by x_{ij}.

Assume that the total iron ore production for all M mines, $s_t = \sum_{i=1}^{M} s_i$, and the total demand for iron ore for N smelters $d_t = \sum_{i=1}^{N} d_i$ is equal to each other. The transportation problem is to compute:

$$\min_{x_{ij}} \sum_{i=1}^{M} \sum_{j}^{N} x_{ij} d_{ij} \quad (6.7)$$

Subject to:

$$\sum_{i}^{M} x_{ij} = s_t \;\; i = 1, \cdots, M \quad (6.8)$$

$$\sum_{j}^{N} x_{ij} = d_t \quad j = 1, \cdots, N \quad (6.9)$$

$$x_{ij} \geq 0 \;\; i = 1, \cdots, M, \quad j = 1, \cdots, N \quad (6.10)$$

If the total supply and demand are not equal, the given constraints are not satisfied. The equality constraints can be written as below:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 & & & & & & & & & & & & \\ & & & & 1 & 1 & \cdots & 1 & & & & & & & & \\ & & & & & & & & \ddots & & & & & & & \\ & & & & & & & & & & & & 1 & 1 & \cdots & 1 \\ 1 & & & & 1 & & & & & & & & 1 & & & \\ & 1 & & & & & 1 & & & & \cdots & & & 1 & & \\ & & \ddots & & & & & \ddots & & & & & & & \ddots & \\ & & & & 1 & & & & 1 & & & & & & & 1 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \\ \vdots \\ x_{m1} \\ x_{m2} \\ \vdots \\ x_{mn} \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \\ d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

The structure of this coefficient matrix can be exploited to improve both the time and the space required by the simplex algorithm on a transportation problem (Rubner et al., 1998)

The transportation simplex algorithm can still be applied when the total supply s_t is not equal to the total demand d_t, which is known as an unbalanced transportation problem. There are two cases: first, if supply s_t is greater than the total demand d_t

The goal is still to find the minimum cost to satisfy all the demand. In this case, however, there will be excess supply after the demand has been satisfied. The LP for the unbalanced case is:

$$\min_{x_{ij}} \sum_{i=1}^{M} \sum_{j}^{N} x_{ij} d_{ij} \quad (6.11)$$

Subject to:

$$\sum_{i}^{M} x_{ij} \leq s_t \;\; i = 1, \cdots, M \quad (6.12)$$

$$\sum_{j}^{N} x_{ij} = d_t \quad j = 1, \cdots, N \quad (6.13)$$

$$x_{ij} \geq 0 \;\; i = 1, \cdots, M, \quad j = 1, \cdots, N \quad (6.14)$$

In order to apply the transportation simplex method, we convert the unbalanced TP to an equivalent balanced TP. This is done by adding a dummy demand $n + 1$ with demand $d_{n+1} = s_t - d_t$, and for which $d_{i,n+1} = 0$ (cost of transporting) for $i = 1, \cdots, M$. The total demand in the modified problem is equal

to the total supply and the minimum cost is the same for the balanced and unbalanced problems. The dummy demand gives the suppliers a place to dump their leftover supply at no cost.

The second is in case demand d_t is greater than the total supply s_t ,
Similar to the first one, we can add a dummy supplier with $s_{n+1} = d_t - s_{t,}$ where the costs associated with dummy supplier is set equal to zero.

A detailed description of the transportation simplex method can be found in Villani, (2003).

6-4 Construction of a dissimilarity distance matrix

One of the most important terminologies in distance-based methods, frequently used in this thesis, is the dissimilarity distance matrix, which is explained here.

If $\hat{f}^{rs}$ is the minimising array of mass equation (6.3), then we define the dissimilarity or distance between realisations r and s to be

$$D_{rs} = \sum_{i,j=1}^{N} d_{ij}\, \hat{f}_{ij}^{rs} \qquad (6.15)$$

Thus, for a set of S generated realisations, the distance matrix $S \times S$ can be constructed by calculating the pairwise distances between S generated realisations. This matrix is called the dissimilarity distance matrix. This matrix reveals an underlying structure of variability between realisations and would be a base for any other calculation, clustering, comparison and visualisation that may be needed to evaluated the space of uncertainty.

As was mentioned in chapter 5, the Kantorovich distance D_{rs} satisfies all properties (nonnegative, symmetry and triangle inequality) expected of a notion of distance in the metric space. In this case, the dissimilarity distance matrix is a square symmetric matrix with zero diagonal elements D_{ii} = 0.

As the dissimilarity matrix is symmetric, the number of calculations of pair distance (n_s) for S generated realisation can be given by this formula $n_s = \mathrm{S}(\mathrm{S}-1)/2$. For example, for 100 realisations 4950 distance pairs calculation should be done to construct the dissimilarity distance matrix.

Furthermore, note that D_{rs} (in the case of evaluating grade uncertainty) has the units of mass × Euclidean distance, so we may interpret the quantity D_{rs}/M as the average distance that 1 mass unit of material is moved to transform realisation r into realisation s. D_{rs} may have different units, for example, time × Euclidean distance.

6-5 Why Kantorovich distance is the robust candidate?

Distance metric is widely used in dissimilarity measurement. Choosing an appropriate distance function to measure the dissimilarity for a given problem is an important step towards solving it. The main purpose of measuring of dissimilarity between the realisations or geological block models (in this study) is to compare two models, namely two different 2D or 3D *spatial mass (grade) distributions* in order to compute a single number which evaluates their dissimilarity between them.

Our target in this study is to quantify the dissimilarity between two K dimensional datasets, namely two block models (K is number of blocks in a model) where each block has certain amount of mass (grade, attribute or any feature) and also individual coordinates (x, y and z).

Geostatistics is about spatial variability of the variables (regionalised variables) that have an attribute value and also a location in a two or three dimensional space. The key point in geostatistics is the assumption of spatial dependency. That means, the location of data with respect to one another plays an important role in any geostatistical analysis such as modelling, simulation, and estimation procedures. Therefore a robust measure of dissimilarity must take into account not only the attributes, but distances between the blocks (spatial information) as well. In the other word, we need to define a distance function that allows us to take into account the differences in locations (spatial information) and attributes together

Although there are quite a few references which used and suggested Kantorovich distances as a robust method for matching multidimensional distributions and dissimilarity measure (Assent et al. 2006), (Tang et al.2013), (Coen 2007), (Jovic et al. 2007) and the most important one (Armstrong et al. 2012); we explain here why Kantorovich is the robust distance for dissimilarity measure of the generated realisations in comparison with Euclidean distance.

The most of distance functions (including Euclidian distance) are normally used in the context of comparing pairs of variables, cases or in general, between two N dimensional vectors (N is number of attributes or features). That means, these attributes or features are that basically contribute to measure a distance

(dissimilarity) between vectors not spatial information. Therefore that would be a very common drawback for the most of well-known distances functions. For example all distance functions are mentioned in Table 5-1 have same drawback.

6-5.1 Kantorovich distance vs. Euclidean distance

We begin by measuring dissimilarity through the Euclidean distance, and present three pairs of models (shown in Figures 6-3, 6-4 and 6-5) to illustrate the robustness of Kantorovich distance over Euclidean distance. We will see that all the three pair models have the same Euclidean distance (dissimilarity) from each other, while they have very different spatial grade distributions.

If there are two vectors X and Y the Euclidean distance can be defined:

$$d(X,Y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2}$$

$$X = (x_1, x_2, \dots, x_N)$$

$$Y = (y_1, y_2, \dots, y_N)$$

Assume there are two block models or realisations X and Y with following attributes (see Table 6-1) or in the other word, there are two vectors (X and Y) of blocks attributes in R^5. Euclidean distance between these vectors is $d(X,Y) = 0.40$.

Table 6-1: Grades of 5 blocks in the realisations X and Y for the pairs of models 1 and 2

Block No.	1	2	3	4	5
Realisation X	0.90	0.65	0.45	0.25	0.80
Realisation Y	1.20	0.55	0.25	0.35	0.70

$$X = (0.90, 0.65, \dots, 0.80)$$

$$Y = (1.2, 0.55, \dots, 0.70)$$

$$d(X, Y) = 0.40$$

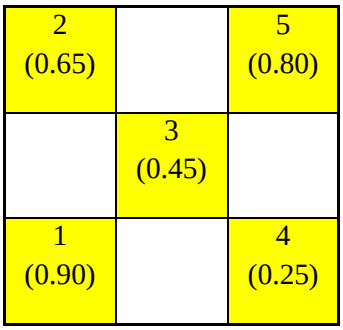
As clearly shown, the dissimilarity based on Euclidean distance between these two vectors (realisations) is only depends on root of square differences between grades (features) of the pair of corresponding blocks, and the Euclidean distance function doesn't take into account anything about *where these blocks are (spatial information) in the model* or *how far these blocks are from each other*, and more important point *what are the grades of neighbouring blocks.*

Now, we apply Kantorovich metric on the same blocks' grades in Table 6-1, but for two different pairs of models are shown in Figures 6-3 and 6-4 (different blocks combinations) to illustrate how well this metric can handle the spatial information and attributes to gather.

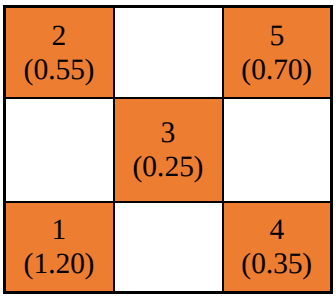
Two realisations X_1 and Y_1 (block models) with 5 blocks (2D) are shown in Figures 6-3. The block size is 20×20, and distance between the centers of 5 blocks are shown in the matrix $M1$. By applying the Kantorovich distance function, and calculating distances $D_{X_1Y_1}$ between 5 blocks of the given realisations a measure of their dissimilarity $D_{X_1Y_1}$=15.66 is obtained.

$M1 =$

Block 1	Block 2	Block 3	Block 4	Block 5	
0	50	28.3	50	70.7	Block 1
50	0	28.3	70.7	50	Block 2
28.3	28.3	0	28.3	28.3	Block 3
50	70.7	28.3	0	50	
70.7	50	28.3	50	0	Block 4



Realisation X_1



Realisation Y_1

Figure 6-3: Realisation X_1 and Y_1 with 5 blocks and their grades, our approach is to computing distance (dissimilarity) between two spatial mass distributions, namely yellow blocks and red blocks. The Kantorovich distance between these models is $D_{X_1Y_1}$=15.66.

Now, we keep the grades the same, but put the blocks closer to each other than what was in Figures 6-3 to get the following block models X_2 and Y_2 (see Figure 6-4).The block size is 20×20, and distance between the centers of blocks are shown in the matrix $M2$. By applying the Kantorovich distance function, and calculating distances $D_{X_2Y_2}$ between 5 blocks of the given realisations a measure of their dissimilarity $D_{X_2Y_2}$=10.83 is obtained.

The difference between $D_{X_1Y_1}$ and $D_{X_2Y_2}$ clearly shows Kantorovich distance is consistent with a reasonable geological notion of "distance" between block models (or realisations). That means if the blocks are spatially close, the Kantorovich distance between them is small and vice versa. That's why we see dissimilarity between models X_2 and Y_2 less than between models $X_1$and Y_1 ($D_{X_2Y_2} < D_{X_1Y_1}$).

It is obvious, in spite of significant and obvious differences between two pairs models, the results of not only Euclidean distance, but also the most of distance functions for those models (Figure 6-3 and 6-4) are not affected by changes in distance, and they are not able to distinguish any deference between these two pairs of models, while the Kantorovich distance appears a meaningful difference (dissimilarity) between the models.

$$M2 = \begin{array}{c} \begin{matrix} \text{Block 1} & \text{Block 2} & \text{Block 3} & \text{Block 4} & \text{Block 5} \end{matrix} \\ \begin{bmatrix} 0 & 20 & 20 & 28.3 & 51 \\ 20 & 0 & 28.3 & 20 & 50 \\ 20 & 28.3 & 0 & 20 & 28.3 \\ 28.3 & 20 & 20 & 0 & 50 \\ 51 & 50 & 28.3 & 50 & 0 \end{bmatrix} \end{array} \begin{matrix} \text{Block 1} \\ \text{Block 2} \\ \text{Block 3} \\ \text{Block 4} \end{matrix}$$

		5 (0.80)
	3 (0.45)	4 (0.25)
	1 (0.90)	2 (0.65)

Realisation X_2

		5 (0.70)
	3 (0.25)	4 (0.35)
	1 (1.20)	2 (0.65)

Realisation Y_2

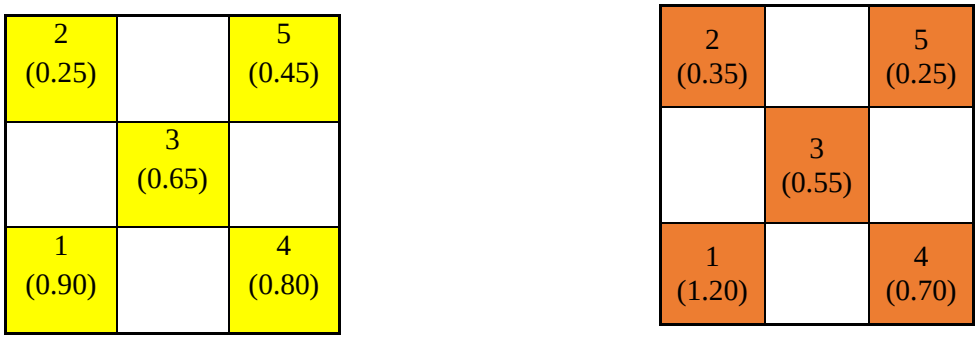
Figure 6-4: Realisation X_2 and Y_2 with 5 blocks and their grades, our approach is to computing distance (dissimilarity) between two spatial mass distributions, namely yellow blocks and red blocks. The Kantorovich distance between these models is $D_{X_2Y_2}$=10.83. The blocks in these models are closer to each other than the blocks in Figure 1 that's why the $D_{X_2Y_2} < D_{X_1Y_1}$.

Now, we present another example to show how Kantorovich distance can even handle swapping grades in the models (without changing the grades) while Euclidean distance cannot reveal any difference. In the following examples, we only shuffle the grades of pair blocks 2, 3, 4 and 5 to have the new pair of models X_3 and Y_3. Table 6-2 and Figures 6-5 show the grades and these pair of models, respectively.

It is obvious these new block models are quite different from the previous one (X_1 and Y_1), as the *spatial mass (grade) distributions are completely changed*, but the Euclidean distance between them remains exactly the same $d(X, Y) = 0.40$. While Kantorovich reveals this difference by showing a higher distance $D_{X_3Y_3} = 18.49$ (dissimilarity) between them than $D_{X_1Y_1}$.

Table 6-2: Grades of 5 blocks in realisations X_3 and Y_3 for pair of models 3

Block No.	1	2	3	4	5
Realisation X_3	0.90	0.25	0.65	0.80	0.45
Realisation Y_3	1.20	0.35	0.55	0.70	0.25



Realisation X_3 Realisation Y_3

Figure 6-5: Realisation X_3 and Y_3 with 5 blocks and their grades, our approach is to computing distance (dissimilarity) between two spatial mass distributions, namely yellow blocks and red blocks. The Kantorovich distance between these models is $D_{X_3Y_3}$=18.49. The high grade blocks in these models are far from low grades in comparison with Figure 1, and that's why the $D_{X_3Y_3} > D_{X_1Y_1}$.

As is clearly illustrated through the examples, Kantorovich distance is very sensitive to any changes in spatial mass distributions of the realisations, namely spatial locations and attributes, and is able to reveal dissimilarly between the models very well. That's what we really need from a good distance function.

But what is the key point in Kantorovich metric that gives it an advantage over the other distances? As that is illustrated, we imagine the blocks grades of realisation X_i (i =1,2 and 3) need to deliver into the realisation Y_i (j =1,2 and 3), therefore there are five options (in model Y_i) for transporting a grade of a block from model X_i into model Y_i, and the distances (spatial information) between blocks play critical role for this mass transportation. Kantorovich distance tries to transport masses (attributes) to the nearest blocks first. That means the Kantorovich distance in contrary to the other distance functions doesn't act as a simple pairwise distance between blocks, but rather is able to take into account all possible options (blocks) where are around to the block, and gives priority to closer ones. That is what we need in geostatistics, considering not only grades of corresponding blocks, but also the grades of neighboring blocks to measure dissimilarity between them.

We believe that these examples proved that Kantorovich distance is a robust candidate for measuring our intuitive notion of dissimilarity between blocks models (realisations) which may have different attributes and features.

6-6 Illustration of the concept with a simple example

Before going through a realistic application of this approach, we give a very simple example to describe the proposed approach.

Assume there are just four generated realisations as given in Figure 6-1 and 6-2. By applying the Kantorovich metric (distance function) and calculating the pairwise distances D_{rs} between the given four images (see Figure 6-6) a measure of their similarity is obtained.

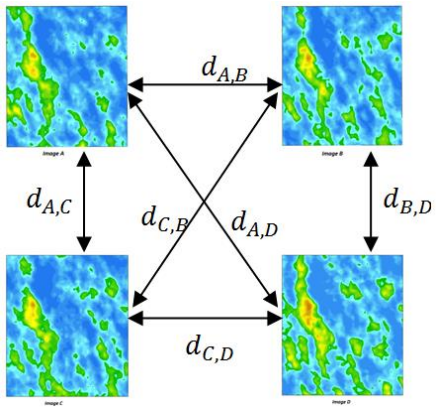


Figure 6-6: Calculating 6 pairwise distances D_{rs} between the given four realisations

As it was mentioned before, the number of calculations of the pair distance (n_s) for 4 realisations would be 6 and the following 4×4 dissimilarity distance matrix can be constructed.

$$\begin{array}{c} \\ ImageA \\ ImageB \\ ImageC \\ ImageD \end{array} \begin{array}{c} \begin{array}{cccc} ImageA & ImageB & ImageC & ImageD \end{array} \\ \begin{pmatrix} 0 & 4,548,600 & 6,798,720 & 5,564,880 \\ 4,548,600 & 0 & 6,668,630 & 4,936,460 \\ 6,798,720 & 6,668,630 & 0 & 9,995,100 \\ 5,564,880 & 4,936,460 & 9,995,100 & 0 \end{pmatrix} \end{array}$$

As it was explained before, D_{rs} has the unit of mass × Euclidean distance; thus, by calculating the quantity of $D_{4\times4}/M$ the ground distances between realisations can be presented.

$$D_{4\times4}/M = \begin{pmatrix} 0 & 4.84 & 7.24 & 5.92 \\ 4.84 & 0 & 7.10 & 5.26 \\ 7.24 & 7.10 & 0 & 10.64 \\ 5.92 & 5.26 & 10.64 & 0 \end{pmatrix}$$

As can be seen in the matrix above, the image A and B (Figure 6-1) are at minimum distance from each other (4.84 m) while images C and D (Figure 6-2) are at maximum distance (10.64 m).

For visualisation purposes, we choose $n = 2$ or 3. Figure 6-7 shows a multidimensional scaling embedding of the 4 realisations in R^2 and R^3. The interpoint distances in R^3 have better approximation for the pairwise realisation distances D_{rs} than R^2.

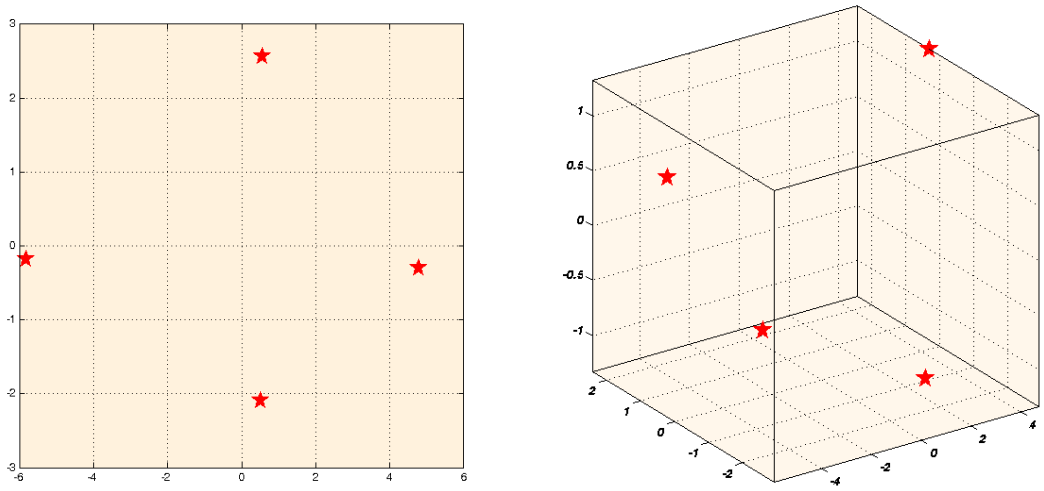


Figure 6-7: Multidimensional scaling embedding of the four realisations in R^2 (left side) and R^3 (right side)

6-7 Realisation reduction

We now suppose that we have constructed a large number of realisations numbered $1, \ldots, \boldsymbol{S}$. With increasing computer performance, it is possible to construct, for example, $S \approx 1000$ realisations for block models of size $N \approx 10^7$ in 48 hours on a desktop PC using techniques such as SGS (it would be faster if a direct block Kriging is used). We will assume that the collection $\{1, \ldots, \boldsymbol{S}\}$ well samples the distribution of all possible realisations and uses this collection as the benchmark against which we compare subcollections.

Remark 1: We remark that in practice N will typically be larger than S so this assumption is in most cases false. Nevertheless, for many applications of realisations, many more realisations can be created than used, whether in an optimisation process (Ramazan and Dimitrakopoulos 2004, Menabde et al. 2004, Froyland et al. 2004, Boland et al. 2008, Tarhan et al. 2009 and Newman et al. 2010) or as part of a small collection of possible geological outcomes interrogated by engineers or geologists. It is, therefore, important to choose S as large as is practicable and then reproduce the variety contained in these S realisations as best as possible with a significantly smaller number of $S \ll \boldsymbol{S}$ realisations to use in optimisations or other applications.

Remark 2: Sampling (realisation reduction) from an original collection $\boldsymbol{S}$ can be done based on different distance criteria, such as the minimum or the maximum distance between the sub-collection (sampled realisations) and the original collection $\boldsymbol{S}$ or any other specified distance criteria which may be defined according to the purpose of sampling.

In Section 2 we identified each realisation s with a probability measure of μ^s on $X = \{1, \dots, N\}$ and used the Kantorovich metric to define a distance function on $M(X)$. We again make this identification and consider the collection $\mathfrak{S} = \{\mu^1, \dots, \mu^{\boldsymbol{S}}\}$. In order to optimally subsample the collection $\mathfrak{S}$ we consider the space of probability measures on $\mathfrak{S}$, $M(\mathfrak{S})$. Principal amongst the elements of $M(\mathfrak{S})$ is our reference measure which gives an equal weight to each of the $\boldsymbol{S}$ generated conditional simulations.

$$\mathcal{V} := \frac{\sum_{r=1}^{\boldsymbol{S}} \mu^r}{\boldsymbol{S}} \qquad (6.16)$$

For $S \leq \boldsymbol{S}$, we seek probability measures $\mathcal{V}_S$ of the form

$$\mathcal{V}_S = \sum_{k=1}^{S} p_{s_k} \mu^{s_k} \quad (6.17)$$

Where $p_{s_k} \geq 0$, $1 \leq k \leq S$, $\sum_{k=1}^{S} p_{s_k} = 1$ which are as close as possible to $\mathcal{V}$.

$$M_S(\mathfrak{S}) = \left\{ \mathcal{V}_S \in M(\mathfrak{S}) : \mathcal{V}_S = \sum_{k=1}^{S} p_{s_k} \mu^{s_k}, \sum_{k=1}^{S} p_{s_k} = 1, \; p_{s_k} \geq 0, 1 \leq s_k \leq \boldsymbol{S} \right\} \quad (6.18)$$

We define as the class of probability measures that use at most S of the building block and probability measures s associated with individual conditional simulations.

In order to define closeness, we again turn to the Kantorovich metric. The definition below is the same as (6.2), except that we have replaced X with $\mathfrak{S}$ and d with D.
For $\mathcal{V}$ as in (6.16) and $\mathcal{V}_S$ as in (6.17), we define

$$D(\mathcal{V}, \mathcal{V}_S) = \min_{\tilde{\mathcal{V}} \in M(\mathfrak{S}\times\mathfrak{S})} \left\{ \sum_{k,l=1}^{\boldsymbol{S}} D(\mu^k, \mu^l)\, \tilde{\mathcal{V}}(\mu^k, \mu^l) : \tilde{\mathcal{V}}(\mu^k, \mathfrak{S}) = 1/\boldsymbol{S},\ \tilde{\mathcal{V}}(\mathfrak{S}, \mu^l) = p_l, 1 \le k, l \le \boldsymbol{S} \right\} \quad (6.19)$$

Following the procedure in the previous section, we have available to us numerical values for the distances between each pair of conditional simulations s and r, namely $D(\mu^s, \mu^r)$, $1 \le r, s \le \boldsymbol{S}$. Thus, in principle, given any $\mathcal{V}_S = M_S(\mathfrak{S})$ we can calculate $D(\mathcal{V}, \mathcal{V}_S)$. We find

$$\mathcal{V}_S^* = arg \min_{\mathcal{V}_S \in M_S(\mathfrak{S})} D(\mathcal{V}, \mathcal{V}_S) \quad (6.20)$$

We will again use the idea of 'transformation' to transform the larger set of $\boldsymbol{S}$ realisations into a smaller set of S realisations with minimal 'work'. Let F_{rs} denote the probability flow from realisation r to realisation $1 \le r, s \le \boldsymbol{S}$. Because we are reducing $\boldsymbol{S}$ realisations to S realisations, there will be at most S special realisations s for which F_{rs} may be greater than zero; that is, only the S special realisations in the reduced collection are allowed to receive a positive probability flow. 'Work' will now become a product of the probability of F_{rs} and the distance D_{rs}.

The special realisations will be chosen using binary variables $x_s, s = 1, \ldots, \boldsymbol{S}$, with $x_s = 1$ signifying that $s \in \{s_1, \ldots, s_s\}$ (the realisation s is selected for the subcollection) and $x_s = 0$ otherwise. We do not insist that each realisation $s_r;\ r = 1, \ldots, S$ have an equal probability, but that the weightings of the chosen realisations can vary. The weight for realisation s will be denoted by ω_s. We state the following mixed integer linear program (MILP) and then describe the effect of the various constraints.

$$z_S = \min_{F, x, \omega} \sum_{r,s=1}^{\boldsymbol{S}} D_{rs} F_{rs} \quad (6.21)$$

Subject to:

$$\sum_{s=1}^{\boldsymbol{S}} F_{rs} = 1/\boldsymbol{S} \qquad 1 \le r \le \boldsymbol{S} \qquad (6.22)$$

$$\sum_{r=1}^{\boldsymbol{S}} F_{rs} = \omega_s \qquad 1 \le s \le \boldsymbol{S} \qquad (6.23)$$

$$\sum_{s=1}^{\boldsymbol{S}} x_s = S \quad (6.24)$$

$$\sum_{s=1}^{\boldsymbol{S}} \omega_s = 1 \quad (6.25)$$

$$\omega_s \le x_s\,, \quad 1 \le s \le \boldsymbol{S} \qquad (6.26)$$

$$F_{rs} \ge 0, \qquad 1 \le r, s \le \boldsymbol{S} \qquad (6.27)$$

$$\omega_s \ge 0\,, \qquad x_s \in \{0,1\}, \qquad 1 \le s \le \boldsymbol{S} \quad (6.28)$$

Equations (6.21) and (6.22) ensure that the probability flow out of each realisation is $1/\boldsymbol{S}$ and that the probability flow into realisation s is ω_s, respectively. Equality (6.24) ensures that exactly S realisations are chosen for the sub-sampled collection. Equality (6.25) guarantees that the sum of the weights ω_s in the chosen sub-collection is 1. Inequality (6.26) forces the weight assigned to realisation s to be zero unless realisation s is selected for the subsample. As in the transportation problem (6.3)-(6.6), the problem (6.21)-(6.28) transports a probability of $1/\boldsymbol{S}$ from the original very large collection of S realisations onto a smaller collection of $\boldsymbol{S}$ realisations. The penalty for the probability flow from realisation r to realisation s is proportional to both the size of the flow F_{rs} and the distance between the realisations D_{rs}. Thus, if there is a set of several realisations with mutually small distances D_{rs}, it is likely that this set will be replaced by one realisation from the set with all the probability of that set owing to the one realisation that now 'represents' that set. Having solved the MILP above, one has the minimising arrays $\hat{F}$, $\hat{x}$, and $\hat{\omega}$. The selected sub-collection consists of those s with $\hat{x_s} = 1$ and the corresponding weights are given by the vector $\hat{\omega_s}$, which has at most S positive entries. The value z_S has the units of D_{rs} (mass×Euclidean distance) as F_{rs} is dimensionless. Thus, z_S has the interpretation of 'work' required transforming all of the S realisations into $S \le \boldsymbol{S}$ realisations. The quantity z_S/M has the units of Euclidean distance and thus may be interpreted as the average distance over which a single unit of mass is moved to effect this transformation.

6-8 Selecting representative schedules

Determining the set of realisation M in generated models set N, which can clearly represent the original set, is a challenging problem even in commonly used clustering methods. In most clustering methods M has to be assumed as a known parameter, which should be chosen by users.

Below we describe how sub-sampling works:

First, it is important to describe how subsampling can be used to estimate how well all N simulations sample the limiting 'true' distribution of simulations.

The next step it to denote the 'true' distribution of simulations as $\mathcal{V}_\infty$. If we wish to determine $I(N,\infty)$, it will be accomplished by extrapolating from known distances.

Denote $\alpha_{P,Q} = d(\mu_{P,N}, \mu_{Q,N})$, $1 \le P \le Q \le N$. In practice, we will compute some of these values, the more the better.

Denote $F = \{(P,Q) \in \mathbb{z}^2, 1 \le P \le Q\}$ we will fit a function $f: F \to \mathbb{R}$ to the $\alpha_{P,Q}$ function obeying the following properties:

1- $f \ge 0$
(Distance must be non-negative),
2- $f(P,P) = 0$ for all $P \ge 1$
(One can perfectly represent P simulations with P simulations),
3- $f(P,Q) > 0$ for all $1 \le P \le Q$
(By Lemma A.1 one cannot perfectly represent Q simulations with P < Q simulations),
4- $f(P,Q) \ge f(P + R, Q)$ for all $1 \le P + R \le Q$
More realisation cannot produce a worse representation.
Once we have fitted a function f obeying these properties to our computed $\alpha_{P,Q}$ values, we then calculate $1 - f(N,\infty) = f(1,\infty)$ which is supposed to approximate the fractional improvement of the N scenarios $\mathcal{V}_N$ over the best deterministic single simulation.

In fact, we are estimating the fractional improvement of the best N scenarios, whereas in practice we merely have some N scenarios. Thus, our estimations on improvement are more favourable than in reality and estimate upper bounds on how well the N scenarios represent the limiting distribution. Two issues arise:

1. How representative would my N scenarios be of $10N$ scenarios?
   Compute $1 - f(N, 10N) = f(1,10N)$.

2. How much better could 10N scenarios represent the limiting distribution?
   Compute $1 - f(10N, 1) = f(1, \infty)$.

Following Dupacova et al. (2003) and Growe et al. (2003), we will report $D(\mathcal{V}, \mathcal{V}_S)$ relative to $d(\mathcal{V}, \mathcal{V}_1)$, the latter representing the "base" distance between the best deterministic approximation of $\mathcal{V}$. We will also report relative to $d(\bar{\mathcal{V}}_N, \mathcal{V}_N)$ where $\bar{\mathcal{V}}_N$ can be the Kriging model, 'E-type' (average of the N realisations) or any other model.
Define $I_{M,N}$ to be the fractional improvement in distance from $\mathcal{V}_N$ that one can achieve by using M scenarios rather than the best single scenario.

$$I_{M,N} = 1 - \frac{d(\mu_{M,N}, \mathcal{V}_N)}{d(\mu_{1,N}, \mathcal{V}_N)} \qquad (6.29)$$

Similarly, define $I'_{M,N}$ to be the fractional improvement in distance from $\mathcal{V}_N$ that one can achieve by using M scenarios rather than the single E-type scenario.

$$I'_{M,N} = 1 - \frac{d(\mu_{M,N}, \mathcal{V}_N)}{d(\bar{\mathcal{V}}_N, \mathcal{V}_N)} \qquad (6.30)$$

Clearly $I_{N,N} = I'_{M,N} = 1$ and $I_{1,N} = I'_{1,N} = 0$

For brevity, we call the quantity $z_S = d(\mu_{M,N}, \mathcal{V}_N)$, $z_1 = d(\mu_{1,N}, \mathcal{V}_N)$. Thus, z_S/z_1 represents the distance between the optimally subsampled S realisations and the full set of $\boldsymbol{S}$ realisations relative to the distance between the best single realisation (deterministic approximation) and the full set of **S** realisations. We call the quantity the *relative accuracy* of the optimal S-subcollection.

$$I_S = \left(1 - \frac{z_S}{z_1}\right) \times 100\% \quad (6.31)$$

Figure 6-8 shows distance z_S between the S-subsample and the full sample of S realiations (blue) and I_S relative accuracy (red) for 400 realisations.

As can be seen, on one hand, by increasing the number of sub-samples the relative accuracy (I_S) of the optimal S-subcollection increases reaching 100%, if all samples are taken. On the other hand, distance z_S between samples and the optimal S-subcollection decreases to zero if all samples are taken.

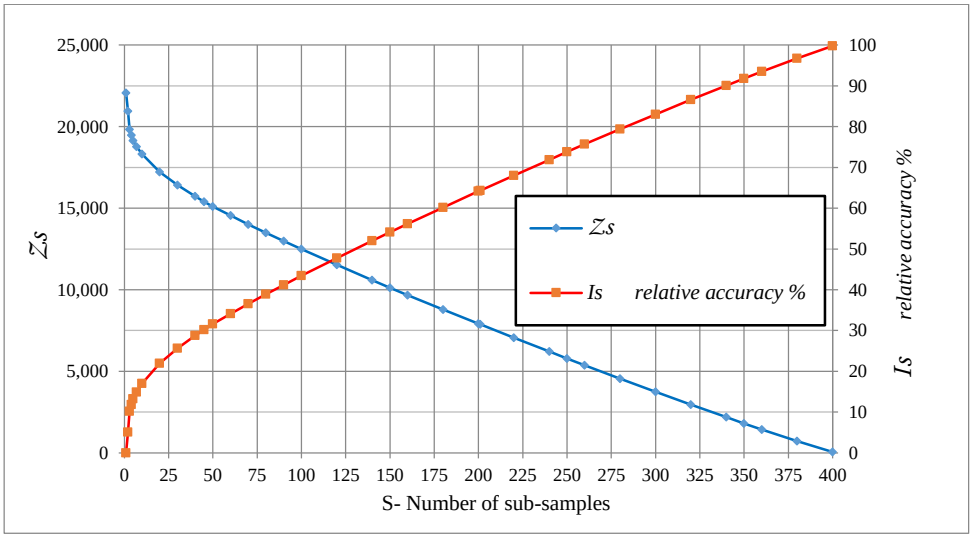


Figure 6-8: Distance $z_S$between the S-subcollection and the full collection of $\boldsymbol{S}$ simulations (red) and relative accuracy (blue)

6-9 Conclusion

In this section we have introduced and applied the Kantorovich distance for the following two main purposes.

Firstly, we have developed a practical quantitative methodology to define and map the space of uncertainty by computing the Kantorovich distance between generated realisations. This metric is a good notion of distance for measuring dissimilarity between geological realisations.

Indeed, the space of uncertainty is quantified by constructing the dissimilarity distance matrix D_{ij}, which contains all pairwise distances. To illustrate, a simple example was presented to show how this method works.

Secondly, in order to select the sub-collection of realisations that best represents the possible outcome of stochastic simulation algorithms, we used the concept of Kantorovich distance and developed a simple optimisation model to find the best samples and quantify how well this high-quality subsample represents the overall uncertainty of the collection. The optimisation model can be used as a general tool, which is able to select any subset of representative realisations against user-defined criteria. For example, our methodology can determine the smallest number of conditional simulations that are required to cover 75% of the total geological uncertainty. Moreover, our approach identifies the corresponding conditional simulations.

Chapter 7

7 Numerical results on geostatistical simulations

7-1 Introduction

As mentioned in pervious chapters, the mathematical algorithm which measures the similarity between realisations can be classified as a transportation problem that is usually solved using the simplex method or other specialised methods. However, we are faced with the time-consuming tasks associated with solving the mathematical problem as we used a very large number of realisations for this study. We believe it is worth spending more time to create a larger number of realisations to attempt to better capture the attributes of the space of uncertainty. This study is quite unique in the way of its generated number of realisations (see Figure 7-1). However, there is definitely no need to generate so many realisations even for mine planning. As we will explain later in this chapter, after generating some realisations the structure of space of uncertainty becomes stable and adding more realisations doesn't significantly change it.

The great number of resource consumption (CPU and memory) required to solve the mathematical problems for this chapter were all accomplished by the Supercomputer of the School of Mathematics and Statistics at UNSW. For example, 143,000 hours of CPU time (this time is summation of all CPU's running time in parallel) were used to solve the mathematical problem for all simulation algorithms and their realisations. That shows the applied algorithm need a huge computation times, but this distance has been widely used in multimedia information retrieval systems in large-scale databases (larger than typical ore bodies in mining industry), so there are quite a few techniques that are used in the implementation of the faster algorithms. These references (M. Shishibori and et al, 2011) and (O. Pele and M. Werman, 2009) are good examples to introduce two powerful algorithms which are able to solve the problem much faster.

Although these fast techniques are able to save the significant time; the plummeting cost of computing power and storage thanks to the trend of cloud computing has radically changed the computing market. It is vastly less expensive to rent cloud computing than the old way (such as on PCs or even on local supercomputers) of doing computational jobs. For example, the cost of renting 240 GB RAM with 32 cores (CPUs) per hour is just about few dollars! And that is expected to reach to a few cents very soon. Therefore that allows us to get as many as cores and RAM that may be needed. Fortunately, the most of calculations for our approach (comparison between realisations) can be carried out simultaneously, so we can take advantage of parallel computing to decrease the computation time very cheaply.

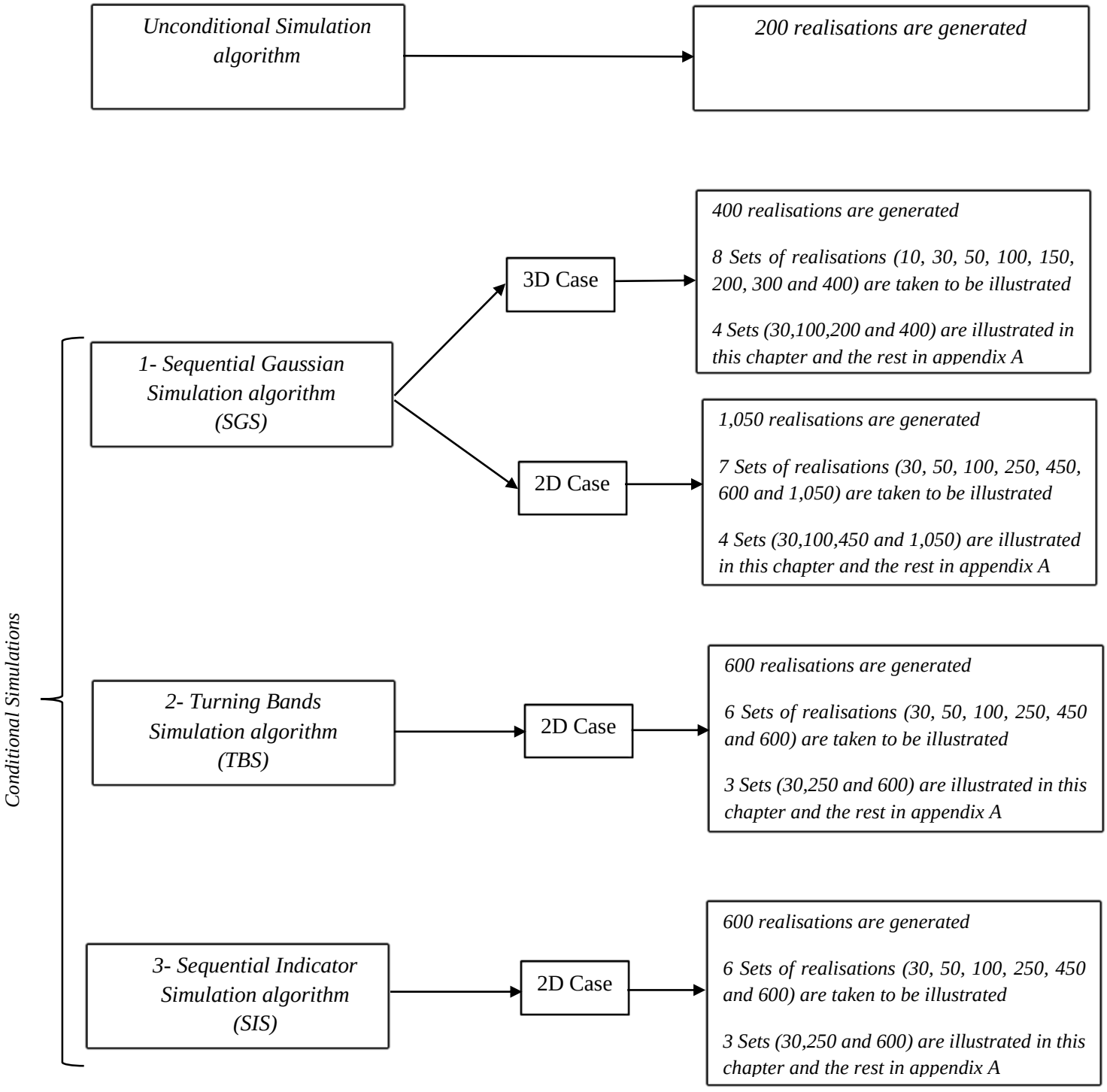


Figure 7-1: The simulation algorithms, data sets and number of generated realisations discussed in this chapter. As can be seen, all three conditional simulation algorithms (SGS, TBS and SIS) are applied on the 2D case; therefore, the comparison between the results can basically reveal any possible differences between the simulation algorithms

Using this technique, in this chapter, we address the following topics: how dissimilar the generated realisations can be to each other; the impact of changing the geostatistical parameters on the space of uncertainty; how far the realisations are from local accuracy (Kriging model); and also controversial issues in geostatistical simulation, such as equal-probability of generated realisations and likelihood.

As mentioned earlier, stochastic simulation algorithms create realisations with identical histograms and covariance matrices; however, these realisations are significantly different in ways that cannot be captured by descriptive geostatistics. To quantify this spatial dissimilarity, in previous chapters we presented the Kantorovich distance as an important and intuitive measure of dissimilarity. Now, we apply the proposed methodology to detect and identify the structural relationships between realisations and visualise the space of uncertainty using two different datasets (see chapter 4).

Moreover, as explained in chapter 3, the simulation algorithms, such as sequential Gaussian simulation (SGS), turning bands simulation (TBS) and sequential indicator simulation (SIS), use completely different techniques and random function (RF) models to generate realisations (the user has to select one of those). In this chapter, by mapping the space of uncertainty made by these algorithms, we compare these three common stochastic simulation algorithms to find out possible differences that they may have.

This chapter contains two different simulation methods (unconditional and conditional) that were selected to illustrate the above mentioned proposals. First, we illustrate some of the basic properties and advantages of our dissimilarity quantification and subsampling approach on unconditional simulations; and then we turn to a realistic sets of conditional simulations, which is the main part of this chapter.

We illustrate and implement the method using two completely different data sets, 2D and 3D, as were described in chapter 4. Figure 7-1 is a snapshot of this chapter and illustrates the simulation algorithms, datasets and number of generated realisations, which are studied here.

7-2 Unconditional simulation

As it was explained in chapter 3, the spatial variability of a measurable geological parameter can be modelled by variograms. A variogram is described by nugget, sill and range and can be used in the estimation of any sort of variogram-based simulations.

The range of a variogram presents the structural part of the variogram model where a higher range shows better mineralisation continuity and also confirms a higher spatial correlation between sample data. We use this parameter in order to better present the impact of changing the geostatistical parameters, that is, the variogram's parameters on the space of uncertainty.

We start with unconditional Gaussian realisations as making variogram models in this type of simulation is easier. We had already generated three series of 3D-realisations with parameters shown in Table 4-4. As mentioned in chapter 4, for each set of realisations the variograms are isotropic spherical, the search neighbourhood is spherical and there are no nugget effects. The block models have a block size of $25 \times 25 \times 12.5$ m (x, y and z direction). Each of the block models contain 2,805 blocks.

Note that the number of realisations in the second series of unconditional simulations is 200, but only 100 of them are used for any comparison between the series of generated realisations. 200 realisations were used simply to show the impact of the number of subsamplings on the relative accuracy (not for comparison), as more realisations were needed to show this impact.

7-2.1 Variation in approximation accuracy with S

Having computed the distance array D_{rs} we solve the realisation reduction problem (6.21)-(6.28) to determine the subset of S weighted realisations that are the best representatives of 200 realisations. In order to know how large S should be and how well an increasing number of optimum points can represent the full collection of realisations, we solve (6.21)-(6.28) for different values of S ranging from 1 to 200. In practice, the number S may be selected based on time or other resource costs.

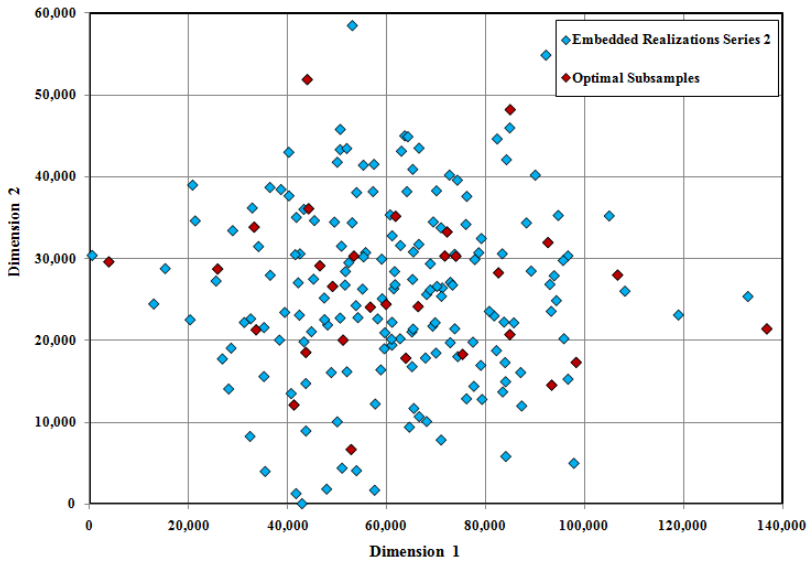


Figure 7-2: Multidimensional scaling embedding of 200 realisations (series 2 of Table 4-4) in R^2 (blue diamonds) and the 30 realisations that form the best subsample of size 30 (red diamonds)

As we mentioned in chapter 6, we report z_S relative to z_1, the latter representing the 'base' distance between the best deterministic approximation of the collection of S realisations, subsequently calculating the relative accuracy of the optimal S-subcollection (I_S). The result of these calculations for thirty different values of S is presented in Figure 7-3.

Figure 7-3 shows, for example, that using just $S = 30$ realisations (15% of all 200 realisations), a relative accuracy of around 52% may be obtained. The red diamonds in Figure 7-2 give a visual representation via multidimensional scaling of the relative position in R^2 of the optimal 30 realisations. Note that the selected (red) points are distributed throughout the entire set of (blue) points to well sample the point of distribution. The relative weights assigned to the red points are not shown in Figure 7-2; but, typically, those points at the periphery of the point set have the lowest possible nonzero weight of $1/\boldsymbol{S} = 1/200$. Points closer to the centre will have a weight above the average of $1/S = 1/30$. We will demonstrate these effects in a real case study in the next section.

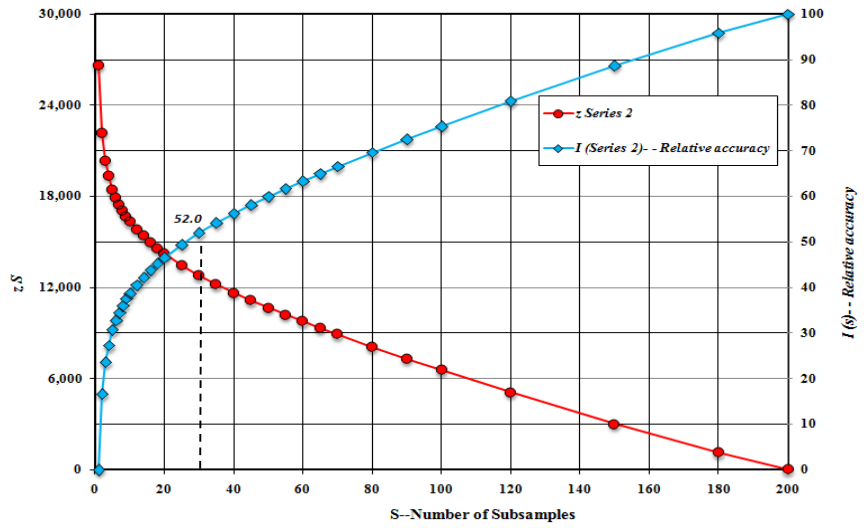


Figure 7-3: Distance z_S between the S-subcollection and the full collection of $\boldsymbol{S}$ simulations (red) and the relative accuracy (blue)

7-2.2 The impact of spatial continuity on the space of uncertainty

We use two additional series (see series 1 and 3 in Table 4-4) to investigate the effects of spatial continuity (variogram range) on the ability to subsample. Repeating the computations used for Figure 7-2 for series 1, and 3 we obtain the results shown in Figure 7-4.

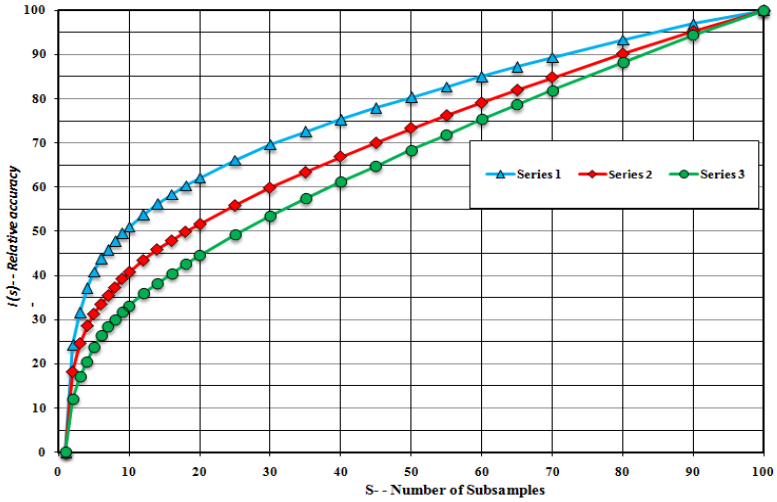


Figure 7-4: Illustration of the impact of spatial continuity (range) on the relative accuracy (I_S) of optimal S-subsamples

From Figure 7-4, we can see that for a fixed S, we can achieve a greater relative accuracy for series 1, which has a larger range. The decreasing range leads to a reduction in relative accuracy. These results are exactly what one would expect; consequently, longer ranges of correlations (with greater spatial continuity) will tend to make the collection of $\boldsymbol{S}$ realisations more structured and less random. This greater structure, encoded via the distances D_{rs}, can be exploited by our optimal subsampling procedure. In other words, simulations with a smaller range will require a larger number of subsamples to achieve the same relative accuracy.

7-2.3 Optimal vs. random subsampling

Figure 7-5 clearly demonstrates the distances z_S that can be achieved by optimally subsampling rather than randomly sampling. The variability of random sampling can also be very high. For example, suppose that rather than generating 100 realisations and sampling the best 5, one merely generated the first 5 realisations of the 100 and then stopped. The variability in distance z_S of the first 5 ranges is from 23,128 to 32,585. Thus, we argue that it is worth investing more time to create a larger number of realisations to attempt to better represent the true resource uncertainty and to then subsample as best possible from that larger collection of realisations. Indeed, in this example, the optimal 5 realisations would have a better relative accuracy than some of the collections of 20 random realisations.

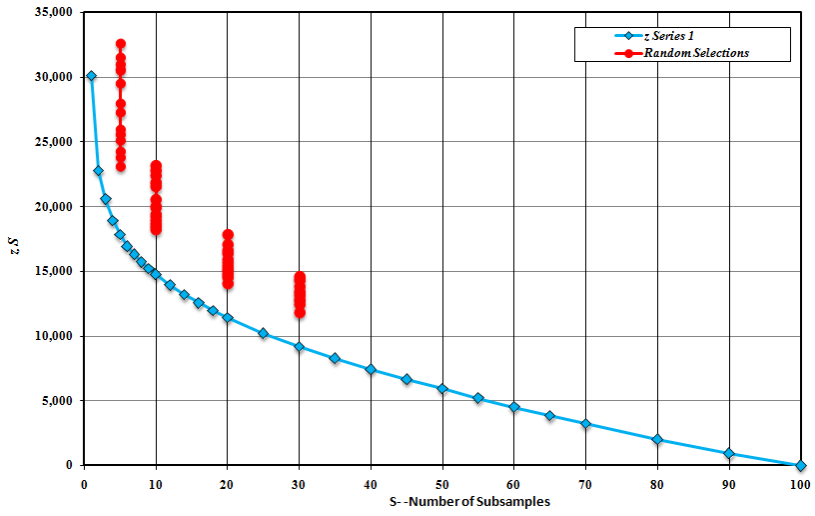


Figure 7-5: Comparison of distances z_S between the optimal subsamples and randomly selected subsamples

7-3 Conditional simulation- SGS algorithm and 3D case

Having illustrated some of the basic properties and advantages of our dissimilarity quantification and subsampling approach on unconditional simulations, we now turn to a realistic set of conditional simulations of a porphyry copper deposit. We try to explain the properties or attributes of the space of uncertainty through conditional simulation algorithms as these types of simulation are common in the geosciences field.

As was explained in chapter 4, three types of block models were estimated by using a block size $25 \times 25 \times 12.5\ m$ ($x, y\ and\ z$ direction). The first type is estimated by Kriging; the second type by generating $S = 400$ sequential Gaussian conditional simulations; and the third type by averaging all 400 simulated models (we call this E-type). Each of these block models contains 2,805 blocks that cover the entire supergene zone.

We now compute the distances D_{rs}, $1 \leq r < s \leq 402$ using (6.3)-(6.6) where, to the 400 realisations, we add the Kriged and E-type block models also computing distances between these models. Interestingly, we note that because of the smoothing effects, the Kriging and E-type models are the closest pair of models, that is $\min_{1 \leq r,s \leq 402} D_{rs} = D_{r^*s^*}$ where $r^* =$ Kriging and $s^* =$ E-type.

7-3.1 Impact of the number of realisations on the space of uncertainty

The characterisation of the space of uncertainty is rendered difficult by the fact that only a limited number of realisations are usually generated. A frequent and still open question relates to the number of realisations needed to characterise this space (Gooveart 1999). In chapter 4, we mentioned a few case studies which addressed the impact of the number of realisations on the space of uncertainty by applying a transfer function. We believe this method loses a significant amount of the underlying structure of the space of uncertainty produced by simulation algorithms.

As was mentioned in chapter 6, for a set of S generated realisations, the dissimilarity distance matrix $S \times S$ can be constructed by calculating the pairwise distances between S generated realisations and this matrix reveals the underlying structure of variability between realisations. The dissimilarity distance matrix consists of a set of points (S) with distances between them (interpoint distance); thus, any property of the matrix needs to be addressed with interpoint distances.

For evaluating the impact of the number of realisations on the space of uncertainty, 40 sets of generated realisations, namely $10,20,\cdots,400$ are taken by stepping 10 increment realisations from 10 to 400, where each set contains all the realisations of the previous set. For each set, the following parameters of its dissimilarity distance matrix are calculated to assess the structure of the space of uncertainty. The first parameters are the probability of distribution function (PDF), Mean, variance and the minimum and maximum of the interpoint distances. This is followed by the average distance of generated realisations from the Kriging model. The Kriging model is used as a fixed point inside the space of uncertainty to constitute a base for any sort of comparison. That is, we check for any possible changes in the location of the other points in respect to this model, in the space of uncertainty, by increasing the number of realisations. The last parameter is the relative accuracy (I_S) of the optimal S-subcollection; this parameter is able to reveal how representative points may be changed by increasing the number of realisations.

Furthermore, for better viewing of these parameters and in order to demonstrate comparative evaluation on the space of uncertainty in detail, 8 sets of realisations (10, 30, 50, 100, 150, 200, 300 and 400) are illustrated. However, for the sake of brevity, only 4 sets with the following numbers of realisations (30, 100, 200, and 400) are presented here; the rest of the sets have been included in appendix A. We employ the MDS technique for visualising (in R^2) the level of similarity of generated realisations for which the inter-point distances points in R^2 approximate the distances D_{rs}. These graphs can show the points of distribution in R^2 that may help to find whether the points are clustered or randomly distributed in the space. Moreover, the approximate location of the Kriging model in the space can be visualised. It should be mentioned here that the Kriging model is always at the centre of the space, which mathematically means that for $z_1 = d(\mu_{1,N}, \mathcal{V}_N),\ z_1 = Kriging\ model$. That means that simulated models have a systematic tendency to configure the space of uncertainty in such a way that smoothing models always have the minimum distance to all realisations. Nevertheless, this may not be clearly shown in some figures because of the approximated distances.

First set - Figure 7-6 shows the result of the interpoint distance calculation for 30 realisations and the Kriging model. As can be seen, the best fitted distribution is lognormal (positively skewed). As it will be shown later, the lognormal distribution[1] can be the best option to describe the interpoint distance distribution

[1] The normal and lognormal distributions are closely related in such way that if X is distributed lognormally with parameters μ and σ, then $log(X)$ is distributed normally with mean μ and standard deviation σ. The density function of lognormal distribution is $f(x) = \frac{1}{x.\sigma\sqrt{2\pi}} \exp(-\frac{1}{2\sigma^2}(\log(x) - \mu)^2)$

in the space of uncertainty. Figure 7-7 shows multidimensional scaling embedding the realisations. As can be seen, the Kriging point is at the minimum distance from the others.

The parameters of the histogram are shown in Table 7-1. The maximum interpoint distance in the dissimilarity distance matrix may be used to describe the size of the space of uncertainty. This parameter is assessed.

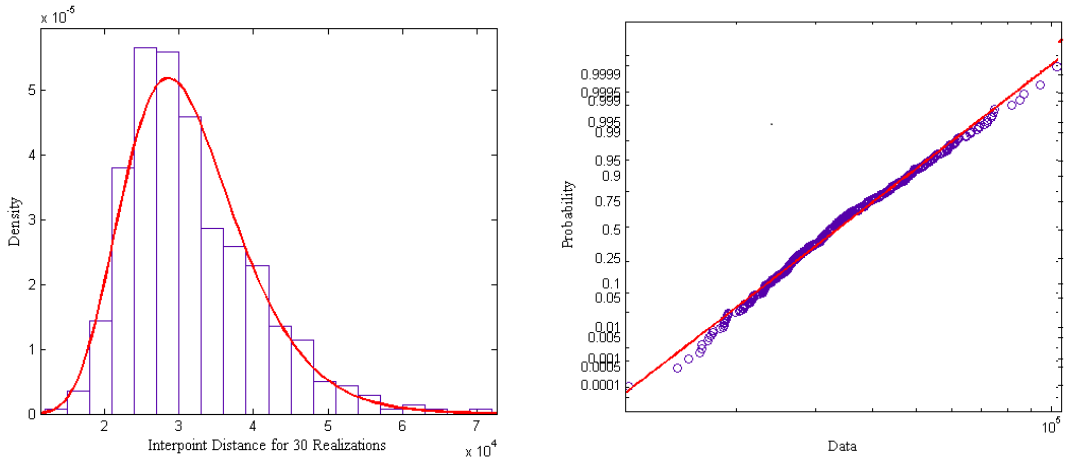


Figure 7-6: The left graph is the histogram of 465 interpoint distances of 30 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them (SGS and 3D case)

Table 7-1:Statistical parameters of the histogram of the interpoint distances for 30 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
465	31,654.14	8,649.40	74,812,131	13,518.60	70,311.60	56,793	1.05	24,723.1

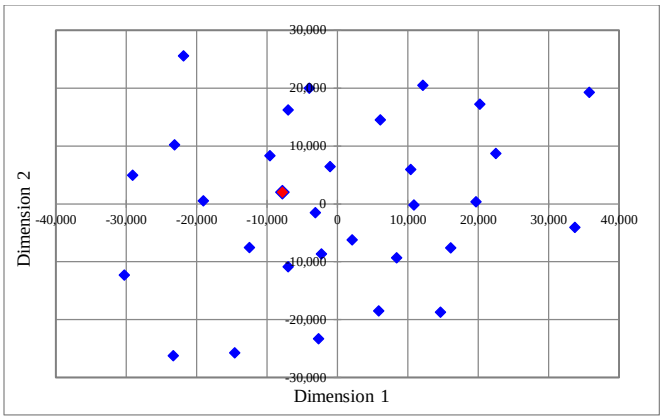


Figure 7-7: Multidimensional scaling embedding of 30 realisations (blue diamonds) and the Kriging model (red diamonds) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

Second set - Figure 7-8 shows the result of the interpoint distance calculations for 100 realisations and the Kriging model; the best distribution that can be fitted on the distribution is still lognormal. The parameters of the histogram are shown in Table 7-2, where the maximum distance (between realisations) is much bigger than the first set. That means that the space is still getting bigger. Figure 7-9 shows multidimensional scaling embedding the realisations. As can be seen, the Kriging point is at the minimum distance from the others.

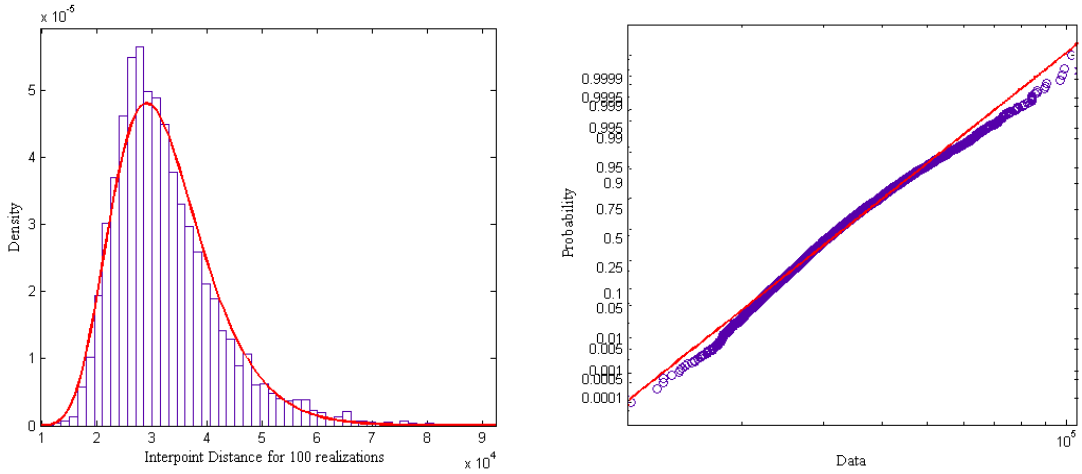


Figure 7-8: The left graph is the histogram of 5,050 interpoint distances of 100 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution which shows a reasonable fitting between them (SGS and 3D case)

Table 7-2: Statistical parameters of the histogram of the interpoint distances of 100 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
5,050	32,551.20	9,662.09	93,355,956	11,582.30	91,022.5	79,440.2	1.35	23,130.0

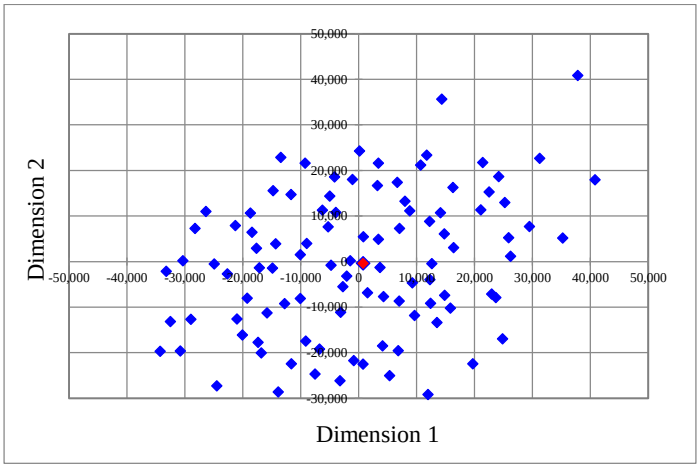


Figure 7-9: Multidimensional scaling embedding of 100 realisations (blue diamonds) and the Kriging model (red diamonds) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

Third set - Figure 7-10 shows the result of the interpoint distance calculations for 200 realisations and the Kriging model; the best fitted distribution is still lognormal. Although the extent of the space has become slightly bigger than the previous one (see Table 7-3), the concentration of points is getting higher than what it was before, as can be seen in Figure 7-11.

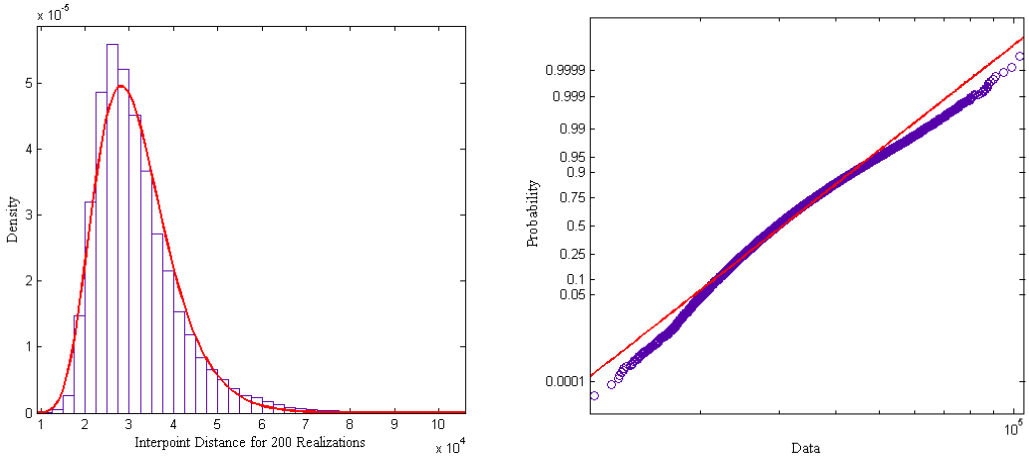


Figure 7-10: The left graph is the histogram of 20,100 interpoint distances of 200 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution showing a reasonable fitting between them (SGS and 3D case)

Table 7-3: Statistical parameters of the histogram of the interpoint distances of 200 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
20,100	31,619.55	9,467.71	89,637,536	11,582.30	103,194.00	91,611.70	1.45	22,349.3

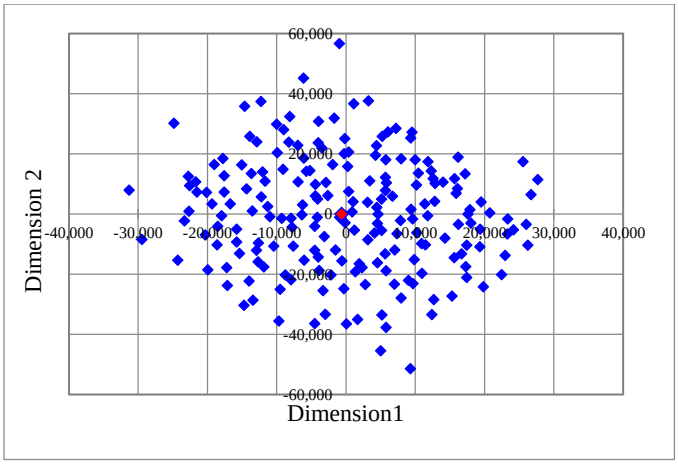


Figure 7-11: Multidimensional scaling embedding of 200 realisations (blue diamonds) and the Kriging model (red diamonds) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

Forth set - Figure 7-12 shows the result of the interpoint distance calculations for 400 realisations and the Kriging model; the best fitted distribution is still lognormal. The extent of the space is the same as in the previous set (see Table 7-4), thus the concentration of points is getting consistently higher and the space has become dense (see Figure 7-13).

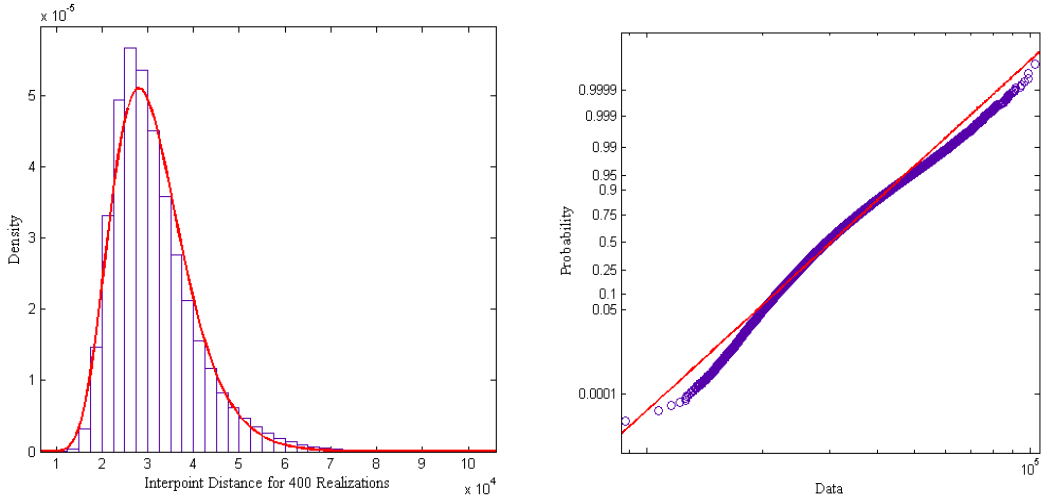


Figure 7-12: The left graph is the histogram of 80,601 interpoint distances of 400 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them (SGS and 3D case)

Table 7-4: Statistical parameters of the histogram of the interpoint distances of 400 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
80,601	31,313.38	9,089.90	82,626,268	11,582.3	103,194	91,611.70	1.35	22,161.0

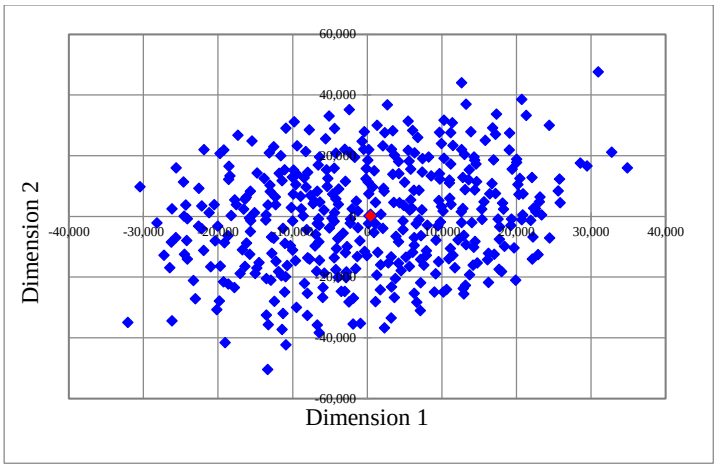


Figure 7-13: Multidimensional scaling embedding of 400 realisations (blue diamonds) and the Kriging model (red diamonds) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

7-3.2 The impact of the number of realisations on size and density of the space of uncertainty (SGS-3D Case)

To evaluate the impact of the number of realisations on size and density of the space of uncertainty, 40 sets of generated realisations, namely $10,20,\cdots,400$ are taken by stepping 10 increment realisations from 10 to 400, each set containing all the realisations of the previous one. For each set the following parameters are calculated: the average and the standard deviation of the interpoint and the minimum and maximum distances.

Figure 7-14 (blue curve) shows the variation of the average of interpoint distances versus the number of realisations. As can be seen, the average distance decreases when the first realisations are generated .In contrast, after 70 realisations the average distance increases, though not significantly, and then reduces to what it was before. Beyond the 150th realisation, the average distance remains constant and shows very insignificant fluctuations around the 31,650.

The red curve in Figure 7-14 shows the variation of the standard deviation of interpoint distances versus the number of realisations. As is illustrated, the standard deviation decreases (same as the average curve) when the first realisations are generated however, after 70 realisations it increases up to a maximum of 9,707. Subsequently, the rate of increase and decrease is reduced until it stabilises in a flat curve; that is, the standard deviation fluctuates to the extent that its rate reduces dramatically by increasing the number of simulations, remaining stable at around 9,100.

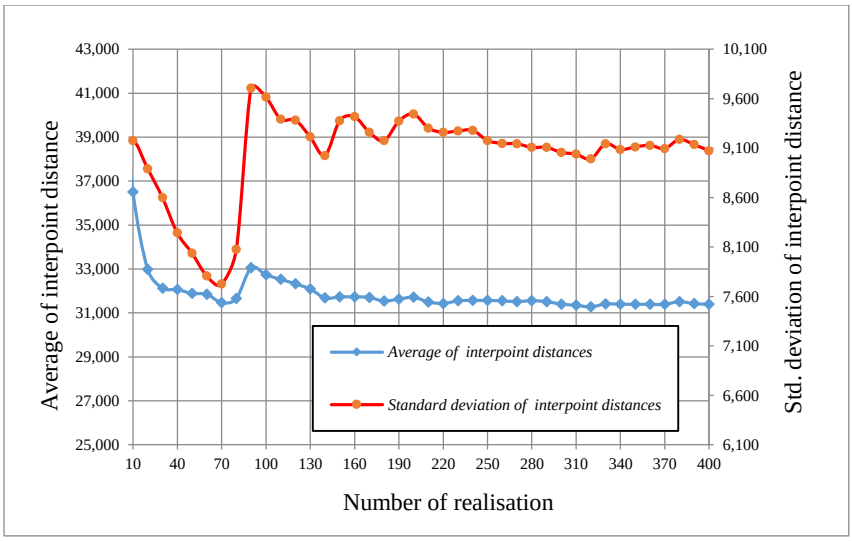


Figure 7-14: The variation of the average (blue curve) and standard deviation (red curve) of interpoint distances versus the number of realisations. Both of them are stabilised by increasing the number of realisations at around 31,650 and 9,100, respectively (SGS and 3D case)

As illustrated earlier, the interpoint distances of histograms in the space of uncertainty follow lognormal distributions, while the mean and standard deviations of the interpoint distances are stabilised by increasing the number of realisations. Thus, it can be concluded with confidence that the simulation algorithm generates retaliations in such a way that the dissimilarity between them (interpoint distances), or precisely the structure of the space of uncertainty, ultimately follows the lognormal distribution below.

$$f(x) = \frac{1}{101,230\sqrt{2\pi}\ \times x} \exp\left(-\frac{1}{2 \times 101,230^2}(log(x) - 13,785)^2\right) \qquad (7.1)$$

In the next section, we will discuss whether this lognormarlity is limited to this case (SGS) or can be valid for other cases and other simulation algorithms.

After finding the interpoint distance distribution, we can now turn to determine minimum and maximum possible dissimilarities in the space of uncertainty and answer the question whether increasing the number of realisations can make any difference to the minimum and maximum interpoint distances. The maximum interpoint distance can be called the size (how big) of the space of uncertainty and the minimum can describe how close the generated realisations can be in the space.

In this case, the size of the space of uncertainty increases with the number of realisations in the three steps (see Figure 7-15) and then maintains the same level after the 150th realisation. It has to be mentioned here that the maximum distance comes from a large dissimilarity between two realisations, thus stabilising the extent of the space of uncertainty even after doubling the number of realisations confirms the tendency of the conditional simulation to generate and maintain more or less similar realisations (in the limited range) instead of quite dissimilar ones.

The same condition is presented for the minimum distance. The minimum distance is approximately stabilised after generating the 50th realisation. That is, the simulation algorithm cannot generate realisations which are very similar to each other. Thus, there is an area (neighbourhood radius $r \leq D_{Min}$) around each retaliation where there are no realisations.

The main reason is that the conditioning data in all generated retaliations, namely sampled points, as explained in chapter 3, have to honour sampled points and, thus, this condition does not let the simulation algorithm create a big or a very small space of uncertainty. By decreasing the sampled points in the original data set, we may expect an expansion of the spaces of uncertainty and larger differences between realisations.

The realisations which contain larger differences from the other ones are usually located on the edges of the space of uncertainty and, consequently, are far from the Kriging model. These realisations can be classified as extreme points in the simulation process.

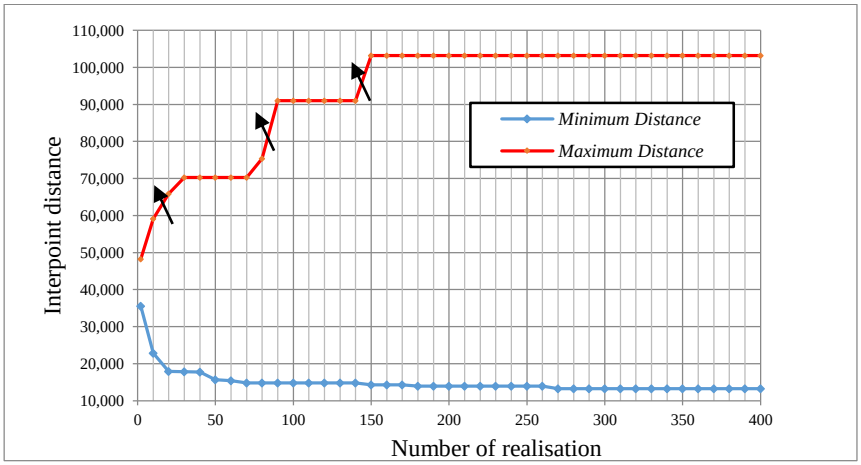


Figure 7-15: Impact of the number of realisations on the maximum and minimum distances between the realisations. As can be seen, both the maximum and minimum distances are stabilised by increasing the number of realisations (SGS and 3D case)

As it has been clearly shown before, in the multidimensional scaling graphs, by increasing the number of realisations the density of the points rises inside the space as a result of not extending the space.

The density of the points can be calculated or compared by using the following two methods. *Method 1* includes the number of points that fall within a fixed area or volume (cell or block). This method cannot apply metric spaces as area or volume and does not have any meaning in these spaces. However, by applying multidimensional scaling embedding of the points into R^2, we are able to reveal and compare the point density inside the space of uncertainty (see Figures 7-16, 7-17, 7-18 and 7-19). As this embedding method is not precisely accurate, we only apply it for visualisation purposes in this case (3D) and to make sense of the point of distribution inside the space of uncertainty.

Method 2 is the number of points that fall within the fixed search radius (neighbourhood radius[2]), for instance, the number of points are in a neighbourhood radius r_i of a fixed point x_i (x_i can be a Kriging or E-type model). This method can be applied on metric spaces (see Figures 7-20 and 7-21). As this method seems to be accurate, it can be applied for all cases (2D and 3D).

Figures 7-16 and 7-17 show the point density of multidimensional scaling embedding of 100 realisations in 2D and 3D graphs, respectively. As illustrated, the points are approximately distributed uniformly in the space and there are few cells where the number of realisations is higher than 2.

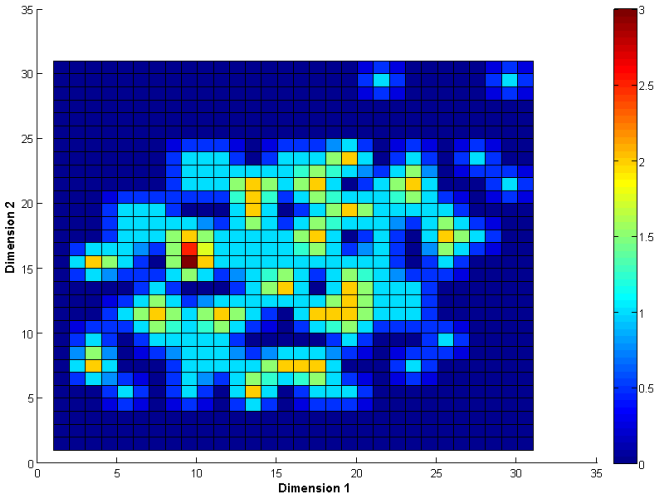


Figure 7-16: The point density of multidimensional scaling embedding of 100 realisations in R^2 the points are approximately distributed uniformly in the space (SGS and 3D case)

[2] See chapter 5.

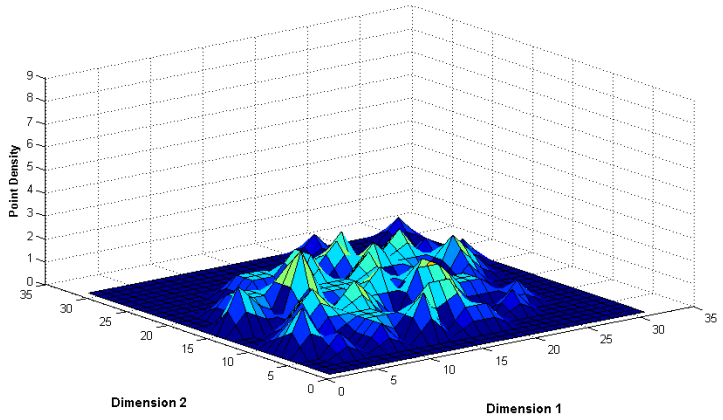


Figure 7-17: Point density of multidimensional scaling embedding of 100 realisations in R^3 . Point density is shown as a continuous surface to better represent the distribution of data. The points are approximately distributed uniformly in the space (SGS and 3D case)

Figures 7-18 and 7-19 show the point density of multidimensional scaling embedding of 400 realisations in 2D and 3D graphs, respectively. The points are distributed in such a way that density in the centre of the space is much higher than the edges. We know that the Kriging or E-type model is located at the centre of the space (see Figure 7-13).

We know that condition simulation does not have any systematic bias; therefore, all generated realisations are equally probable and fairly represent the entire uncertainty space. However, it can be seen here that there is a larger probability to generate realisations in the neighbourhood of the Kriging or E-type model rather than the edge of the space. This means that conditional simulation (in this case) has a systematic tendency to configure the space of uncertainty in such a way that the probability to find or generate the realisations decreases from the centre to the edge of the space. This trend occurs in all directions but it does not seem to have a perfect symmetric shape.

In the majority of geoscience applications only a few realisations can be chosen. If this selection is based on a random selection from all possible outcomes, the set of chosen realisations is not fairly sampled and the realisation around the Kriging or E-type models have a higher chance to be collected.

The main purpose of using the geostatistical simulation method is to generate the realisations which are equally probable. Therefore, the chance of being selected inside the space of uncertainty has to be equal for all retaliations. But as can be seen in Figures 7-18 and 7-19, those graphs do not confirm that the chance of being selected is equal for all generated realisation.

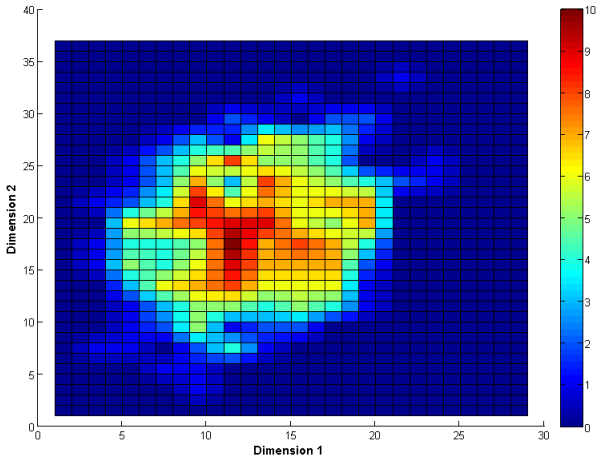


Figure 7-18: The point density of multidimensional scaling embedding of 400 realisations in R^2. The points are distributed in such way that the density in the centre of the space is much higher than on the edges. The number of realisations in the centre cells is higher than the marginal ones (SGS and 3D case)

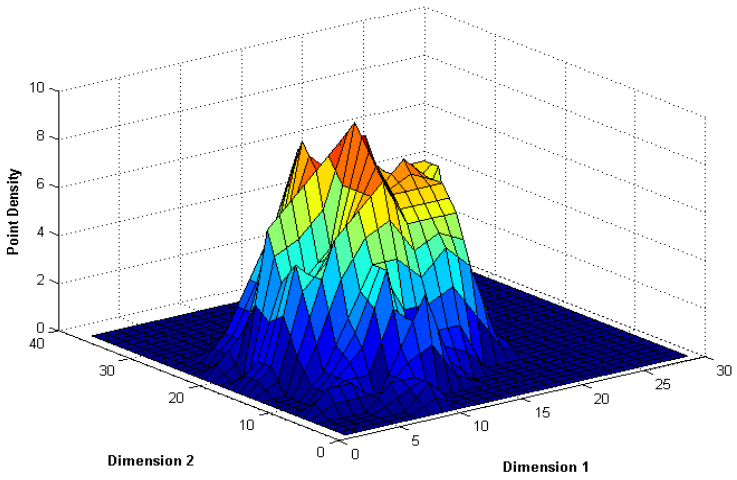


Figure 7-19: The point density of multidimensional scaling embedding of 400 realisations in R^3. The point density is shown as a continuous surface to better represent the distribution of data. The points seem to be approximately distributed into a bell shape although it is not symmetric (SGS and 3D case)

Using this approach to find the density of points inside the pace would be valid if only the approximation distances, which are made by MDS, were acceptable in the point of Stress factor (for more details see chapter 5). There are different standards regarding the amount of Stress to tolerate. The procedure we could find in the literature considers that the stress under 0.1 is excellent while any value over 0.15 is unacceptable. If the stress factor is high, instead of going through the approximated distance, we can use the neighbourhood radius to evaluate the density of points around any desired point.

Figure 7-20 shows the number of realisations that fall within the neighbourhood radius $r \leq a,\ 0 < a \leq Max\ D_{rs}$ for the Kriging model and the four selected realisations which are sorted based on the distance from the Kriging model. That is, realisations no.31 and no.350 are closest and farthest to the Kriging model, respectively. As can be seen, the number of realisations that fall within a fixed neighbourhood radius r for the Kriging model are much higher than the others. For example, 160 realisations are in neighbourhood radius $r \leq 20{,}000$ for the Kriging model, while they are less than 80 for realisation no.31 and less than 3 for realisation no.350.

Furthermore, for the Kriging model, all realisations (all space) fall within neighbourhood radius $r \leq 55{,}600$, while for realisations no.31 and no.350 the neighbourhood radius has to be more than 58,000 and 80,000, respectively.

Moreover, Figure 7-21 shows the histogram (red) of the distances of 400 realisations from the Kriging model. The histogram (blue) is the distance of 400 realisations from realisation no.350 (extreme point), all confirming that the number of realisations in the vicinity of the Kriging model are considerably higher than in the other parts of the space. In addition, by moving away from the Kriging model, the number of realisations (for the same neighbourhood radius) decreases. The result is exactly the same as what had already been illustrated through the multidimensional scaling, but this approach is more reliable.

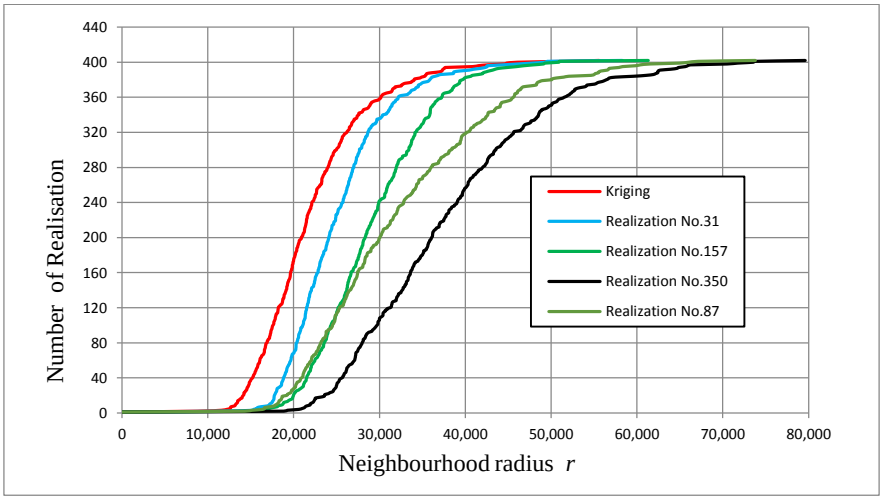


Figure 7-20: The number of realisations that fall within the neighbourhood radius r for the Kriging model and the four selected realisations which are sorted, based on the distance from the Kriging model (SGS and 3D case)

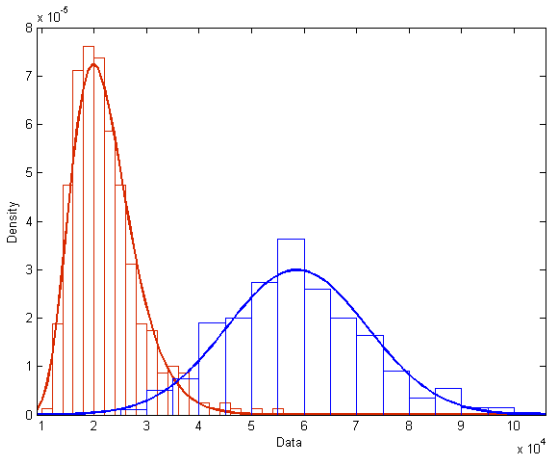


Figure 7-21: The red histogram is the distance of 400 realisations from the Kriging model with the best fitted lognormal distribution (red line). The blue histogram is the distance of 400 realisations from realisation no.350 (extreme point) with the best fitted lognormal distribution (blue line) (SGS and 3D case)

As the Kriging model is a fixed point in the space, we check any possible changes in the location of the other points with respect to this model (in the space of uncertainty) by increasing the number of realisations. Figure 7-22 shows the variation of the average of distance and distances variance from the Kriging versus the number of realisations. As can be seen, both of them are stabilised by increasing the number of realisations, although they fluctuate highly before the 50th realisation. Thus, generating more realisations does not make a significant change in the structure of the space of uncertainty and the distance histogram from a fixed point (Kriging) would remain unchanged.

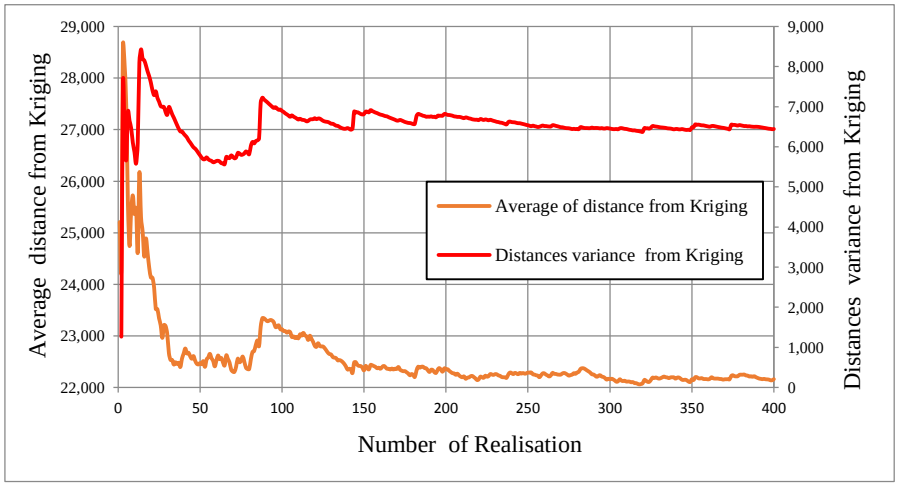


Figure 7-22: The variation of the average of distance and distances variance from the Kriging versus the number of realisations. Both of them are stabilised by increasing the number of realisations (SGS and 3D case)

7-3.3 Impact of the number of realisations on relative accuracy (SGS and 3D case)

Before evaluating the impact of the number of realisations on relative accuracy, we first determine the subset of S weighted realisations that are the best representatives of 400 realisations.

We compute the distances D_{rs}, $1 \leq r < s \leq 402$ using (6.3)-(6.6) (to the 400 realisations, we add the Kriging and E-type block models and compute distances between these models as well). Interestingly, we note that because of the smoothing effects, the Kriging and E-type models are the closest pair of models, that is $\min_{1 \leq r,s \leq 402} D_{r,s} = D_{r^* s^*}$ where r^* = Kriging and s^* =E-type. We then solve (6.21)-(6.28) with $S = 30$, which we find produces a relative accuracy of 25.64% (Figure 7-23).

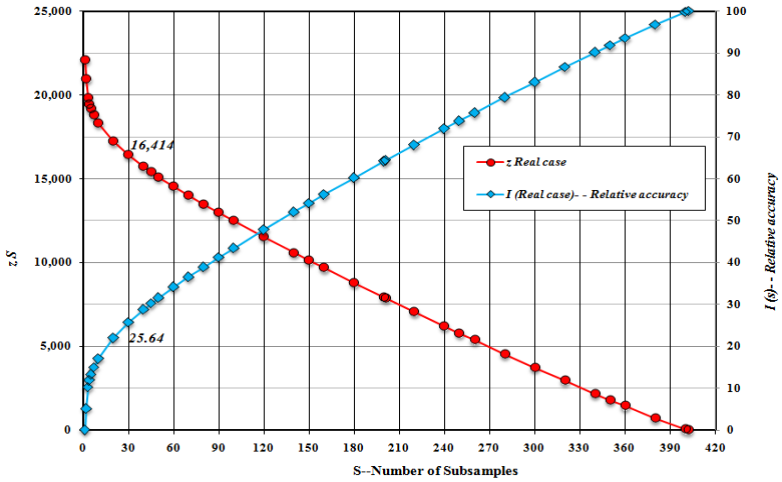


Figure 7-23: Distance z_s between the S-subsample and the full sample of $\boldsymbol{S}$ realisations (red) and relative accuracy (blue) (SGS and 3D case)

To visualise in an approximate way the relative positions of the realisations, we employ the MDS technique to produce points in R^2 (see Figure 7-23) for which the inter point distances approximate the distances D_{rs}. Figure 7-24 shows the MDS embedding of 400 simulations, the E-type and the Kriging models with the optimal weights assigned to each of the 30 subsamples shown as ball heights. The Kriging and E-type models are shown on the graph.

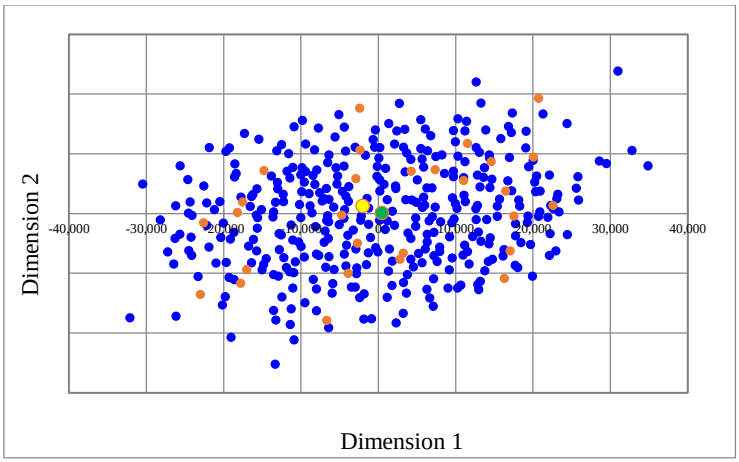


Figure 7-24: Multidimensional scaling embedding of 400 realisations (blue balls), the E-type model, (yellow ball) and the Kriging (green ball) in R^2 and the 30 realisations that form the best subsample of size 30 (red balls) (SGS and 3D case)

These peripheral selections are effectively chosen to represent themselves only, while the selected realisations nearer to the centre of the figure represent not only themselves, but those realisations nearby, effectively absorbing the weights from these nearby realisations. This unequal weight distribution is another advantage that our approach has over simply taking the first S realisations computed (we remark that if one requires $w_s = 1/S$ in (6)-(13), the optimisation model (6)-(13) is easily modifiable to achieve this).

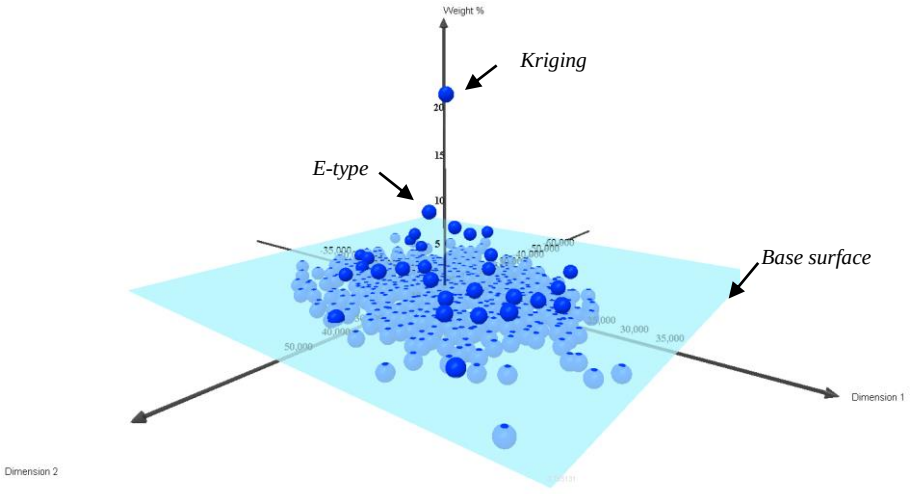


Figure 7-25: Multidimensional scaling embedding of 400 realisations, the E-type model, and the Kriging model in R^3. The 30 subsamples that best represent the 402 points are shown above the base surface with the height of the 'ball' indicating the relative weight given to each of the 30 subsamples. The Kriging and E-type models are shown on the graph (SGS and 3D case)

Two issues are apparent from Figures 7-24 and 7-25 .Firstly, the distances D_{rs} for r = Kriging or r = E-type and $s = 1, \dots, \boldsymbol{S}$ are small on average when compared to the overall average of the distances D_{rs}, $1 \leq r < s \leq \boldsymbol{S}$. Formally, $(1/\boldsymbol{S}) \sum_{s=1}^{\boldsymbol{S}} D_{rs} = 23{,}606$ for r = Kriging and 22, 073 for r = E-type, while the overall average of distances is $\frac{1}{S(S-1)/2} \sum_{s=1}^{S} \sum_{r<s} D_{rs} = 31{,}313$.

Secondly, the weights given to the selected realisations tend to be higher nearer to the centre of the figure, while those selected realisations near the periphery tend to have lower weights (and often, the lowest possible nonzero weight of $(1/S = 1/30)$.

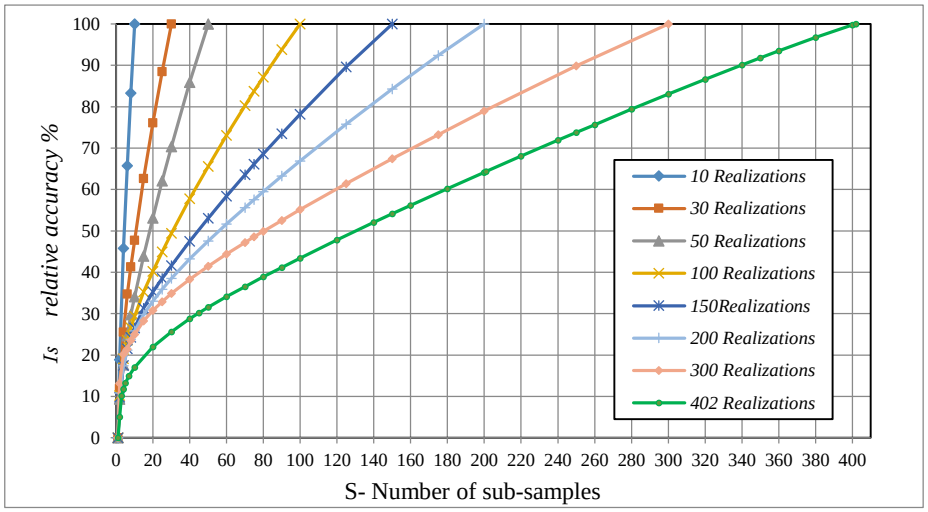


Figure 7-26: Impact of the number of realisations on relative accuracy, namely sub-sampling for the 8 sets with the following number of realisations 10, 30, 50, 100, 150, 200, 300 and 400 (SGS and 3D case)

More of the overall distribution of the realisations may be seen by computing a larger number $\boldsymbol{S}$; further, this information is able to be used to not only choose a good subsample, but also to allocate optimal weights to the subsample. We now compute what was done above for the rest of the collected sets, namely 10, 30, 50, 100, 150, 200, 300, and 402), to evaluate the impact of the number of realisations on relative accuracy (see Figure 7-26).

For a better comparison between the different sets, we compute the percentage of the number of sub-samples for each set to get same x axis. Figure 7-27 shows the impact of the percentage of the number of sub-samples $\left(\frac{S}{N}\%\right)$ on relative accuracy of the 8 sets. As can be seen, after 50 realisations (the third set) the differences between relative accuracy of the sets become insignificant. For instance, for 20% of the sub-

samples, the difference between the relative accuracy of 100 realisations (the fourth set) and 400 realisations (the eighth set) is less than 10 percent. This number would be around 5 percent for 40% of sub-samples.

The mean reason for these insignificant changes can be addressed by equation 6.31. As we explained in chapter 6, z_S/z_1 ratio represents the distance between the optimally subsampled S realisations and the full set of $\boldsymbol{S}$ realisations. If this ratio does not vary significantly by increasing the number of realisations (for a fixed amount of $\frac{S}{N}\%$) that means the distance between the best single realisation (z_1) and the set of collected S realisations does not change. That is, the structure of the space of uncertainty almost remains the same after generating a certain number of realisations.

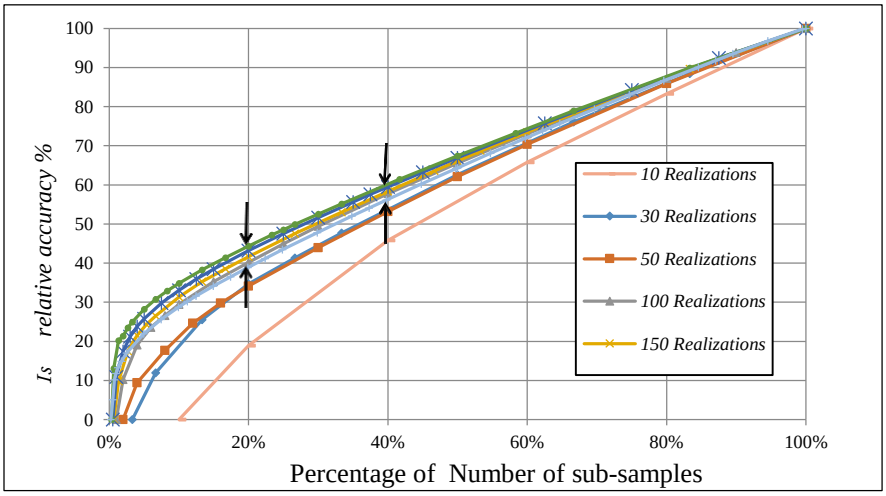


Figure 7-27: Impact of the percentage of the number of sub-samples on relative accuracy, namely sub-sampling for the 8 sets with the following number of realisations: 10, 30, 50, 100, 150, 200, 300 and 400. After 50 realisations (red Curve), the differences between the relative accuracy of the different set has become insignificant (SGS and 3D case)

7-4 Impact of the simulation algorithms and the number of realisations on the space of uncertainty (2D case)

In the previous sections, we explained some properties of the space of uncertainties with an example of the three dimensional (3D) data set and the SGS algorithm.

This section compares the space of uncertainty generated by three of the most commonly used algorithms: sequential Gaussian simulation (SGS); turning bands simulation (TBS); and sequential indicator simulation (SIS) by applying a different data set, namely Walker Lake data set (see chapter 4). As was explained in chapter 3, these algorithms use completely different techniques and random function (RF) models to generate realisations.

All three simulation algorithms are applied on the Walker Lake data set; therefore, the results of the comparison between their spaces of uncertainty can basically reveal any possible difference between these simulation algorithms.

The main reasons to use the Walker Lake data set as a 2D case study are as follows. First, we want to use at least two completely different data sets to conduct a better evaluation of the properties of the space of uncertainty. Other models and cases will likely show the same properties in case these two data sets in 2D and 3D share common properties. The second reason lies in the Walker Lake case being an open source data set which is well- known in the geostatistics field; therefore, all computations and mentioned results can be repeatable for those who may be interested in applying this method. Finally, the Walker Lake case has an exhaustive data set which is presented here as a real model. Thus, including the Kriging model, there would be two fixed points where the distance between them (in all types of simulation algorithms) always remains constant. That gives a better sense of the structure of the space of uncertainty (made by the algorithms) where we are meant to compare.

Similar to what has been presented in the previous section for the 3D case, we first generate different realisations using each of the simulation algorithms for the 2D data set; after that the characteristics of the space of uncertainty of each of them are calculated individually; and, finally, we compare these spaces to find any possible differences between them.

Furthermore, we calculate deterministic geological reserve estimations, produced by Kriging, and also the real model (exhaustive data set) into the space of uncertainty to reveal how dissimilar other realisations are to the estimated and the real model. In other words, we measure how close global accuracies (different realisations) are to the local accuracy (estimation) and the real model.

7-4.1 Evaluating the space of uncertainty generated by SGS algorithm[3] (2D Case)

In this example, 1,050 realisations are generated by SGS algorithms. This huge number of generated realisations would help to find any possible changes that may have occurred inside the space of uncertainty by increasing the number of realisations.

[3] In this section, as the numbers in Figures and Tables are in form of scientific format $a \times 10^6$ for sake of brevity they are shown without 10^6.

To evaluate the impact of the number of realisations on the space of uncertainty, 52 sets of generated realisations, namely $10,30,\cdots,1050$ are taken by stepping 20 increment realisations from 10 to 1050, each set containing all realisations of the previous set.

Furthermore, similar to the three dimensional (3D) data set, for better illustration of the impact of the number of realisations on the space of uncertainty, 7 sets with the following number of realisations 30, 50, 100, 250, 450, 600 and 1050 are taken from 52 sets to obtain a better view of these parameters and conduct a comparative evaluation on the space of uncertainty in detail (Kriging and also the real model are included in all sets). However, for the sake of brevity, only 4 sets with the following number of realisations 30, 100, 450, and 1050 are presented here; the rest of the sets are detailed in appendix A. Similar to the previous case, the parameters for each set are assessed (as explained for the 3D case).

First set -Figure 7-28 shows the histogram of the interpoint distance for 30 realisations. As can be seen, the best fitted distribution is lognormal. The parameters of the histogram are shown in Table 7-5.

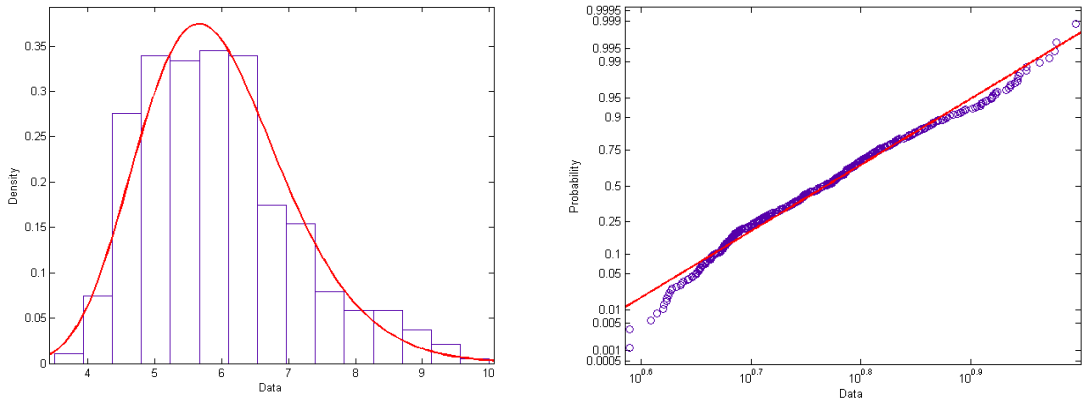


Figure 7-28: The left graph is the histogram of 435 interpoint distances of 30 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution which shows a reasonable fitting between them ($a \times 10^6$)-(SGS and 2D case)

Table 7-5: Statistical parameters of the histogram of the interpoint distances of 30 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
435	5.962	1.142	1.304	3.882	9.905	0.785	5.940	4.897

$(a \times 10^6)$

Figure 7-29 shows multidimensional scaling embedding of 30 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. That is, the real model is close to the edge of the space of uncertainty. This may not be a good signal for the simulation algorithm (SGS) to fail to generate close retaliations to the real model. Being at the edge of the pace of uncertainty means the model is quite dissimilar to the other models, thus it is called the extreme model.

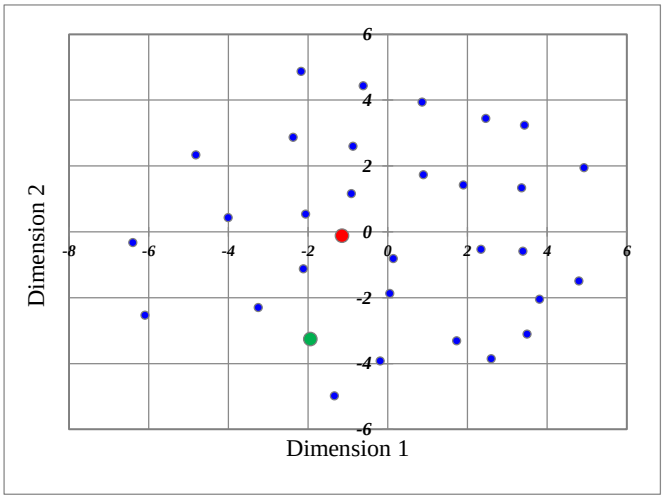


Figure 7-29: Multidimensional scaling embedding of 30 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations (SGS and 2D case)

Second set - Figure 7-30 shows the histogram of the interpoint distance for 100 realisations. As can be seen, the best fitted distribution is lognormal. The parameters of the histogram are shown in Table 7-6.

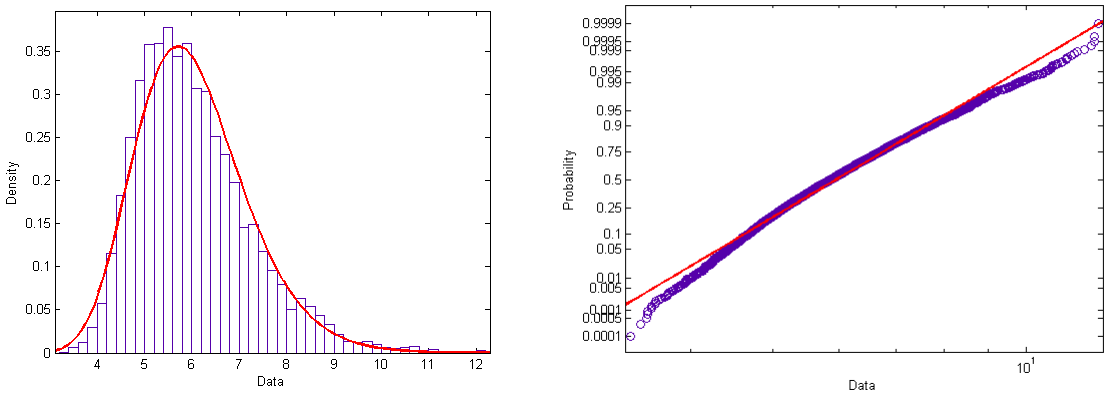


Figure 7-30: The graph on the left is the histogram of 4950 interpoint distances of 100 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them ($a \times 10^6$) - (SGS and 2D case)

Table 7-6: Statistical parameters of the histogram of the interpoint distances of 100 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
4,950	6.040	1.221	1.491	3.397	12.166	0.996	6.034	4.938

$(a \times 10^6)$

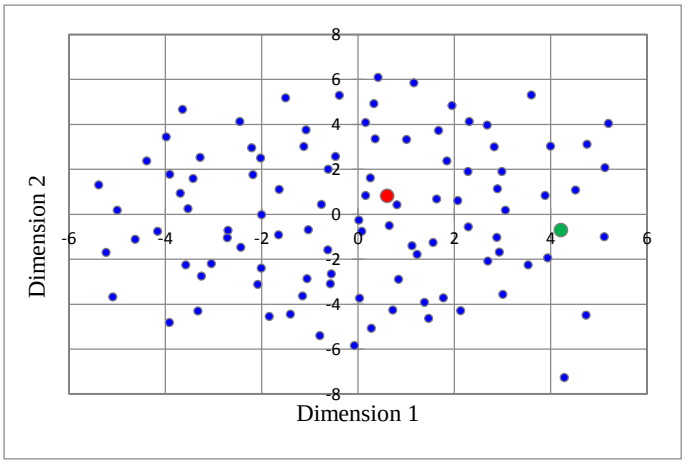


Figure 7-31: Multidimensional scaling embedding of 100 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and real model is quite far from the generated realisations (SGS and 2D case)

Figure 7-31 shows the multidimensional scaling embedding of 100 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

For better viewing and to achieve more accuracy in the embedded interpoint distances, the realisations are embedded into R^3 as well. Figure 7-32 shows the embedded points in 3D. As illustrated, the real model is close to the edge of the space and there are very few points around it, while the Kriging is in the centre of the space. The 3D Figure 7-32 is very similar to 2D Figure 7-30, but its Kruskal stress factor is 25% less than 2D, which may give a better view of the space of uncertainty.

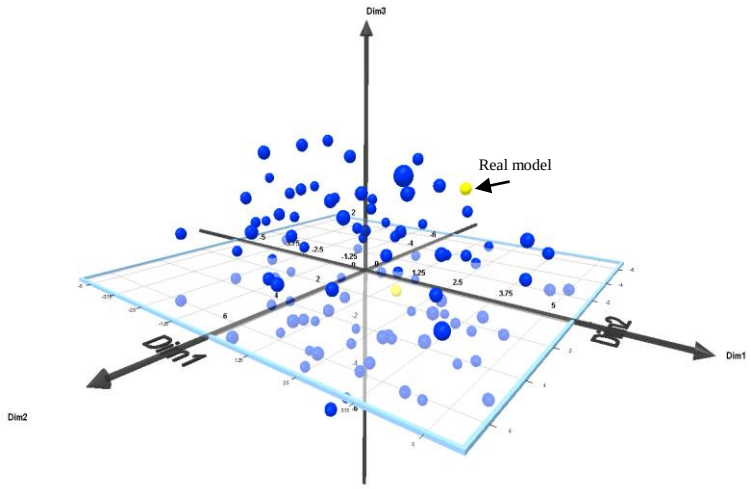


Figure 7-32: Multidimensional scaling embedding of 100 realisations (blue balls), the Kriging model and the real model (yellow balls) in R^3. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations. Embedding points in the 3D may give better accuracy and illustration than 2D (SGS and 2D case)

Third set - Figure 7-33 shows the histogram of the interpoint distance for 450 realisations. As can be seen, the best fitted distribution is lognormal. The parameters of the histogram are shown in Table 7-7.

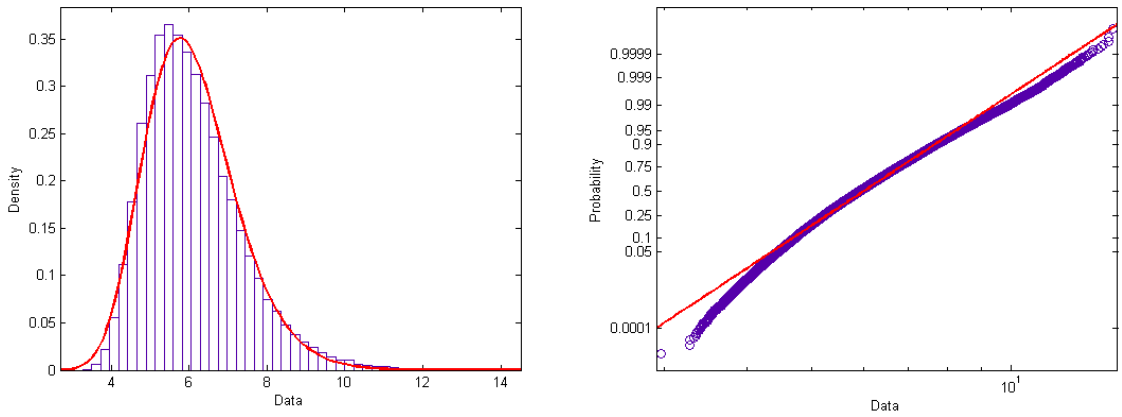


Figure 7-33: The graph on the left is the histogram of 101,025 interpoint distances of 450 realisations with the best fitted lognormal distribution (red line). The one on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, which shows a reasonable fitting between them (a $\times$ 10^6)- (SGS and 2D case)

Table 7-7: Statistical parameters of the histogram of the interpoint distances of 450 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
101,025	6.096	1.232	1.518	3.397	14.235	0.963	6.093	5.001

($a \times 10^6$)

Figure 7-34 shows multidimensional scaling embedding of 450 realisations. As be illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is still far from the generated realisations. The space is getting dense as a result of not extending the space by increasing the number of realisations.

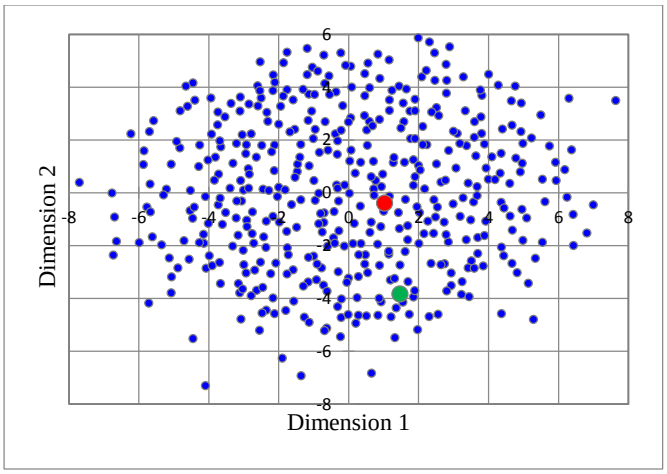


Figure 7-34: Multidimensional scaling embedding of 450 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations (SGS and 2D case)

Fourth set -Figure 7-35 shows the histogram of the interpoint distance for 1,050 realisations. As can be seen, the best distribution that can be fitted on the output distribution is lognormal. The parameters of the histogram are shown in Table 7-8

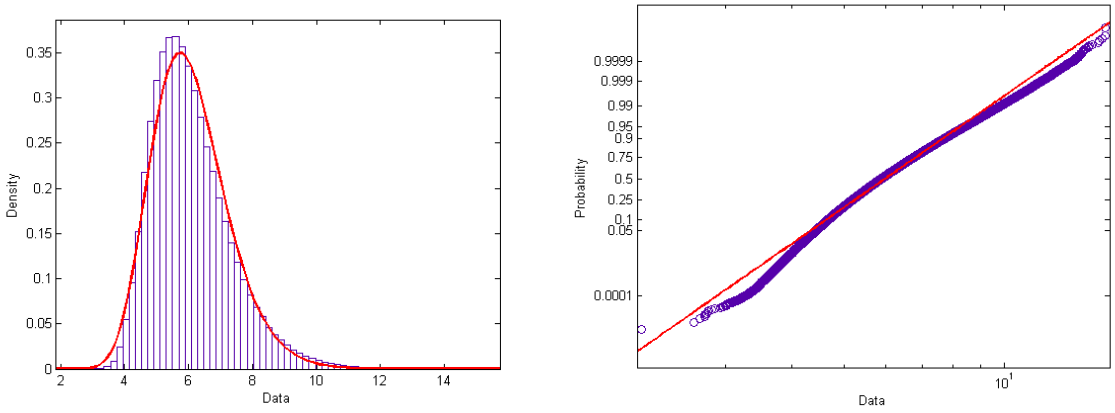


Figure 7-35: The graph on the left is the histogram of 554,931 interpoint distances of 1,050 realisations with the best fitted lognormal distribution (red line). The right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them (a $\times$ 10^6)-(SGS and 2D case)

Table 7-8: Statistical parameters of the histogram of the interpoint distances of 1,050 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
554,931	6.076	1.240	1.539	3.397	15.406	0.994	6.131	4.996

$(a \times 10^6)$

Figure 7-36 shows multidimensional scaling embedding of 1,050 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

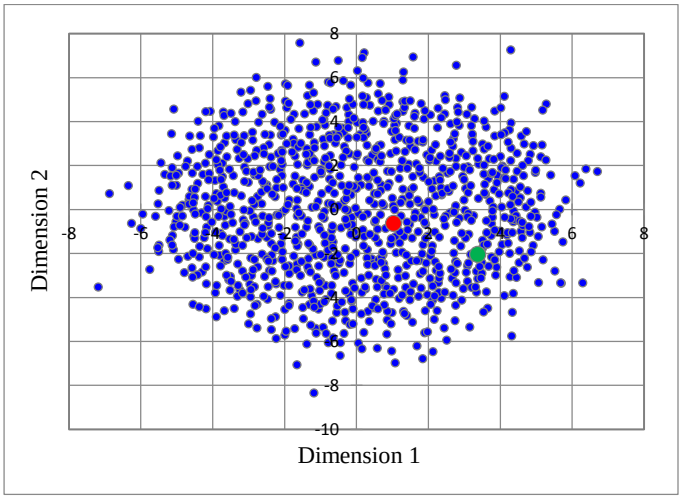


Figure 7-36: Multidimensional scaling embedding of 1,050 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations (SGS and 2D case)

7-4.1.1 Impact of the number of realisations on the size and density of the space of uncertainty (SGS-2D case)

Figure 7-37 (blue curve) shows the variation of the average of interpoint distances versus the number of realisations. As can be seen, the average distance increases when the very first realisations are generated, but after 120 realisations this number slightly decreases to be reinstated soon after. Beyond 300 realisations,

the distance remains constant and shows very insignificant fluctuations around 6.08×10^6. The red curve in Figure 7-37 shows the variation of the standard deviation of interpoint distances versus the number of realisations. As can be seen, the standard deviation increases (same as the average curve) when the first realisations are generated; but after 120 realisations it decreases to the minimum 1.2×10^6 when the rate of increase and decrease reaches a plateau. That is, the standard deviation fluctuates in a manner that is dramatically reduced by increasing the number of simulations, remaining stable at around 1.23×10^6.

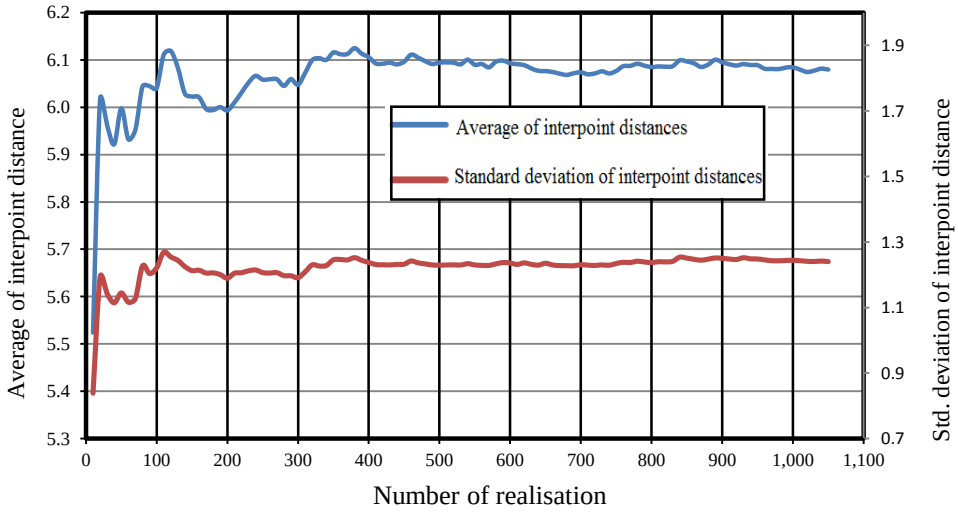


Figure 7-37: The variation of the average (blue curve) and the standard deviation (red curve) of the interpoint distances versus the number of realisations. Both of them are stabilised by increasing the number of realisations around 6.08×10^6 and 1.23×10^6, respectively (SGS and 2D case)

As illustrated earlier, the interpoint distances in the space of uncertainty follow a lognormal distribution and the mean and standard deviations of the interpoint distances are stabilised by increasing the number of realisations. Thus, it can be concluded with confidence that the simulation algorithm generates retaliations in such a way that the dissimilarly between the interpoint distances, or, more precisely, the structure of the space of uncertainty ultimately follows below the lognormal distribution.

$$f(x) = \frac{1}{(1.23 \times 10^6) \times \sqrt{2\pi} \times x} \exp\left(-\frac{1}{2 \times (1.23 \times 10^6)^2} (log(x) - 6.1 \times 10^6)^2\right) \qquad (7.2)$$

Similar to the first case (3D), the size of the space of uncertainty increases with the number of realisations but becomes stabilised by increasing (more than 730) the number of realisations (see Figure 7-38). It has to

be mentioned here that jump points in Figure 7-38 derive from very few generated realisations, namely the extreme points that are on the edge of the space of uncertainty.

The same condition occurs for the minimum distance, which is approximately stabilised after generating 300 realisations. That is, the simulation algorithm cannot generate realisations which are closer than $D_{Min} = 3.397 \times 10^6$. The main reason for that, as described before, is the conditioning data which limits the variation to the specific range $[D_{Min}, D_{Max}]$ and generating more realisations cannot extend this limit.

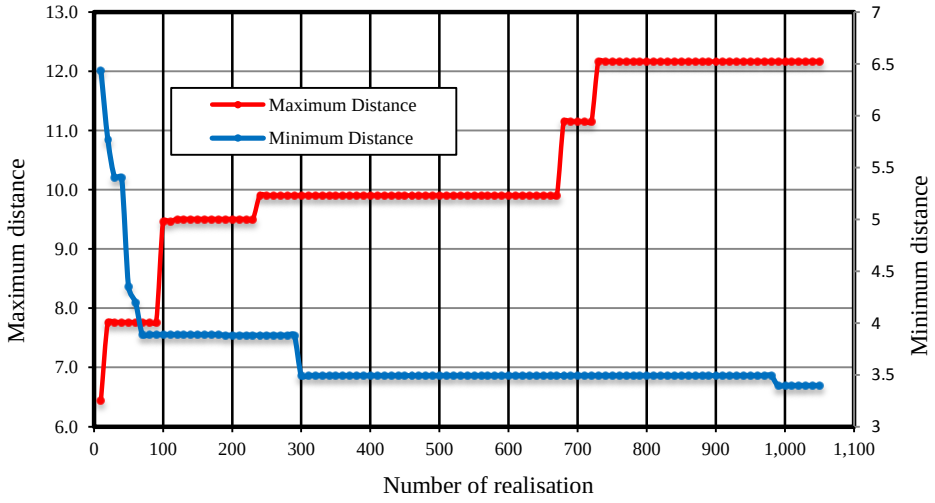


Figure 7-38: Impact of the number of generated realisations on the maximum and minimum distances between the realisations. Both the maximum and the minimum distances are stabilised by increasing the number of realisation (SGS and 2D case)

As the stress factor (in MDS) is more than 32% in this case, instead of going through the approximated distance, we can use the neighbourhood radius to evaluate the density of the points around any desired point. Figure 7-39 shows the number of realisations that fall within the neighbourhood radius $r \leq a,\ 0 < a \leq Max\ D_{rs}$ for the Kriging, the real model and the four selected realisations which are sorted based on the distance from the Kriging model. That is, realisations no.304 and no.504 are closest and farthest to the Kriging model, respectively. As can be seen, the number of realisations that fall within a fixed neighbourhood radius r for the Kriging model are considerably higher than the others. For example, there are 600 realisations in neighbourhood radius $r \leq 5 \times 10^6$ for the Kriging model (more than 57% of all realisations), while there are less than 360 for realisation no.504 and less than 120 for the real model and realisation no.504.

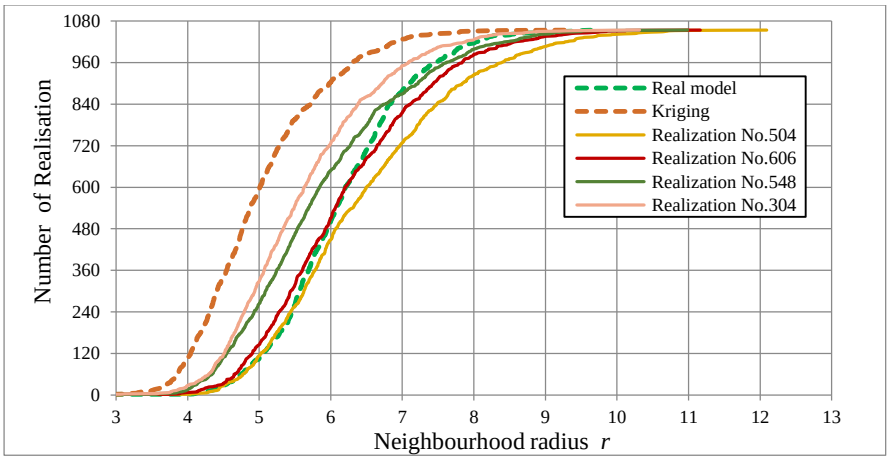


Figure 7-39: The number of realisations that fall within the neighbourhood radius r for the Kriging, the real model and the four selected realisations, which are sorted based on the distance from the Kriging model ($r \times 10^6$) (SGS and 2D case)

Moreover, Figure 7-40 shows the histogram (red) of the distances of 1,050 realisations from the Kriging model, the histogram (blue) from realisation no.504 (extreme point), and the histogram (green) from the real model. Those histograms confirm that the number of realisations in the vicinity of the Kriging model is significantly higher than the other parts of the space and by moving away from the Kriging model, the number of realisations (for the same neighbourhood radius) decreases.

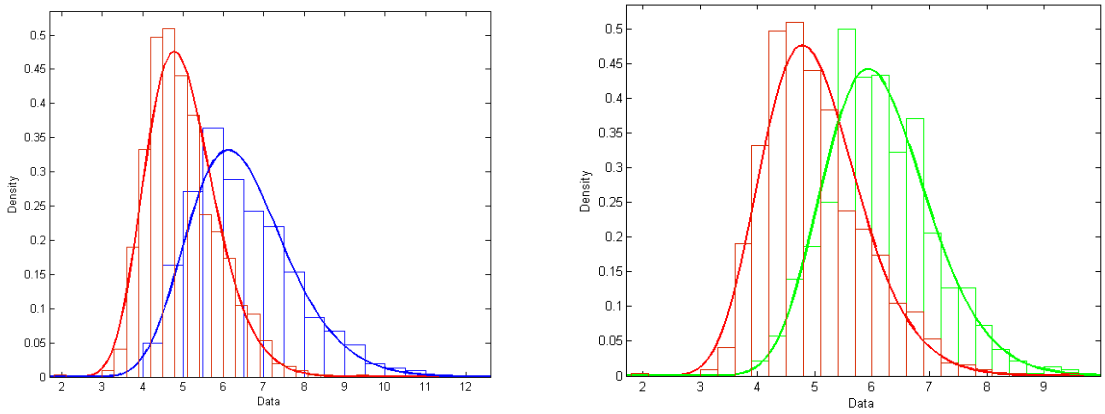


Figure 7-40: The red histogram is the distance of 1,050 realisations from the Kriging model with the best fitted lognormal distribution (red line); the blue histogram is the distance of 1,050 realisations from realisation no.504 (extreme point) with the best fitted lognormal distribution (blue line); and the green histogram is the distance of 1,050 realisations from the real model with the best fitted lognormal distribution (green line) (SGS and 2D case)

As the Kriging and the real models are fixed points in the space, we check any possible change in location of the other points with respect to these models by increasing the number of realisations. Figure 7-41 shows the variation of the average of distance and distances variance from the Kriging and the real models versus the number of realisations.

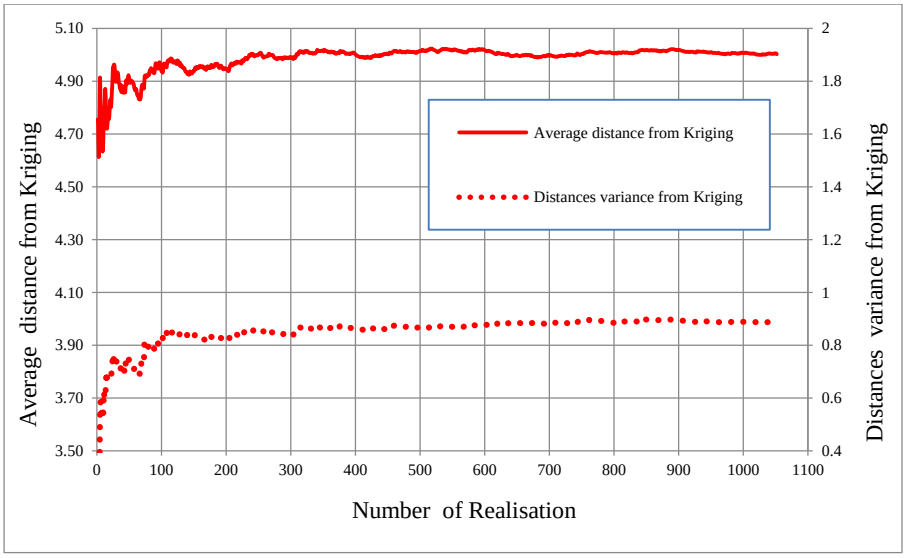


Figure 7-41: The variation of the average of distance and distances variance from the Kriging (red lines) and real models (green lines) versus the number of realisations. All of the graphs are stabilised by increasing the number of realisations (SGS and 2D case)

As can be seen, all of the graphs are stabilised by increasing the number of realisations, although they highly fluctuate before the 100th realisation. Thus, generating more realisations does not make significant changes in the structure of the space or the interpoint distance histograms from the fixed points (Kriging and real model); therefore, the structure of the space would remain unchanged.

7-4.1.2 Impact of the number of realisations on relative accuracy (SGS-2D case)

The impact of the number of realisations on relative accuracy (sub-sampling) of SGS realisations is evaluated by determining following seven sets of generated realisations from 14, 34, 54, 104, 204, and 304 to 404. It has been concluded that using more sets of realisations does not change the results. The result of these calculations is presented in Figure 7-42.

For better comparison between those sets, we compute the percentage of the number of sub-samples for each set to get the same x axis. Figure 7-43 shows the impact of the percentage of the number of sub-samples $\left(\frac{S}{N}\%\right)$ on the relative accuracy of those seven sets.

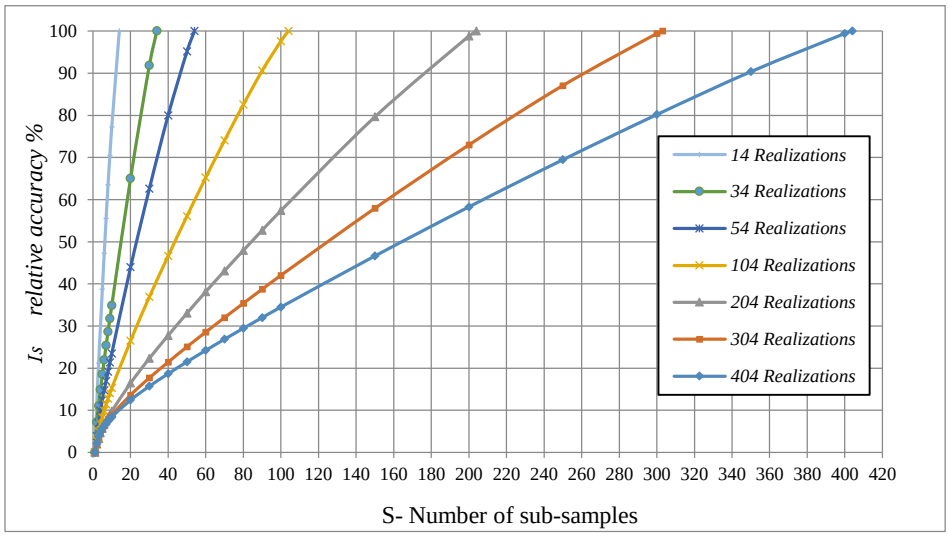


Figure 7-42: Impact of the number of realisations on relative accuracy, namely sub-sampling for the 7 sets with the following number of realisations (14, 34, 54, 104, 204, 304, and 404) (SGS and 2D case)

As can be seen, after 54 realisations (the third set) the differences between the relative accuracy of the sets become insignificant. For instance, for 20% of the sub-samples, the difference between the relative accuracy of 104 realisations (the fourth set) and 404 realisations (the seventh set) is less than 5 percent. This number would be around 2.5 percent for 40% of sub-samples. As it was explained before, that means that the distance between the best single realisation (z_1) and the set of collected S realisations does not change by increasing the number of realisations. That is, the structure of the space of uncertainty of SGS almost remains constant after generating a certain number of realisations.

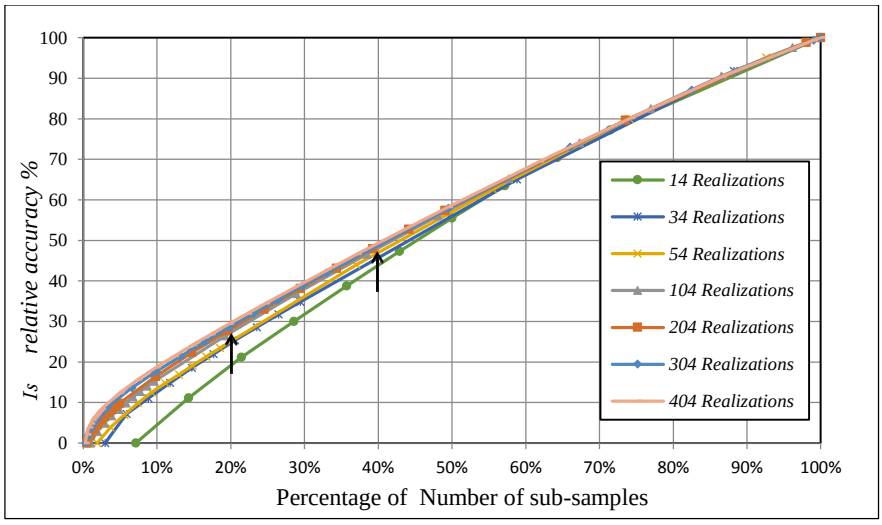


Figure 7-43: Impact of the percentage of number of sub-samples on the relative accuracy, namely sub-sampling for the 7 sets with the following number of realisations (14, 34, 54, 104, 204, 304, and 404) (SGS and 2D case)

7-4.2 Evaluating the space of uncertainty generated by TBS algorithm (2D Case)

As was described in chapter 3, the Turning Band simulation (TBS) is an unconditional Gaussian simulation algorithm, which is designed to reduce the dimensional of simulation from 3D to one-dimensional. After completing unconditional simulation procedures, the model is conditioned by sampled data. Although TBS and SGS can both be classified as Gaussian simulation, they use completely different ways, or random functions, to generate realisations (see chapter 3).

To evaluate the impact of the number of realisations on the space of uncertainty, where it is created by Turning bands simulation algorithms (TBS), 600 realisations were generated (see chapter 4) and 60 sets of generated realisations, namely $10,30,\cdots,600$ were taken by stepping 10 increment realisations from 10 to 600, each set containing all the realisations from the previous set. For each set, we repeat exactly what was done for SGS algorithm to assess the structure of the space of uncertainty.

Furthermore, similar to the previous section, for better illustration of the impact of the number of realisations on the space of uncertainty, 6 sets with the following numbers of realisations 30, 50, 100, 250, 450 and 600 are taken from 60 sets to obtain a better view of these parameters and to produce a comparative

evaluation on the space of uncertainty in detail (both the Kriging and also the real model are included in all sets). For the sake of brevity, only 3 sets with the following number of realisations (30, 100, and 600) are presented here; the rest of the sets are in appendix A. The possible interpoint distribution between the generated realisations and its statistics are assessed so a comparison can be drawn between them.

First set - Figure 7-44 shows the histogram of the interpoint distance for 30 realisations. As can be seen, the best fitted distribution is lognormal. The parameters of the histogram are shown in Table 7-9.

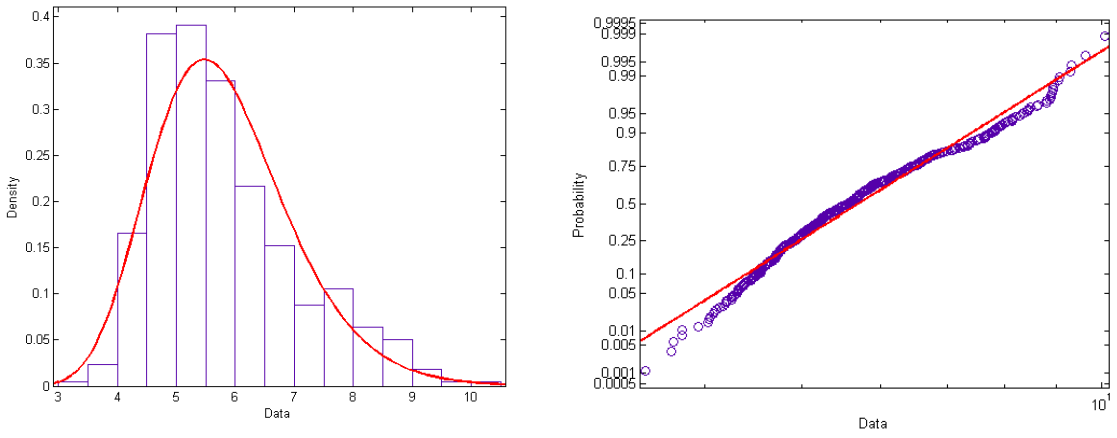


Figure 7-44: The left graph is the histogram of 435 interpoint distances of 30 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution which shows a reasonable fitting between them ($a \times 10^6$)- (TBS and 2D case)

Table 7-9: Statistical parameters of the histogram of the interpoint distances of 30 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
435	5.821	1.237	1.530	3.484	10.086	0.914	5.693	4.548

($a \times 10^6$)

Figure 7-45 shows multidimensional scaling embedding of 30 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are more or less the same as the SGS algorithm.

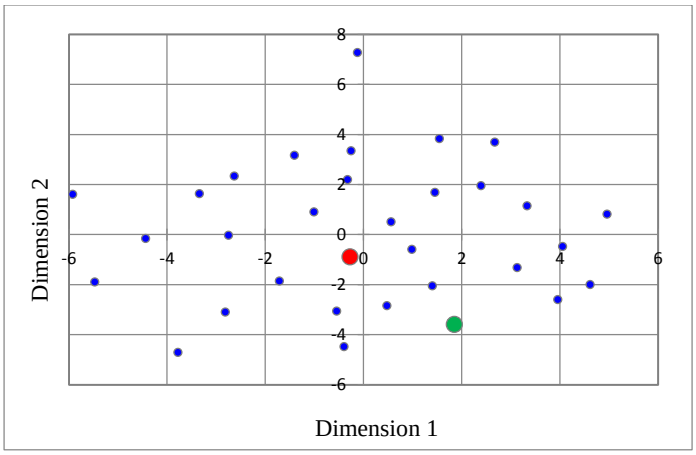


Figure 7-45: Multidimensional scaling embedding of 30 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance of the others and the real model is considerably distant from the generated realisations (TBS and 2D case)

Second set -Figure 7-46 shows the histogram of the interpoint distance for 100 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table 7-10.

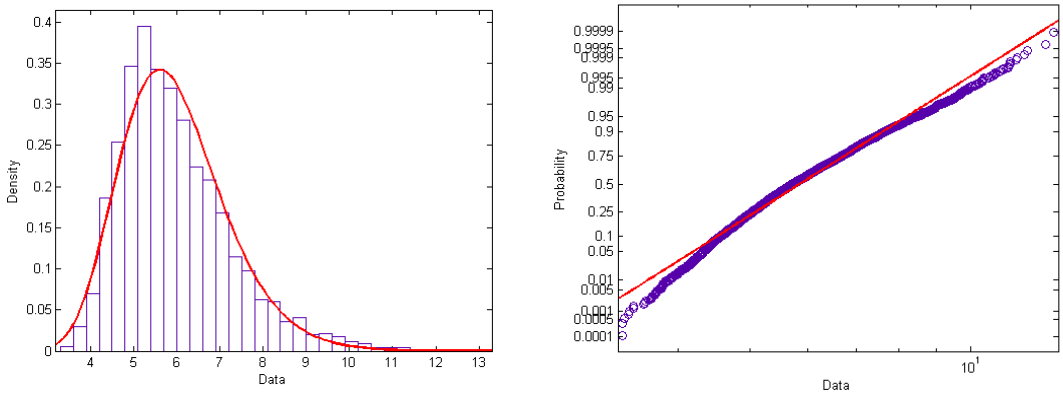


Figure 7-46: The left graph is the histogram of 4950 interpoint distances of 100 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them (a $\times$ 10^6)- (TBS and 2D case)

Table 7-10: Statistical parameters of the histogram of the interpoint distances of 100 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
4,950	5.956	1.281	1.642	3.352	13.004	1.059	5.794	4.682

Figure 7-47 shows multidimensional scaling embedding of 100 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are comparable to the SGS algorithm.

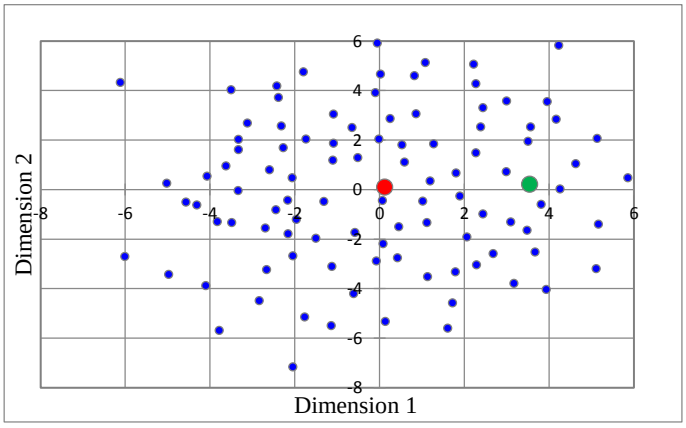


Figure 7-47: Multidimensional scaling embedding of 100 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations (TBS and 2D case)

For better viewing and to achieve more accuracy in the embedded interpoint distances, the realisations are embedded into R^3 as well. Figure 7-48 shows the embedded points in 3D. As illustrated, the real model is close to the edge of the space with a few points around it, while the Kriging is in the centre of the space. The 3D figure 7-48 is similar to the 2D Figure 7-47, but its Kruskal stress factor is 20% less than the 2D, which may give a better view of the space of uncertainty.

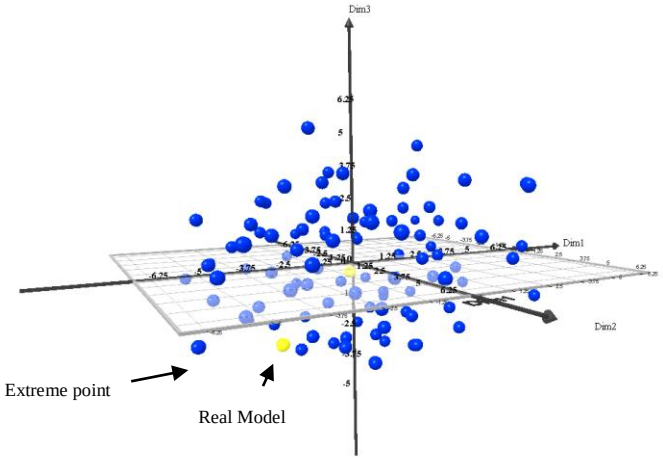


Figure 7-48: Multidimensional scaling embedding of 100 realisations (blue circles), the Kriging model and the real model (yellow circles) in R^3. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations. Embedding points in the 3D may give better accuracy and precision illustration than 2D (TBS and 2D case)

Third set -Figure 7-49 shows the histogram of the interpoint distance for 600 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table 7-11.

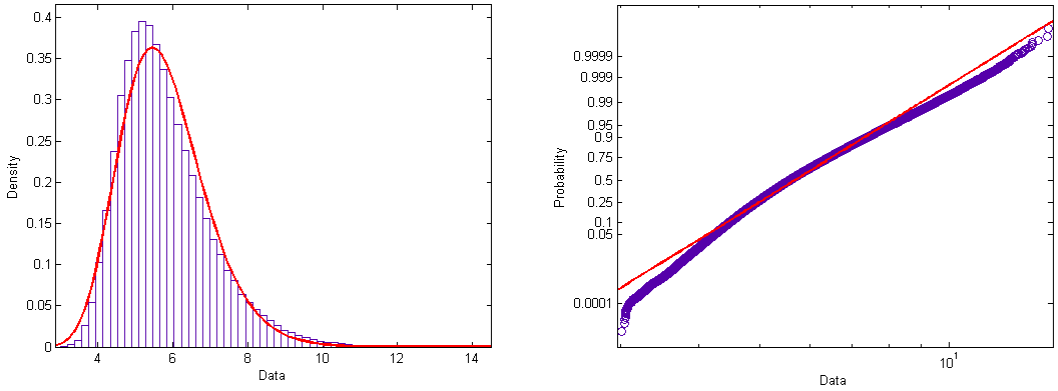


Figure 7-49: The left graph is the histogram of 179,700 interpoint distances of 600 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, demonstrating a reasonably adjusted fitting between them (a × 10^6)

Table 7-11: Statistical parameters of the histogram of the interpoint distances of 600 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
179,700	5.800	1.199	1.437	3.052	14.357	1.064	5.724	4.615

(a × 10^6)

Figure 7-50 shows multidimensional scaling embedding of 600 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are consistent with the SGS algorithm.

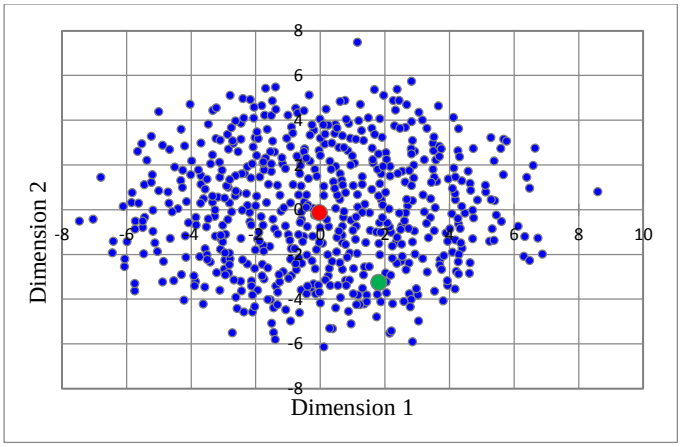


Figure 7-50: Multidimensional scaling embedding of 600 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others and the real model is quite far from the generated realisations (TBS and 2D case)

7-4.2.1 Impact of the number of realisations on size and density of the space of uncertainty (TBS-2D case)

Figure 7-51 (blue curve) shows that the variation of the average of interpoint distances versus the number of realisations. As can be seen, the average distance increases when the very first realisations are generated; however, after 5 realisations the distance decreases significantly, and then rises slightly up to 6.0×10^6. Beyond the 250th realisation, it remains constant and shows very insignificant fluctuations around 5.85×10^6. The red curve in Figure 7-51 shows that the variation of the standard deviation of interpoint distances versus the number of realisations. As can be seen, the standard deviation increases (same as the average curve) when the first realisations are generated, but after 5 realisations it decreases to a minimum of 1.2×10^6; then the rate of fluctuation is reduced and reaches a plateau. That is, the standard deviation fluctuates in a manner in which its rate reduces dramatically by increasing the number of simulations, remaining stable at around 1.21×10^6.

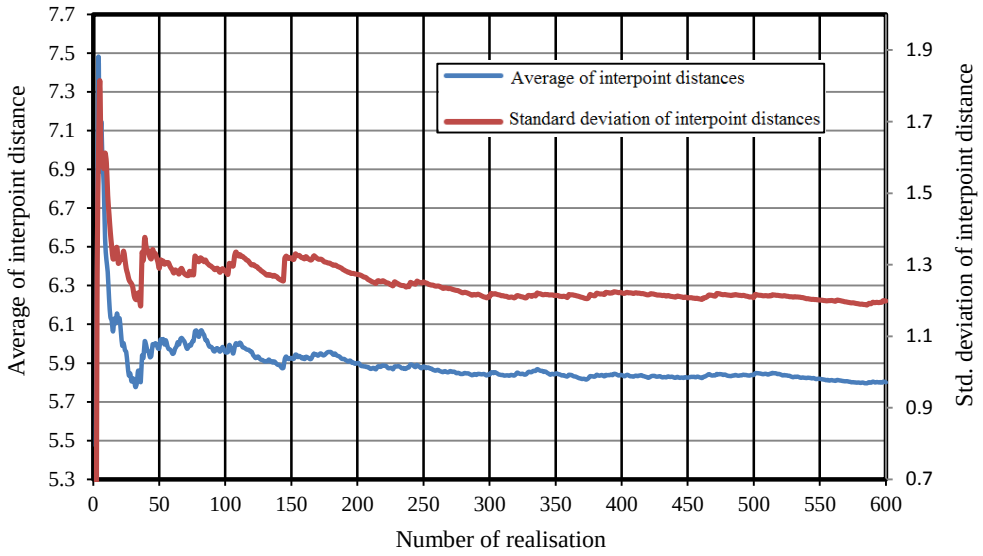


Figure 7-51: The variation of the average (blue curve) and standard deviation (red curve) of interpoint distances versus the number of realisations. Both graphs are stabilised by increasing the number of realisations around 5.85×10^6 and 1.21×10^6, respectively (TBS and 2D case)

As previously illustrated, the interpoint distances in the space of uncertainty (in this case) follow a lognormal distribution where the mean and standard deviation of the interpoint distances are stabilised by increasing the number of realisations. Thus, it can be concluded with confidence that the simulation algorithm generates retaliations in such a way that the dissimilarity between the interpoint distances, or, more precisely, the structure of the space of uncertainty ultimately follows below the lognormal distribution.

$$f(x) = \frac{1}{(1.21 \times 10^6) \times \sqrt{2\pi} \times x} \exp\left(-\frac{1}{2 \times (1.21 \times 10^6)^2} (log(x) - 5.85 \times 10^6)^2\right) \qquad (7.3)$$

Similarly to what was explained for the SGS algorithm, the size of the space of uncertainty increases with the number of realisations (see Figure 7-52), but becomes stabilised after generating 150 realisations. That is, the simulation algorithm cannot generate realisations which are far from $D_{Max} = 14.35 \times 10^6$.

The condition occurs with minimum distance, which is stabilised after generating 330 realisations. That is, the simulation algorithm cannot generate realisations which are closer than $D_{Min} = 3.05 \times 10^6$.

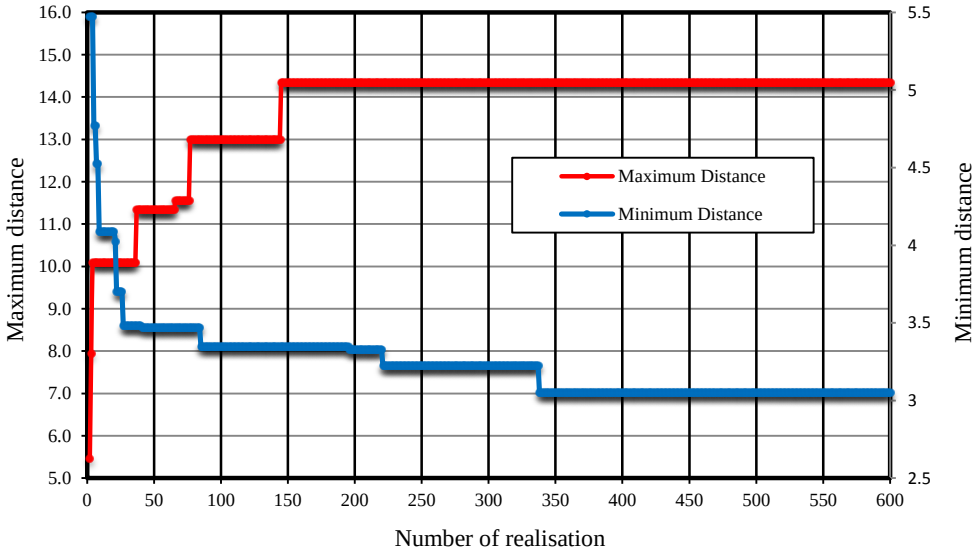


Figure 7-52: Impact of the number of generated realisations on the maximum and the minimum distances between the realisations. Both the maximum and the minimum distances are stabilised by increasing the number of realisations (TBS and 2D case)

To evaluate the density of the points around any desired points, similar to the SGS algorithm, we used method 2, called neighbourhood radius (see section 7-3.2), instead of going through the approximated distance (the stress factor is more than 32 %). Figure 7-53 shows the number of realisations that fall within the neighbourhood radius $r \leq a,\ 0 < a \leq Max\ D_{rs}$ for the Kriging, the real model and the four selected realisations, which are sorted based on the distance from the Kriging model. That is, realisations no.554 and 81 are closest and farthest to the Kriging model, respectively. As can be seen, the number of realisations that fall within a fixed neighbourhood radius r for the Kriging model are much higher than the others. For example, 450 realisations are in neighbourhood radius $r \leq 5 \times 10^6$ for the Kriging model (more than 75% of all realisations), while they are less than 225 for realisation no.554, and less than 120 for the real model. This confirms that the number of realisations in the vicinity of the Kriging model is considerably larger than in the other parts of the space. In addition, by moving away from the Kriging model, the number of realisations (for the same neighbourhood radius) decreases.

Moreover, Figure 7-54 shows the histogram (red) of the distances of 1,050 realisations from the Kriging model, the histogram (blue) from realisation no.81 (extreme point) and the histogram (green) from the real model. Those histograms confirm that the number of realisations in the vicinity of the Kriging model is substantially higher than the other parts of the space, and by moving away from the Kriging model the number of realisations (for the same neighbourhood radius) decreases.

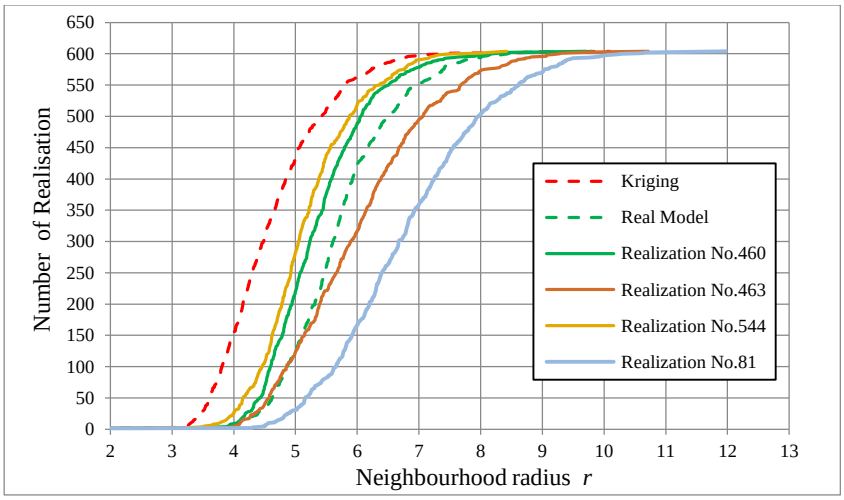


Figure 7-53: The number of realisations that fall within the neighbourhood radius r for the Kriging, the real model and the four selected realisations, which are sorted based on the distance from the Kriging model ($r \times 10^6$) (TBS and 2D case)

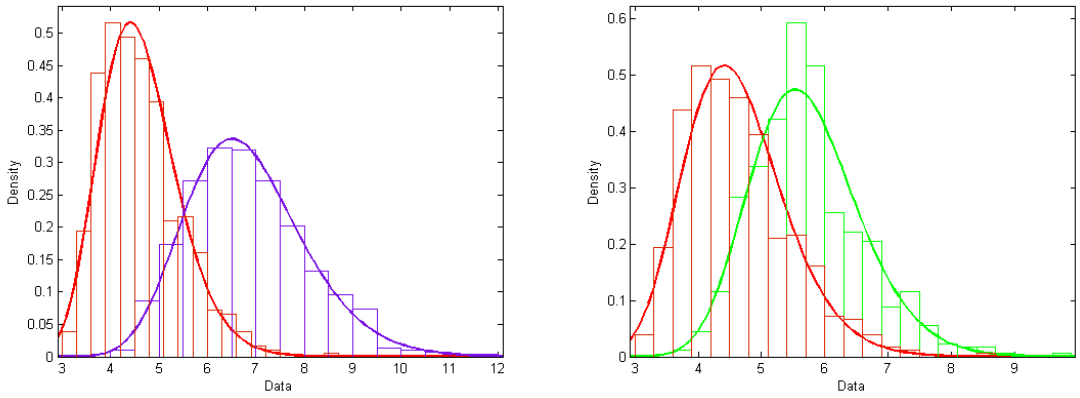


Figure 7-54: The red histogram is the distance of 600 realisations from the Kriging model with the best fitted lognormal distribution (red line); the blue histogram is the distance of 600 realisations from realisation no.81 (extreme point) with the best fitted lognormal distribution (blue line); and the green histogram is the distance of 600 realisations from the real model with the best fitted lognormal distribution (green line) (TBS and 2D case)

As the Kriging and the real model are fixed points in the space, we check any possible change in the location of the other points with respect to these models by increasing the number of realisations. Figure 7-55 shows the variation of the average distance and their variances from the Kriging and the real model versus the number of realisations.

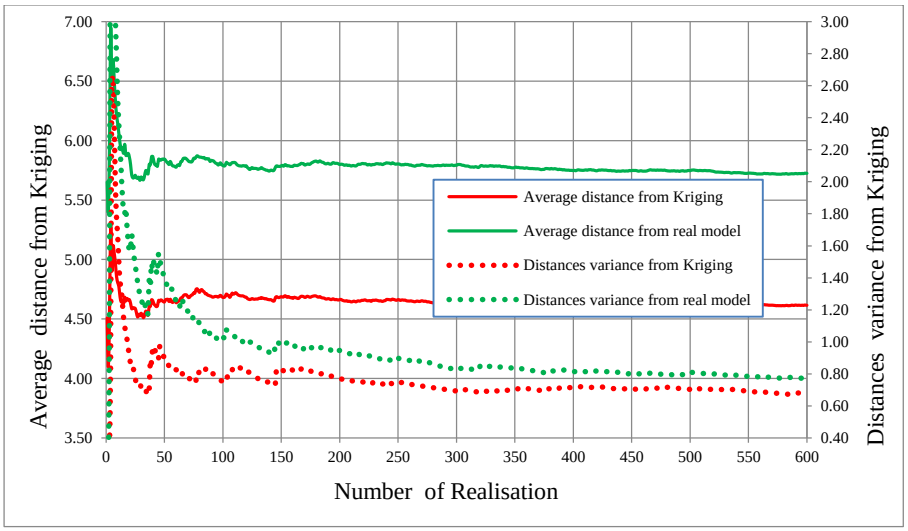


Figure 7-55: The variation of the average of distance and distances variance from the Kriging (red lines) and the real model (green lines) versus the number of realisations. All of the graphs are stabilised by increasing the number of realisation (TBS and 2D case)

As can be seen, all of the graphs are stabilised by increasing the number of realisations although they highly fluctuate before the 150th realisation. Thus, generating more realisations does not create a significant change in the structure of the space of uncertainty and distance histograms from fixed points (Kriging and real model) would be unchanged.

7-4.2.2 Impact of the number of realisations on relative accuracy (TBS-2D case)

To evaluate the impact of the number of realisations on relative accuracy (sub-sampling) of TBS realisations, we determine that simply following these seven sets of generated realisations (11, 31, 51, 101, 201, 301 and 401). The result of these calculations is presented in Figure 7-56.

For better comparison between those sets, we compute the percentage of the number of sub-samples for each set to get the same x axis. Figure 7-57 shows the impact of the percentage of the number of sub-samples $\left(\frac{S}{N}\%\right)$ on relative accuracy of the 7 sets.

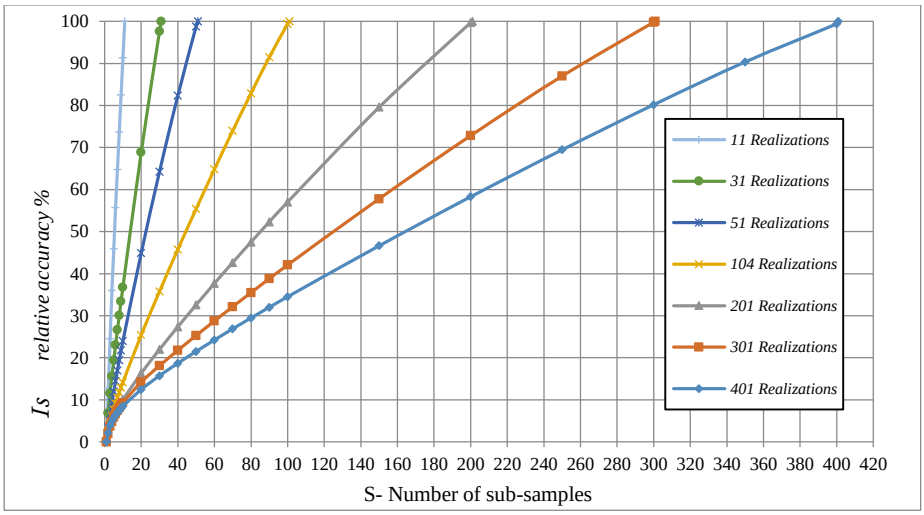


Figure 7-56: Impact of the number of realisations on relative accuracy, namely sub-sampling for the 7 sets with the following number of realisations 11, 31, 54, 101, 201, 301 and 401 (TBS and 2D case)

As can be seen, after 51 realisations (the third set) the differences between the relative accuracy of the sets become insignificant. For instance, for 20% of the sub-samples, the difference between the relative accuracy of 104 realisations (the fourth set) and 401 realisations (the seventh set) is less than 8 percent. This number would be around 5 percent for 40% of sub-samples.

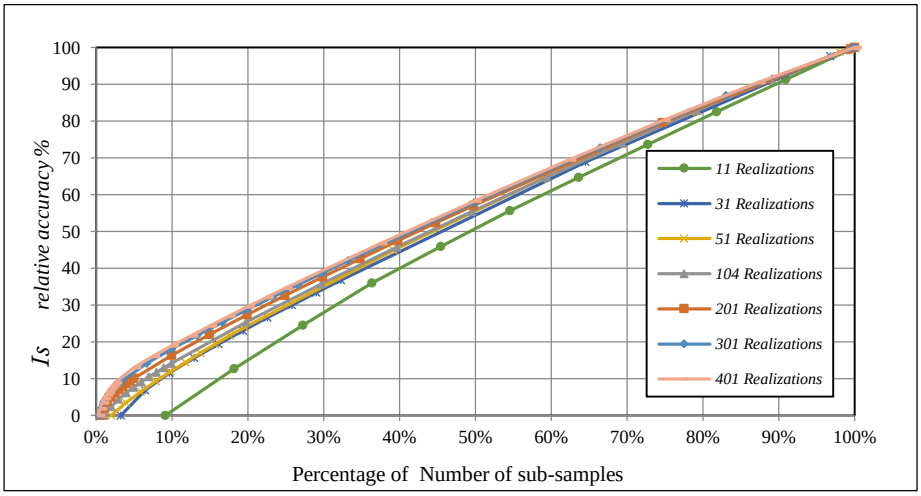


Figure 7-57: Impact of the percentage of the number of sub-samples on relative accuracy, namely sub-sampling for the 7 sets with the following number of realisations 11, 31, 51, 101, 201, 301 and 401 (TBS and 2D case)

7-4.3 Evaluating the space of uncertainty generated by SIS algorithm (2D case)

The sequential indicator simulation (SIS) is the most widely used non-Gaussian simulation algorithm (see chapter 3) which, in this point, (non-Gaussian random function) is very different from SGS and TBS.

To evaluate the impact of the number of realisations on the space of uncertainty where it is created by the sequential indicator simulation algorithm (SIS), 600 realisations were generated (see chapter 4) and 60 sets of generated realisations, namely $10,30,\cdots,600,$ are taken by stepping 10 increment realisations from 10 to 600. Also, each set of the realisations contains all the realisations of the previous set. For each set, the parameters of its dissimilarity distance matrix are calculated.

Furthermore, similar to the three dimensional (3D) data set, for better illustration of the impact of the number of realisations on the space of uncertainty, 6 sets with the following number of realisations -30, 50, 100, 250, 450 and 600 - are taken from 60 sets to obtain a better view of these parameters. They also perform a comparative evaluation on the space of uncertainty in detail (the Kriging and also the real model are included to all sets), but for the sake of brevity just 4 sets with the following number of realisations (30, 100, and 600) are presented here, the rest of the sets being introduced in appendix A. Similarly to the previous cases, we assess the possible interpoint distribution between the generated realisations and its statistics to obtain a comparison between them.

First set - Figure 7-58 shows the histogram of the interpoint distance for 30 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table 7-12.

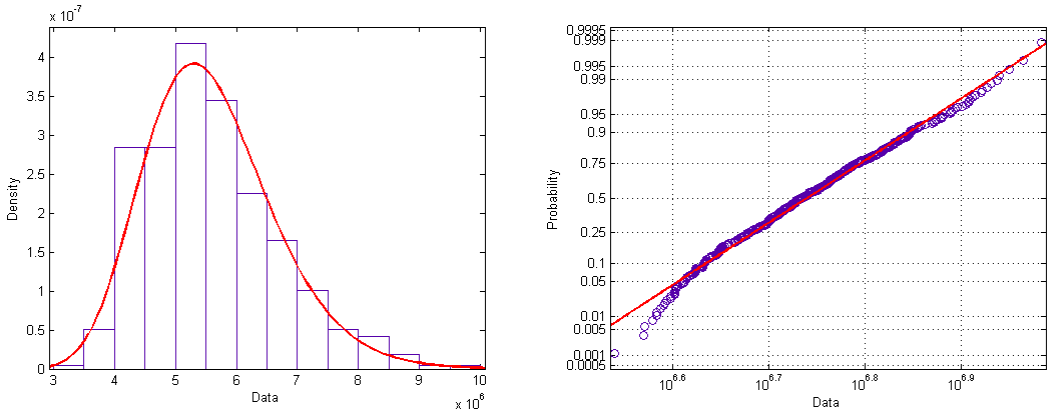


Figure 7-58: The left graph is the histogram of 435 interpoint distances of 30 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them (a $\times$ 10^6)- (SIS and 2D case)

Table 7-12: Statistical parameters of the histogram of the interpoint distances of 30 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
435	5.591	1.088	1.183	3.466	9.637	0.752	3.971	6.310	3.971

($a \times 10^6$)

Figure 7-59 shows multidimensional scaling embedding of 30 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are similar to the SGS algorithm.

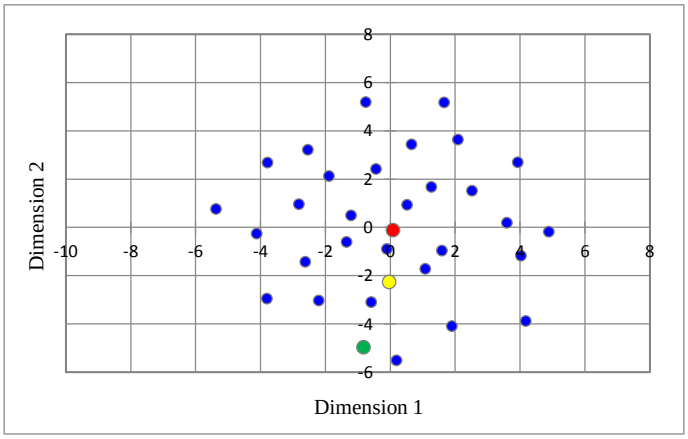


Figure 7-59: Multidimensional scaling embedding of 30 realisations (blue circles); average of 30 realisations (red circle); the Kriging model (yellow circle); and the real model (green circle) in R^2. The average model is at the minimum distance from the others and the real model is quite far from the generated realisations (SIS and 2D case)

Second set - Figure 7-60 shows the histogram of the interpoint distance for 100 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table 7-13.

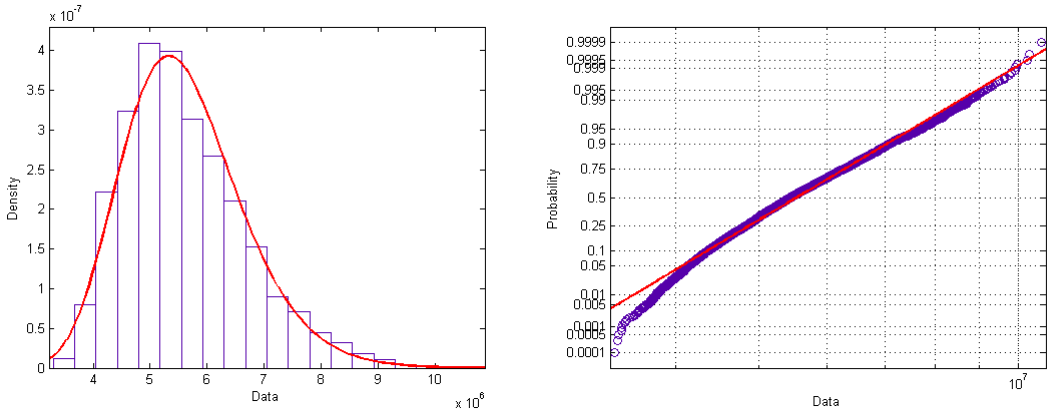


Figure 7-60: The left graph is the histogram of 4950 interpoint distances of 100 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing an adequate fitting between them (a $\times$ 10^6)- (SIS and 2D case)

Table 7-13: Statistical parameters of the histogram of the interpoint distances of 100 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
4950	5.617	1.098	1.186	3.395	10.64	0.815	4.004	6.344	4.78

($a \times 10^6$)

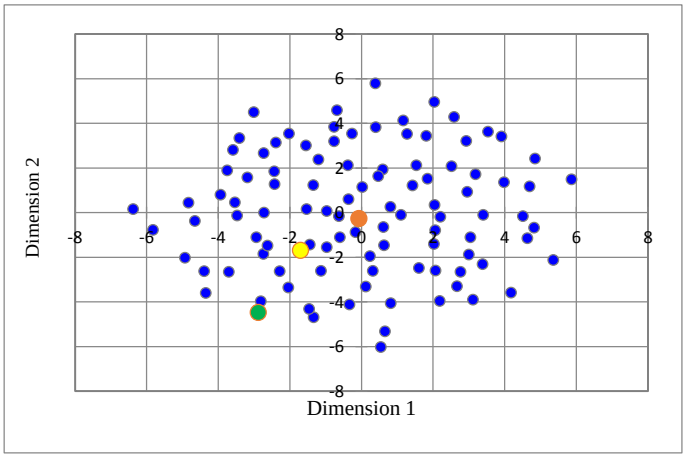


Figure 7-61: Multidimensional scaling embedding of 100 realisations (blue circles); average of 100 realisations (red circle); the Kriging model (yellow circle); and the real model (green circle) in R^2. The average model is at the minimum distance from the others and the real model is quite far from the generated realisations (SIS and 2D case)

Figure 7-61 shows multidimensional scaling embedding of 100 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

For better viewing, and in order to achieve more accuracy in the embedded interpoint distances the realisations are embedded into R^3 as well. Figure 7-62 shows the embedded points in 3D. As illustrated, the real model is close to the edge of the space with few points around it, while the Kriging is in the centre of the space. The 3D figure 7-62 is similar to 2D Figure 7-61, but its Kruskal stress factor is 20% less than 2D, which may provide a better view of the space of uncertainty.

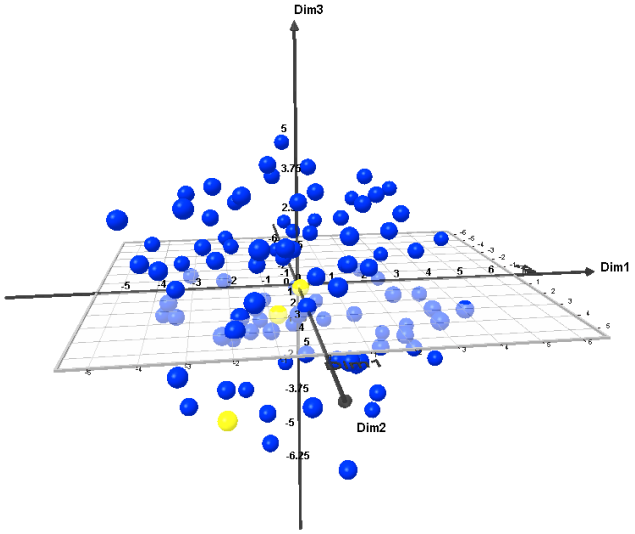


Figure 7-62: Multidimensional scaling embedding of 100 realisations (blue circles), the Kriging model and the real model (yellow circles) in R^3. The Kriging point is at the minimum distance from the others and real model is quite far from the generated realisations. Embedding points in the 3D may give a better accuracy and illustration than 2D (SIS and 2D case)

Third set -Figure 7-63 shows the histogram of the interpoint distance for 600 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table 7-14

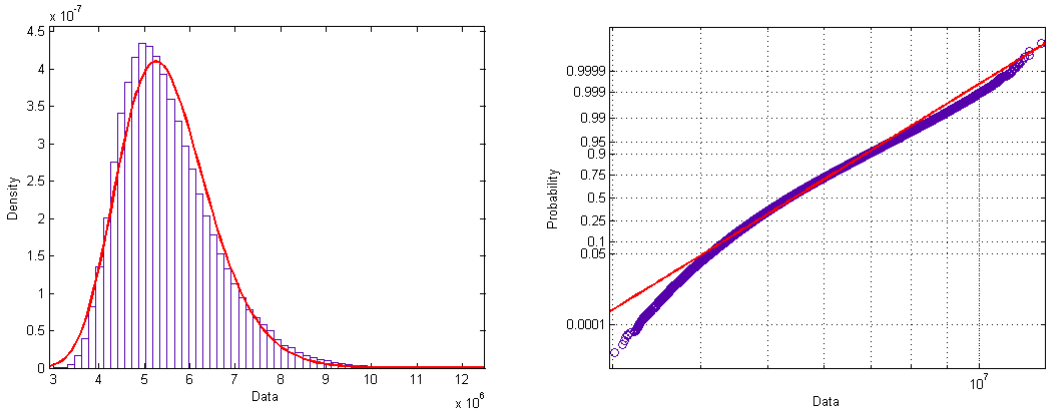


Figure 7-63: The left graph is the histogram of 210,925 interpoint distances of 600 realisations with the best fitted lognormal distribution (red line). The right graph shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, which shows a reasonable fitting between them (a $\times$ 10^6)- (SIS and 2D case)

Table 7-14: Statistical parameters of the histogram of the interpoint distances of 600 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
179,700	5.528	1.049	1.100	3.015	12.27	0.93	3.935	6.312	4.74

($a \times 10^6$)

Figure 7-64 shows multidimensional scaling embedding of 600 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are similar to the SGS and TBS algorithms.

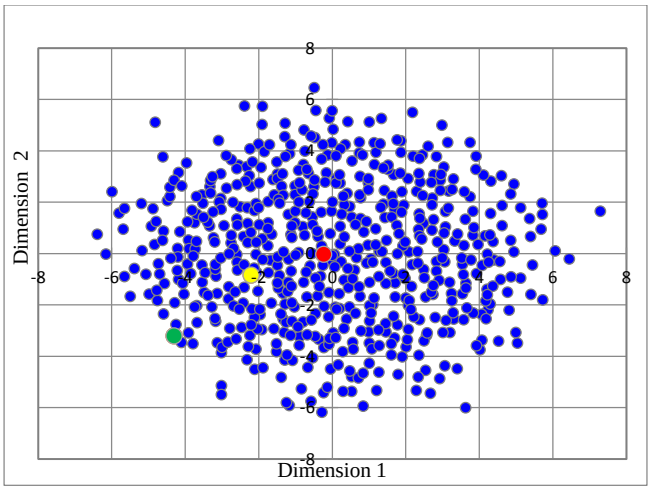


Figure 7-64: Multidimensional scaling embedding of 600 realisations (blue circles); average of 600 realisations (red circle); the Kriging model (yellow circle); and the real model (green circle) in R^2. The average model is at the minimum distance from the others and real model is quite far from the generated realisations (SIS and 2D case)

7-4.3.1 Impact of the number of realisations on the size and density of the space of uncertainty (SIS-2D case)

Figure 7-65 (blue curve) shows the variation of the average of interpoint distances versus the number of realisations. As can be seen, the average distance increases when the very first realisations are generated, but after 40 realisations it decreases significantly, and subsequently increases slightly up to 6.0×10^6. Beyond 200 realisations, it remains constant and shows insignificant fluctuations around 5.55×10^6. The red curve in Figure 7-65 shows the variation of the standard deviation of interpoint distances versus the number of realisations. As can be seen, the standard deviation increases (same as the average curve) when the first realisations are generated, but after 25 realisations it decreases to a minimum of 1.0×10^6. Subsequently, the rate of increase and decrease reduces until it plateaus. That is, the standard deviation fluctuates until it reduces dramatically by increasing the number of simulations; it finally remains stable at around 1.05×10^6.

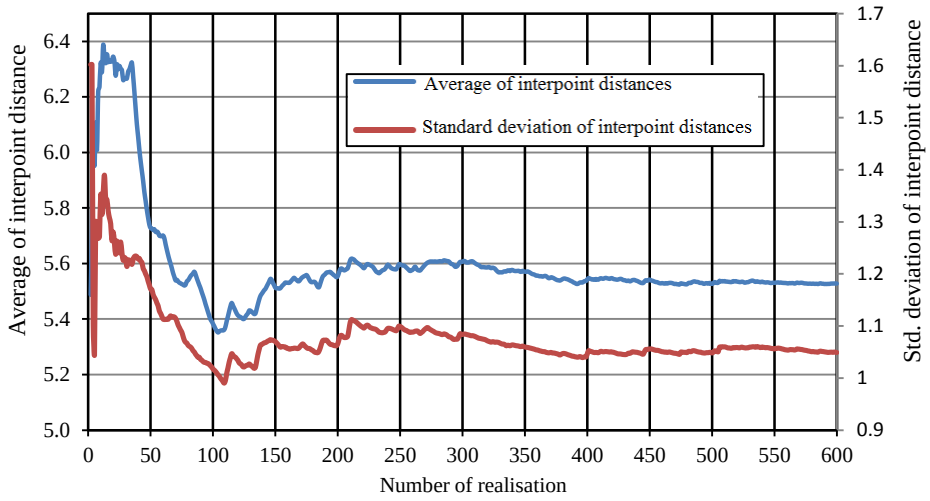


Figure 7-65: The variation of the average (blue curve) and standard deviations (red curve) of interpoint distances versus the number of realisations. Both of them are stabilised by increasing the number of realisations around 5.5×10^6 and 1.05×10^6, respectively (SIS and 2D case).

We illustrated early the interpoint distances in the space of uncertainty (in this case) following a lognormal distribution As the mean and standard deviations of the interpoint distances are stabilised by increasing the number of realisations, it can be concluded with confidence that the simulation algorithm generates retaliations in such a way that the dissimilarity between the interpoint distances, or precisely the structure of the space of uncertainty, ultimately follows below the lognormal distribution.

$$f(x) = \frac{1}{(1.05 \times 10^6) \times \sqrt{2\pi} \times x} \exp\left(-\frac{1}{2 \times (1.05 \times 10^6)^2} (log(x) - 5.50 \times 10^6)^2\right) \qquad (7.4)$$

Similar to what was explained for the SGS algorithm, the size of the space of uncertainty increases with the number of realisations in the five steps (see Figure 7-66) and is stabilised after generating 150 realisations. That is, the simulation algorithm cannot generate realisations which are far from $D_{Max} = 14.35 \times 10^6$.

There is the same condition for the minimum distance, where the graph stabilised after generating approximately 330 realisations. That is, the simulation algorithm cannot generate realisations which are closer than $D_{Min} = 3.05 \times 10^6$.

To evaluate the density of the points around any desired points, similar to the SGS and TBS algorithms, we used method 2, namely the neighbourhood radius (see section 7-3.2), instead of going through the

approximated distance (the stress factor is more than 32 %). Figure 7-67 shows the number of realisations that fall within the neighbourhood radius $r \leq a,\ 0 < a \leq Max\ D_{rs}$ for the Kriging, the real model and the four selected realisations, which are sorted based on the distance from the Kriging model.

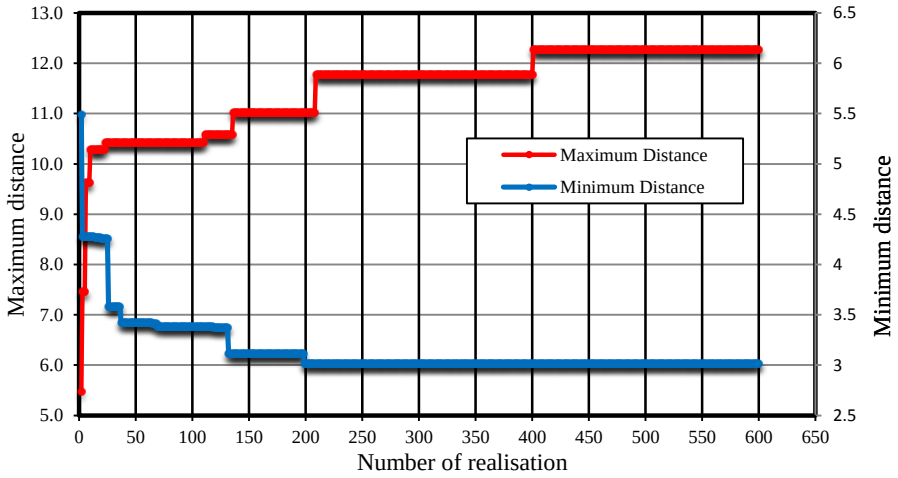


Figure 7-66: Impact of the number of generated realisations on the maximum and minimum distances between the realisations. Both of them are stabilised by increasing the number of realisations (SIS and 2D case)

That is, realisations no.554 and no.81 are closest and farthest to the Kriging model, respectively. As can be seen, the number of realisations that fall within fixed neighbourhood radius r for the Kriging model are much higher than in the other parts of the space. For example, 450 realisations are in neighbourhood radius $r \leq 5 \times 10^6$ for the Kriging model (more than 75% of all realisations), while they are less than 225 for realisation no.554 and less than 120 for the real model.

This confirms that the number of realisations in the vicinity of the Kriging model is considerably higher than in the other parts of the space and by moving away from the Kriging model, the number of realisations (for the same neighbourhood radius) decreases. The result is the same as what has already been illustrated for the Gaussian algorithms.

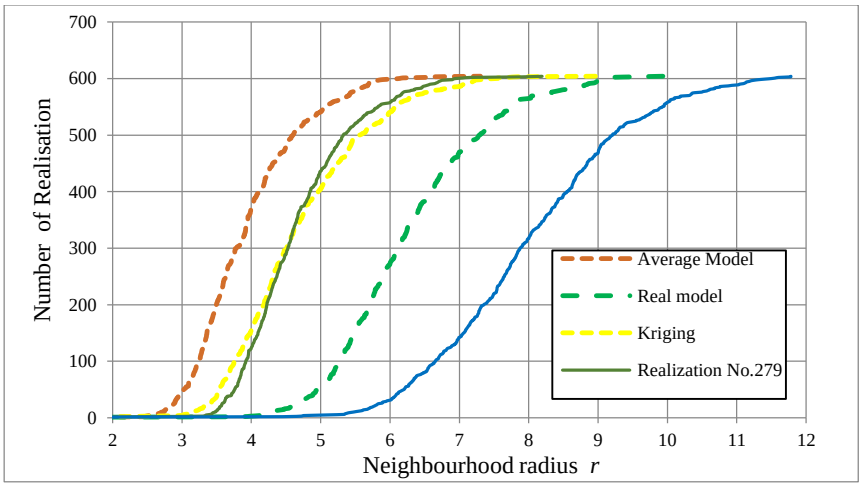


Figure 7-67: The number of realisations that fall within the neighbourhood radius r for average, the Kriging and the real models, as well as the two selected realisations no.279 and no.277, which have closest and farthest distances from the Kriging model, respectively (SIS and 2D case)

Moreover, Figure 7-68 shows the histogram (red) of the distances of 600 realisations from the Kriging model, the histogram (blue) from realisation no.277 (extreme point), and the histogram (green) from the real model. Those histograms confirm that the number of realisations in the vicinity of the Kriging model is much higher than in the other parts of the space, and by moving away from the Kriging model the number of realisations (for the same neighbourhood radius) decreases.

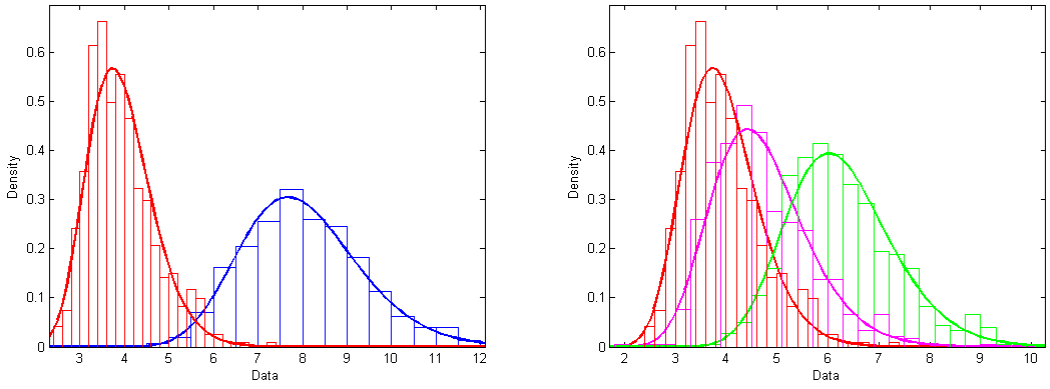


Figure 7-68: The red histogram is the distance of 600 realisations from the Kriging model with the best fitted lognormal distribution (red line); the blue histogram is the distance of 600 realisations from realisation no.277 (extreme point) with the best fitted lognormal distribution (blue line); and the green histogram is the distance of 600 realisations from the real model with the best fitted lognormal distribution (green line) (SIS and 2D case)

7-4.3.2 Impact of the number of realisations on relative accuracy (SIS-2D case)

To evaluate the impact of the number of realisations on relative accuracy (sub-sampling) of TBS realisations, we determined that by simply following seven sets of generated realisations (11, 31, 51, 101, 201, 301 and 401) and using more sets does not change the results. The result of these calculations is presented in Figure 7-69.

For better comparison between those sets, we compute the percentage of the number of sub-samples for each set to achieve the same x axis. Figure 7-70 shows the impact of the percentage of the number of sub-samples $\left(\frac{S}{N}\%\right)$ on relative accuracy of the 7 sets. As can be seen, after 34 realisations (the third set), the differences between the relative accuracy of the sets has become insignificant. For instance, for 20% of sub-samples, the difference between the relative accuracy of 104 realisations (the fourth set) and 404 realisations (the seventh set) is less than 2.5 percent. This number would be around 1 percent for 40% of sub-samples.

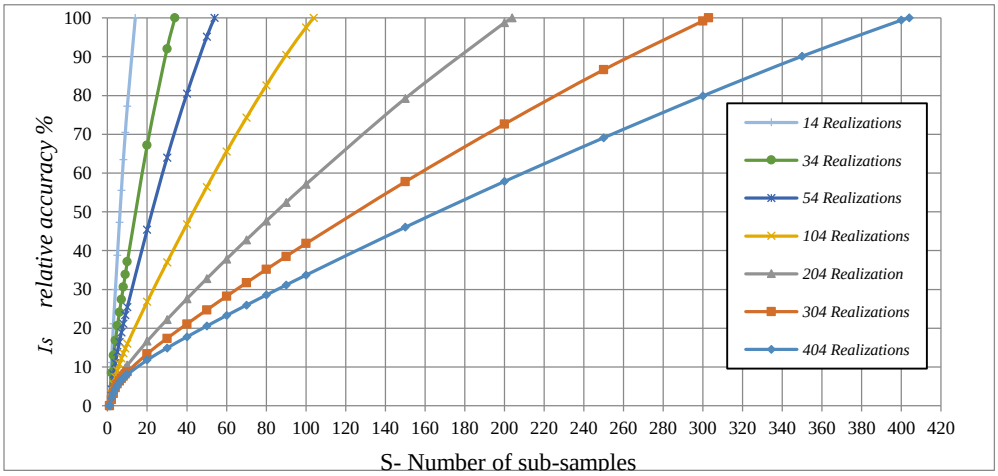


Figure 7-69: Impact of the number of realisations on the relative accuracy, namely the sub-sampling for the 7 sets with the following number of realisations 14, 34, 54, 104, 204, 304, and 404 (SIS and 2D case)

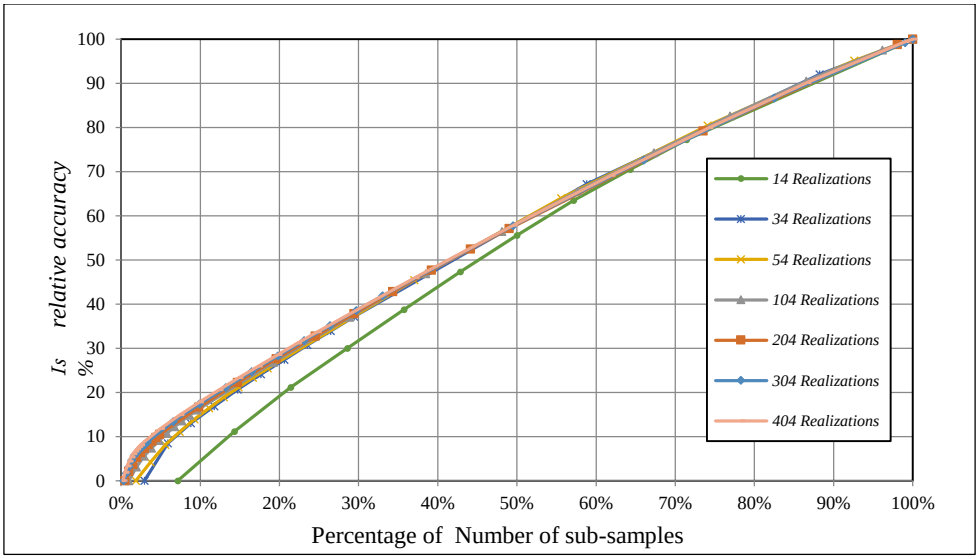


Figure 7-70: Impact of the percentage of the number of sub-samples on the relative accuracy, namely sub-sampling for the 7 sets with the following number of realisations (14, 34, 54, 104, 204, 304 and 404) (SIS and 2D case)

After 34 realisations, the differences between the relative accuracy of the different sets have become insignificant. As previously explained, that means that the distance between the best single realisation (z_1) and the set of collected S realisations does not change by increasing the number of realisations. That is, the structure of the space of uncertainty of SIS remains practically unchanged after generating a certain amount of realisations.

7-5 Conclusion

Although the conclusions that can be presented by this chapter, to some degree, would be specific to the two different used data sets (2D and 3D), it is clear that the space of uncertainty generated by three simulation algorithms (SGS, TBS and SIS) when the same information (conditioning data, histogram, variogram) is being used, may not vary significantly from one algorithm to another. As was shown, in spite of significant and obvious differences between the two data sets used, the results of those cases are almost the same (in the point of comparison between the simulation algorithms).

The factors that influence the characteristics of the space of uncertainty and have been addressed in this study are, as follows:

1. We know that condition simulation does not contain any systematic bias and all generated realisations are equally probable and fairly represent the entire space of uncertainty. However, we can see in this study that there is a larger probability to generate realisations in the neighbourhood of the Kriging or the E-type models rather than on the edge of the space. This means that there is a systematic tendency to configure the space of uncertainty in such a way that the probability to find or generate the realisations decreases from the centre to the edge of the space.

2. In the majority of geoscience applications only a few realisations can be chosen. If this selection is based on a random selection (which normally is), the set of chosen realisations is not fairly sampled and the realisations which are around the Kriging or the E-type model have a higher chance to be collected.

3. The interpoint distance histograms (in the space of uncertainty) follow lognormal distributions and the mean and the standard deviations of the interpoint distances are stabilised by increasing the number of realisations. Thus, it can be concluded with confidence that the simulation algorithms generate retaliations in such a way that the dissimilarity between them (interpoint distances), or precisely the structure of the space of uncertainty, ultimately follows a lognormal distribution. It is not easy to explain why the similarity or interpoint distances in the space of uncertainty tend to be approximately lognormally distributed. The main reason may be that all the lognormal algorithms are conditional to sampled points; therefore, that would not allow realisations to be generated with so many differences (dissimilarity). That is, the generated realisations tend to be similar (close) to each other rather than far from, and this may create the positive skew distribution.

4. By increasing the number of generated realisations the size of the space of uncertainty would not change, only the point of density increases.

5. We know that the condition simulation does not have any systematic bias, so all generated realisations are equally probable and fairly represent the entire space of uncertainty. However,

there is a larger probability to generate realisations in the neighbourhood of the Kriging or the E-type model.

6. By increasing the number of generated realisations, the difference between the relative accuracy of the sets, namely sub-sampling, would be insignificant. That means that the structure of the space of uncertainty remains relatively unchanged.

7. As can be seen in Table 7-15, there are no major differences between the means, the standard deviations, the variances and the minimum and maximum of the simulation algorithms. However, SGS shows a higher variation (in interpoint distances) than the other algorithms while the SIS shows the lowest. That means that the ability to generate different realisations (that are not similar to each other) for SGS is the highest and for SIS is the lowest. That can be confirmed again by comparing the minimum and the maximum graphs of the interpoint distances of the simulation algorithms (see Table 7-15).

8. As we explained before, for a better comparison between the simulation algorithms, two points are fixed in all three spaces of uncertainty, namely the real model (exhaustive data set) and the Kriging model. As can be seen in Table 7-15, the average distances between the generated realisations (by simulation algorithms) and these fixed points are close to each other.

Table 7-15: Statistical parameters of the interpoint distances distributions of the simulation algorithms (All 2D – cases)

Simulation Algorithm	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
SGS	6.076	1.240	1.539	3.397	15.406	0.994	6.131	4.996
TBS	5.800	1.199	1.437	3.052	14.357	1.064	5.724	4.615
SIS	5.528	1.049	1.1	3.015	12.27	0.930	6.312	4.740

$(a \times 10^6)$

Chapter 8

8 Stochastic mine design and risk analysis using metric space

8-1 Introduction

As explained in the literature review, traditional methods of mine planning involve the creation of a schedule using information from a single deterministic block model. Often the schedule is optimised to maximise, for example, the net present value (NPV). Currently, multiple randomly generated sets of block models, called realisations, are available. Each realisation is compatible with sampled data and satisfies various geostatistical restrictions. For over a decade methods that are based on selecting a single schedule to exploit the orebody, using multiple realisation information, have slowly been developed. Many of these rely on constructing a collection of schedules and assessing these schedules in a limited way against the available realisations. In this chapter we highlight the shortcomings of such approaches and put forward a novel methodology to mitigate these shortcomings. Our new approach works directly with a candidate set of schedules; for example, those generated by individually optimising each realisation. We introduce a means for measuring the difference between pairs of schedules in this collection and, using this information, select a small set of candidate schedules for further analysis by NPV. We argue that this small set of candidate schedules produce more robust outcomes than schedules selected by other existing approaches.

While conditional simulations are relatively commonly used by the mining industry to assess grade uncertainty in mine optimisation, mine scheduling and mine planning activities, there has been very little work, if any, on quantifying how well a given collection of realisations represent the total grade of uncertainty in mine designs.

Although quantifying the space of uncertainty of all kinds of stochastic simulations and subcollecting realisation(s) that best represent the space of grade uncertainty have been well developed and addressed in chapter 6, the nonlinearity of the optimisation process (as a transfer function) is still a big challenge when applying the representative realisations to the mine design process. This is because only a limited number of realisations are generated and used (as a finite collection of all possible outcomes) by any sort of risk based mine design method.

Risk based mine design methods usually generate different mine designs based on a stochastically simulated orebody. Mathematically, this means that the transfer functions (such as pit optimisation and any kind of mine scheduling algorithms) are applied on the space of grade uncertainty and not only produce a distribution of possible project indicators, but also make a new space, which is commonly named 'response'.

We differentiate between two types of responses. The first is response data that can be represented by number(s) (type I), for example any project performance indicator, such as NPV, Internal Rate of Return (IRR), cash flow, ore tonnage or metal quantity. The second type of response can be represented by sets or series of numbers (type II). Type II can be a series of optimal extraction sequences (OES) of blocks or even a set of spatially connected blocks with a pit geometry, such as a series of nested pits, designed pushbacks or even the set of blocks which is mined during a fixed period of time, for example, yearly. That is, type II responses show how a mine has to be mined to get the highest possible NPV, for example.

In the past, several research groups have addressed stochastic optimisation and mine planning to evaluate the risk of project performance indicators (type I) using conditional simulations; however, they have two drawbacks. The first is that the distribution of the project indicators (type I) cannot provide any information about different optimised pits, pit designs or mine scheduling; in addition, it is likely that different pit designs have approximately the same project indicators. Therefore, this sort of response variable (type I) is unable to reflect the dissimilarities between different optimised pits or pit designs, which come from different realisations. Generally, type I highlights the impact of grade uncertainty (in mine planning activities) on the assessment of the financial or technical indicators of the blocks. Type II, however, indicates this impact on mine design and assesses the uncertainty in sequences or the time extracting the blocks-uncertainty in place (which block) and time (when to mine).

The second drawback is that the chosen design(s) would no longer be equiprobable. This means that some designs are highly likely to occur while others would be less able to represent the actual mine design. According to Dimitrakopoulos et al., (2007) "*Although the simulated orebody models are equally probable, the corresponding designs are not*" (p.76). The common risk based methods assume that chosen design(s) are all equiprobable and this assumption may cause misleading results in inaccurate design selection.

Although we developed and described an application of the distance based method in chapter 6 to quantify the dissimilarity of different realisations, we very briefly summarise and restate the methodology and findings of chapter 6 here in order to apply them to pit optimisations, mine designs or scheduling using a type II response. We also use this information to optimally subsample a collection of the designs, quantify

how well the high-quality collection represents the overall uncertainty and measure the weights of the chosen designs. In this chapter we address the need to take into account the probability or weight of the chosen design(s) that can be produced by any risk based mine design method. This approach can be readily generalised to promote a more robust performance of risk based mine design methods.

8-2 Risk based methods

Some risk based methods (Dimitrakopoulos et al., 2007) and (Leite and Dimitrakopoulos, 2007) generally try to choose a single or modified mine design among the other designs, which may provide the best or better results based on the quantified risk of the project indicators. They deal with a limited number of designs without considering or identifying the structural relationships between them. Notably, there are spatial differences (dissimilarities), which may affect the results of further processing.

The maximum upside or minimum downside (Dimitrakopoulos et al., 2007) approach, as described in chapter 2, is one such risk-based method. This method, briefly, first applies traditional optimisation and mine scheduling on each random generated realisation to design a pit. Second, it generates the distributions of any other project indicator by applying a given mine plan on each realisation. Third, it discards the designs that may not meet user defined criteria and selects one design that can capture the maximum upside reward or minimum downside risk. The design selection in the approach outlined above is based on type I of the response variable, while the designs are significantly different in ways that cannot be captured only by type I. Considering the use of type II in these methods may avoid the two following issues.

The first simple example that addresses the first issue, is shown in Figure 8-1, where there are four ore blocks which have a Gaussian grade distribution of which means are shown on the blocks (the standard deviations are the same). We randomly draw grade values from their grade distributions to generate 10,000 equal probable realisations; then, we try to find all possible mining combinations (block sequences) in such a way that the high grade blocks are mined first if their top block has already been mined. As can be seen in Figure 8-1, there are only six possible sequences with the given frequencies. It is clear that the first extraction sequence is highly likely to occur with a probability of more than 68%, while the fifth is less likely to occur, with a probability of 0.36%.

As risk based methods do not consider any probability or weight for the chosen design(s), it is highly likely that a low probable design, for example path 5, can satisfy the conditions, such as maximising the upside and minimising the downside, while highly probable designs might be discarded. It is clear that higher probable design(s) have to be focused on doing any further assessment.

This example, albeit fabricated, reveals the importance of considering the probability or weight of each design before making any decisions about them. In contrast to this example, in real case, all possible mining combinations (block sequences) may be completely different from each other. This means that the probability that two different realisations have exactly the same optimal extraction sequences (OES) is very low. In such cases, we try to find or measure the similarity (or dissimilarity) between them, instead of exactly matching them. This means that although there are many different OES, some are more similar to each other than others.

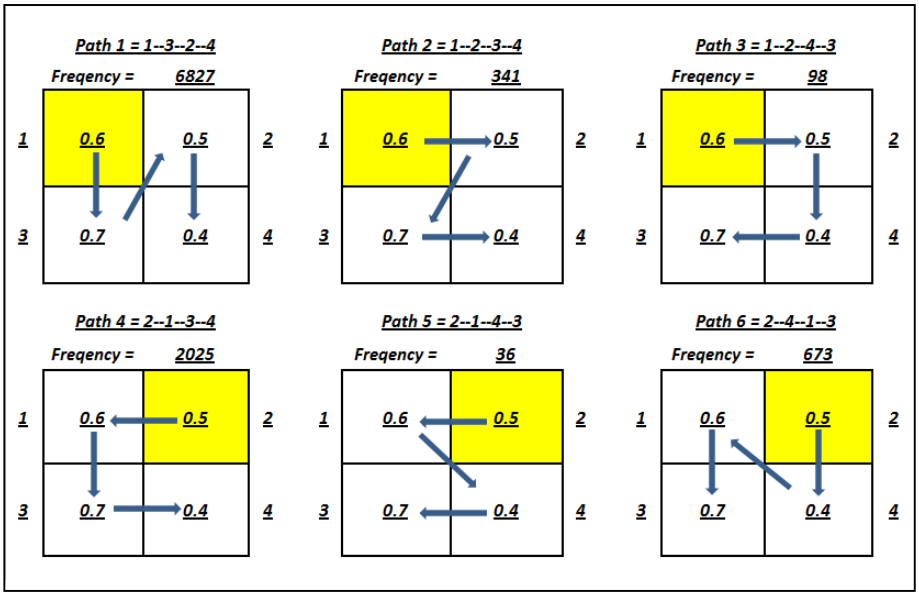


Figure 8-1: Six possible mining block sequences (schedules) and their probabilities

This approach addresses the second issue to be avoided. In the second issue, each schedule is individually optimised regarding NPV, production targets (such as ore production and yearly target grade) and mining constraints. However, applying these optimised schedules to the other realisations incurs a high risk of not meeting production targets and, therefore, the new schedule which is not optimised would not guarantee practical solutions. Although they may show high NPV in the point of maximum upside and minimum downside, they would not be achievable.

Moreover, it would be a time consuming procedure to go through each realisation one by one, checking production targets or mining constraints. For example, in this theoretical case, more than 10,000 schedules would have to be checked. We explain in section 8-5.2 how to apply the distance based method to select a set of schedules that can solve this issue.

In this fabricated example we have simulated 4-block models and have shown the relative percentages of each possible extraction sequence (path) to find out which one would most likely happen. Now, we take an advantage of this example to explain how to select the most popular extraction sequence(s).

As we explained on chapter 6, our sub-sampling method based on Kantorovich distances treats extraction sequences as vectors or objects having distances from each other. It helps us to partition the extraction sequences into a few clusters, such that extraction sequences within each cluster are as similar to each other as possible, and dissimilar from objects in other clusters as possible. Size of each cluster depends on how many points (extraction sequences) sit around (nearby) the cluster centre point. A cluster centre point is a point which minimises the sum of distances between it and other points in cluster. We will take these points (or only one point) as representative(s) or the most popular extraction sequences for mine design.

In this example, there are only six possible extraction sequences, therefore numbers of possible cluster would be six. As each extraction sequence appears many times, so the size of each cluster would be equal to the given frequencies (see Table 8-1), namely weight($\widehat{\omega}_s$). Each cluster centre is a vector in R^4 , and we are able to calculate Kantorovich distance (dissimilarity) between them based on extraction sequences and geometric spatial information of the blocks. However, as the most popular extraction sequences, namely cluster centre 1 has more than half of total weight, it doesn't need to calculate dissimilarity matrix for these 6 extraction sequences.

Table 8-1: shows the extraction sequences (vectors) and their weights of each cluster centre.

	Block 1	Block 2	Block 3	Block 4	Weight ($\widehat{\omega}_s$)
Path 1/Centre of cluster 1	1	3	2	4	68.27%
Path 2/Centre of cluster 2	1	2	3	4	3.41%
Path 3/Centre of cluster 3	1	2	4	3	0.98%
Path 4/Centre of cluster 4	2	1	3	4	20.25%
Path 5/Centre of cluster 5	2	1	4	3	0.36%
Path 6/Centre of cluster 6	2	4	1	3	6.73%

We have shown these clusters as circles with different sizes ($\widehat{\omega}_s$) in Figure 8-2, as the cluster 1 is the largest cluster (the most popular), picking up this extraction sequences (if we are looking for a single extraction sequences) as an outcome can cover more than 68% of the total uncertainty, and this amount would increase up to 88.5% if we add the second largest cluster, namely cluster 4.

In the real example later in this chapter, we will show that number of extraction sequences is equal to number of realisations (instead of having only six possible sequences), but we are able to partition them into a few clusters same as what has been shown in Figure 8-2 based on their similarities, and then the pick up the centre point of one of these clusters which is the most popular as a final outcome.

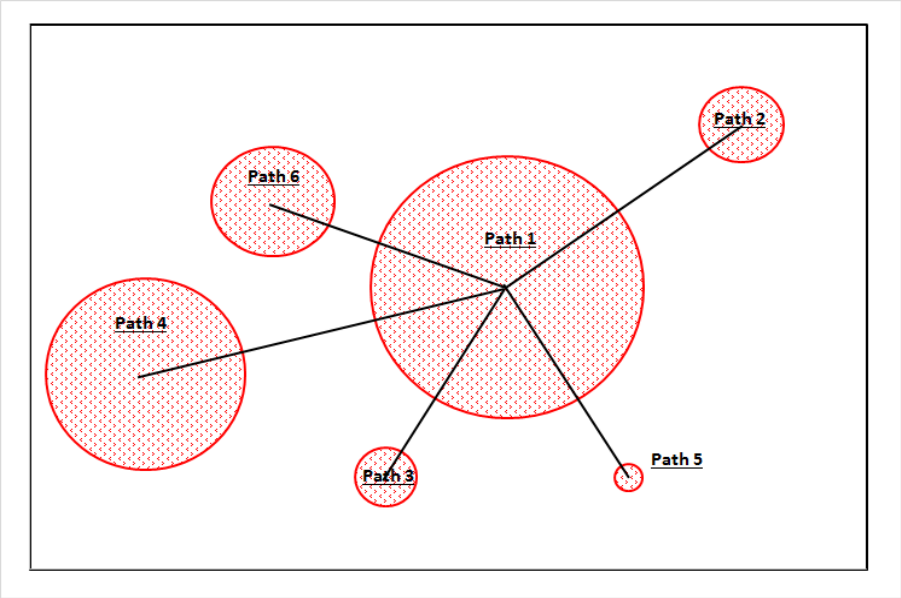


Figure 8-2: Six clusters of the fabricated example, size of the clusters are schematically equal to their weights, the centre point of the cluster 1 is the final outcome, as cluster 1 has more than half of total weight

Thus far, we know the probability of the schedules occurring for the given realisations (see Table 8-1). Now, we turn to calculate the NPV of these schedules to understand how these probabilities relate to the NPV. For calculating the NPVs, we assume the annual interest rate is 20%, value ($) of each block for one percent grade is about 20,000$, and each block takes one year to get mined.

Table 8-2 shows NPV statistics for the six representative schedules. As seen, schedule no.1 (path 1) not only has the highest probability between the six other schedules, but it also has the highest individual NPV (29,367 $) between the schedules. Therefore, we can say that the schedule no.1 is robust in terms of NPV as well.

Table 8-2: NPVs for the six representative schedules

Schedule	First block's grade	Second block's grade	Third block's grade	Fourth block's grade	First block's Value $	Second block's Value $	Third block's Value $	Fourth block's Value $	NPV $	Probability %
Path 1 (1-3-2-4)	0.60	0.70	0.50	0.40	12,000	14,000	10,000	8,000	29,367	68.27
Path 2 (1-2-3-4)	0.60	0.50	0.70	0.40	12,000	10,000	14,000	8,000	28,904	3.41
Path 3 (1-2-4-3)	0.60	0.50	0.40	0.70	12,000	10,000	8,000	14,000	28,325	0.98
Path 4 (2-1-3-4)	0.50	0.60	0.70	0.40	10,000	12,000	14,000	8,000	28,626	20.25
Path 5 (2-1-4-3)	0.50	0.60	0.40	0.70	10,000	12,000	8,000	14,000	28,047	0.36
Path 6 (2-4-1-3)	0.50	0.40	0.60	0.70	10,000	8,000	12,000	14,000	27,584	6.73

Although, in this case, the most probable schedule has the highest NPV, that is worth considering what if the NPV was not the highest for this schedule. That means, another schedule with less probability has the highest NPV.

It is clear that there is no any unique answer for this situation, and as a general rule, the more money we stand to make, the more money we stand to lose. That means, we typically need to take on more risk to achieve higher returns (NPV). As a matter of fact, there are different levels of risk attached to different types of mine schedules, and therefore we have to decide first what level of risk we are comfortable with. A mining company may select the highest probable schedule (less NPV) because the probability of low block values corresponding to this schedule is the highest, so it is highly likely to happen, namely more realistic. On the other hand, other company may choose the highest NPV (less chance) because the difference between their NPVs (less probable and high probable schedules) is high enough to outweigh the risk, and this risk is within the acceptable risk range for that company.

We will study a real case for a better comparison between risk (probability) and NPV later in section 8-5.

8-3 Definition of dissimilarity and realisation reduction

We restate here, in a more abridged way, the methodology which was more thoroughly described in chapter 6, by measuring the distance between two extraction sequences, namely type II of response variables (a metric on the space of extraction sequences). Let $s = \{s_1, \dots s_N\}$ and $r = \{r_1, \dots r_N\}$ be two extraction sequences of two mine designs where each set shows the order of digging up the individual blocks from the first scheduled block (s_1, r_1) to the last one (s_N, r_N) .

We suppose that the extraction sequences are of the same length; if the sequences are of different lengths, strategies to handle this are discussed in Remark 8-1. A simple way to define distance between s and r could be to consider them as N-vectors and compute $\sum_{i=1}^{N}|s_i - r_i|$ or $\sum_{i=1}^{N}(s_i - r_i)^2$. However, such an approach ignores the spatial locations of blocks. For example, suppose that $s = r$, except at two positions i^* and j^* where $r_{i^*} = s_{j^*}$ and $r_{j^*} = s_{i^*}$ (to create r, we take s and just swap the extraction number of two blocks in the sequence). Using the simple sums above, the 'distance' computed between s and r is completely independent of how spatially close block i^* and block j^* are. However, for a useful definition of distance, if the blocks are physically close, the sequence distance should also be close; while if the blocks are spatially distant, then the sequence distance should be larger.

The method we now propose incorporates geometric spatial information from the block model. We denote the three-dimensional coordinates of the centre of block i, by c_i, $i = 1,\ldots, N$, and set $d_{ij} = |c_i - c_i|$, the Euclidean distance between the centres of blocks i and j. To define the distance between these two sequences, in which we denote D_{rs}, we compute the minimal 'work' required to transform sequence r into sequence s. There are two issues here: the 'work' and 'transforming'. First, let's examine 'transforming': we imagine that we have a bipartite graph (Figure 8-3) with $2N$ nodes and N^2 arcs, shown below. On the left N nodes we write the values of sequence s and on the right N nodes we write the values of sequence r. We then 'transform' sequence s into sequence r by flowing 'sequence *value'* along the N^2 arcs. If $f_{ij} > 0$ denotes the flow from block i on the left to block j on the right, then mathematically we require for each $i = 1,\ldots, N$, $\sum_{j=1}^{N} f_{ij} = s_i$ (total flow leaving block i is s_i) and for each , $j = 1,\ldots, N$, $\sum_{i=1}^{N} f_{ij} = r_j$ (total flow entering block j is r_i).

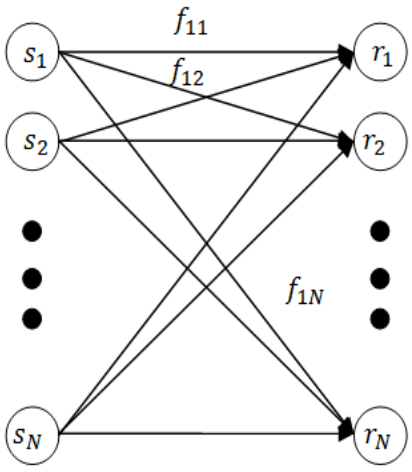


Figure 8-3: The Kantorovich process, the minimum work for transferring the left distributions to the right side

There are several different sets of flows that satisfy these $2N$ equations (we have N^2 variables and only $2N$ equations). However, we want a particular flow, one that minimises 'work'. We define effort as $\sum_{i,j=1}^{N} d_{ij}\ f_{ij}$; that is, the total effort is a weighted sum of the flows, weighted according to the spatial distance between the block pairs i and j. In summary, we define

$$\min_{f^{rs}} \quad \sum_{i,j=1}^{N} d_{ij} f_{ij}^{rs}$$

$$\text{Subject to} \quad \sum_{j=1}^{N} f_{ij}^{rs} = m_i^r, \qquad 1 \le i \le N \qquad (8\text{-}1)$$

$$\sum_{i=1}^{N} f_{ij}^{rs} = m_j^s, \qquad 1 \le j \le N \qquad (8\text{-}2)$$

$$f_{ij}^{rs} \ge 0, \qquad 1 \le i\ , j \le N \qquad (8\text{-}3)$$

We have already explained in chapter 6 how to solve this standard transportation problem using linear programs.

Remark 8-1. In the situation where sequences s and r have lengths N and M, respectively, $M - N$ is assumed without loss of generality that $M < N$. One possible approach is to 'pad' sequence r to create $r' = \left[r_{1,\dots,} r_M\ , M+1 \dots, M+1\right]$, where there are $N - M$ copies of $M + 1$ at the end of r', and then apply the methodology above to the N -sequences s and r'.

We now have a way of measuring the distance between two extraction sequences. Suppose that we are provided with a collection of extraction sequences from which we wish to, for example, extract some representative sequences for a more intensive analysis, or wish to compute which extraction sequences are '*outlier 's*. This collection of extraction sequences could arise via the individual optimisation of a collection of conditional simulations, such as in the maximum upside and minimum downside scheme. If we have $\boldsymbol{S}$ extraction sequences, we may compute z_S for the $\boldsymbol{S}\ \boldsymbol{S}(\boldsymbol{S}-1)/2$ distinct pairs of extraction sequences. We call this $\boldsymbol{S}(\boldsymbol{S}-1)/2\ \times \boldsymbol{S}(\boldsymbol{S}-1)/2$ matrix, the similarity matrix. The value z_s has the units of D_{rs}. Thus, z_S has the interpretation of 'Work' required to transform all of the $\boldsymbol{S}$ schedules into $S \ll \boldsymbol{S}$ schedules.

We may now use the method developed in chapter 6 to extract a subcollection of $S \ll \boldsymbol{S}$ extraction sequences that best represent the larger collection of $\boldsymbol{S}$ sequences. We called this method '*realisation reduction.* Indeed, the realisation reduction method is based on transporting the probability of $1/\boldsymbol{S}$ from the original large collection of $\boldsymbol{S}$ extraction sequences into a smaller collection of S extraction sequences

(representative extraction sequences); therefore, each of the representative extraction sequences $s^1, \dots, s^S$ is given the weight $\widehat{\omega}_s$, which satisfies the following properties:

$$0 < \widehat{\omega}_s \leq 1 \qquad (8-4)$$

$$\sum_{i=1}^{S} \widehat{\omega}_{s^i} = 1 \qquad (8-5)$$

We refer to chapter 6 for details. Once these S representative extraction sequences $s^1, \dots, s^S$ have been found, one can straightforwardly create a *'cluster'* of sequences around each of these representatives (cluster centres) by assigning each non-representative sequence to the representative sequence it is closest to, according to the distance D. That is, assign sequence r to the representative sequence s^{k*} if $D_{r,s^{k*}} \leq D_{r,s^k}$ for all $1 \leq k \leq S$. Because (i) each sequence is more similar, in the sense of the distance D, to sequences within the same cluster than to sequences in different clusters, and (ii) the distance D penalises spatial distance, we expect that the NPV and other type-I statistics will also be similar within a cluster. It may indeed happen that sequences in different clusters may have similar NPVs or other type-I statistics; in fact, this property demonstrates that using type-I statistics is a very poor way of determining the similarities of the sequences.

Remark 8-2. As we know, $\widehat{\omega}_s$ are proportional to the size of each cluster; therefore the weight or the probability of any cluster centre ($\widehat{\omega}_s$) varies with the size of the cluster. We can apply a similar idea to find the weight or the probability ω_r of the any chosen extraction sequences r inside the cluster $C_i, r \neq s^i$. That means we re-cluster the space by enforcing the chosen extraction sequences r as a centre of a cluster. The variability of ω_r, for this kind of selection can be $1/\boldsymbol{S} \leq \omega_r < \widehat{\omega}_{s^i}$. Now, we are able to compare ω_r against $\widehat{\omega}_s$ to calculate how much weight can be lost, if r is chosen instead of S.

Applying grade uncertainty in mine design produces physical differences in optimised pit limits, pushback designs or the scheduling patterns. For example, Figure 8-4 schematically presents one section of a mine with three different scheduling patterns (which come from three different realisations). As seen, each section has 122 blocks, which are classified into four pushbacks. By visually comparing these sections, we conclude that design 1 is closer to design 2 than design 3. Quantifying this similarity can be a powerful technique to risk assessment in mine designs.

For example, to quantify the dissimilarities between these sections, first, we apply the best case scenario which consists of mining out pushback 1 from the top down, before starting the next pushbacks. That can

generate three different optimal extraction sequences of the blocks. By calculating D_{rs} between the given three sequential series, the following optimal pairwise distances (dissimilarities) are calculated $D_{12} = 54{,}244.2$, $D_{13} = 151{,}041$, and $D_{23} = 98{,}148.9$. Schedules 1 and 2 are at a minimum distance from each other, while schedules 1 and 3 are at a maximum distance. This is what can be visually seen.

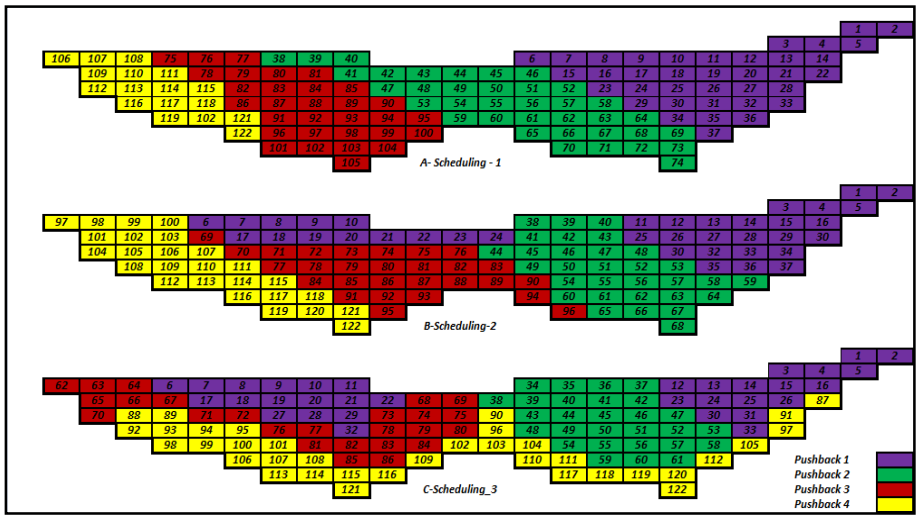


Figure 8-4: The schematic mine section of three different mine designs (based on three realisations) with different pushback designs (coloured blocks) and block sequences (number in each block)

8-4 Type II of response variables

In this section, type II of the response variables and the application of this term to quantify the dissimilarity between the different designs are briefly explained. The open pit mine planning stages are shown in Figure 8-5; as can be seen, they are divided into three different stages.

By applying optimisation and the mine design procedure over each realisation the following series of responses of type II can normally be obtained. These series could be used for measuring the dissimilarity between the designs.

1- Lerch-Grossman (LG) phases, Nested pits

2- Practical pushback sequences

3- Block extraction sequences

4- Block extraction time

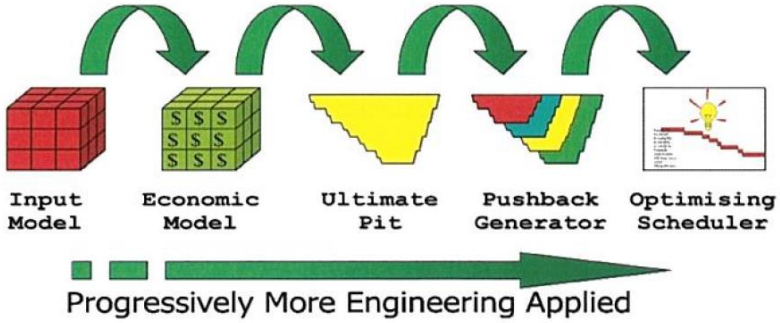


Figure 8-5: Open pit mine design stages from the geological model to mine scheduling[9]

LG phases: The starting point for mine design is to generate *LG* Phases. By stepping (for example m steps) the economic parameters, such as the revenue factor and the commodity price factor in percentage increments a series of nested pits is generated and all blocks are divided into maximum m sequential optimal pits $(1,2,\dots,m)$. By applying the *LG* phases on $\boldsymbol{S}$ realisations, $\boldsymbol{S}$ different sequences $\{(s_{1,1},\dots,s_{1,m}),(s_{2,1},\dots,s_{2,m}),\dots,(s_{\boldsymbol{S},1},\dots,s_{\boldsymbol{S},m})\}$ of these nested pits can be generated. It is clear that the similarity between the n series would be smaller if the number of steps (m) becomes smaller. Further, extremely small steps tend to give block optimal extraction sequences instead of optimal nested pits. In such a case the dissimilarity becomes much larger.

Practical pushback sequences: This is the same as the *LG* phases series, but the number of practical pushbacks (chosen by a mine designer) would be much less than m. For example, in Figure 8-5, there are four different pushbacks for each design where all blocks in pushback n.1 (purple) have the number 1, in the second pushback (green) are 2, and so on.

Block extraction sequences: Optimal Extraction Sequence of the blocks (OES) is the sequential numbers usually generated by the scheduling algorithms (or even the optimisation and pushback design algorithms) to maximise, for example, NPV. In Figure 8- 4, the numbers inside each block in each mine design show the extraction sequences of the block. In this chapter, we use these same sequences.

Block extraction time: This is almost the last step of a mine design procedure and it can generally classify the blocks into yearly scheduling by considering practical mining constraints; therefore, the blocks would be given the numbers $\{1,2,\dots,\boldsymbol{N}\}$ while $\boldsymbol{N}$ is the mine life.

Making any change in the financial and technical parameters, constraints and algorithms would alter these sequence numbers. For example, the pushback sequences can be altered by changing the number of pushbacks, pushback size and minimum distance between pushbacks. Also, the block extraction times can be altered, for instance, by bounds and target values of grade.

In addition, mine designs are progressively developed into practical designs by applying different constraints, from optimisation to mine scheduling (or even stockpiling); consequently, the similarities between the corresponding designs at different stages are not the same, although there may be some correlation between them. For example, Figure 8-6 shows a correlation between the distances (similarities) in the sequences of the pushback designs and the block extraction sequences of corresponding scheduling for 101 realisations, which were noted in section 8-5. As can be seen, they have a correlation coefficient of 51 %. Selecting the appropriate response variables highly depends on accuracy, sensitivity and the stages that we need to measure the dissimilarities. For example, in contrast to the other mentioned type II, the OES series usually contain many sequential numbers (which may range from thousands to millions), and therefore the similarity between the two OES series is usually smaller than the similarity between the corresponding pushbacks or block extraction times.

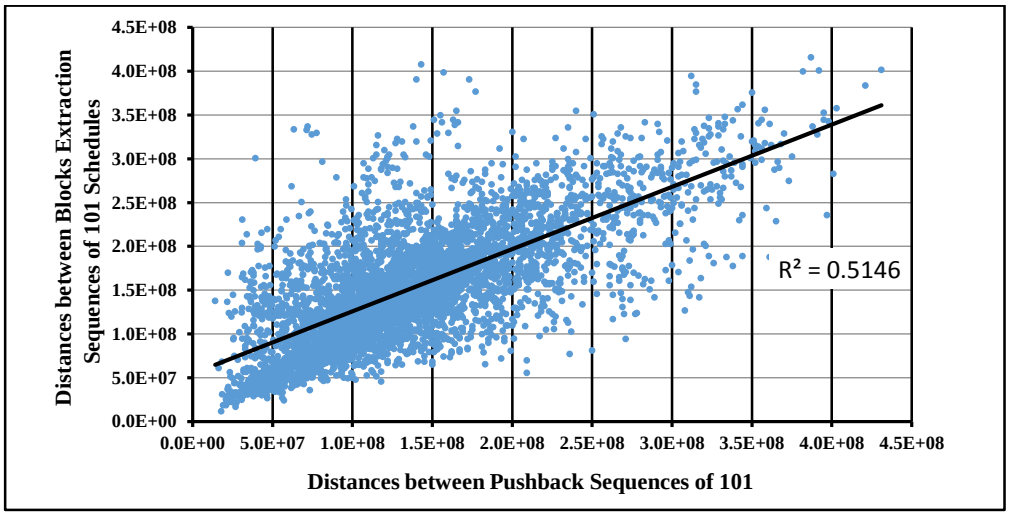


Figure 8-6: Correlation between the distances of pushback sequences and the block extraction sequences

Note that the block extraction sequences and the block extraction times behave in a similar manner to each other (see correlation coefficient ~92% in Figure 8-7). However, the block extraction times (for example, yearly) do not consider any block sequence variation within a year and give one number to all blocks which

are mined in the same year. This makes the schedules closer to each other than OES. Figure 8-7 shows a correlation between the distances in the block extraction times and the blocks extraction sequences of corresponding scheduling for the 101 realisations which were mentioned in section 8-5. As described above, the x axis range (dissimilarity between yearly block extraction times) is smaller (two times) than y axis (dissimilarity between OES).

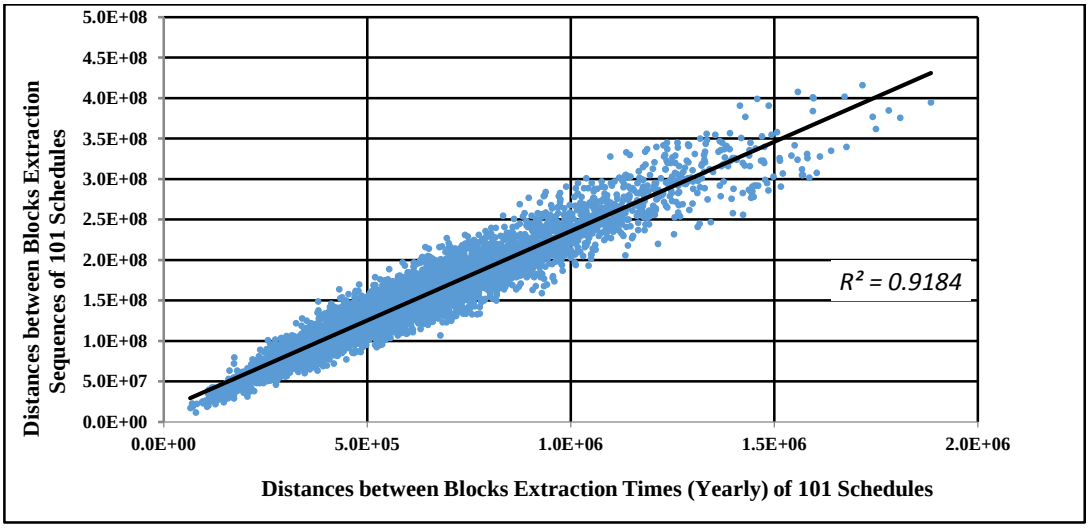


Figure 8-7: Correlation between the distances of yearly block extraction times and the block extraction sequences

8-5 Application at a copper porphyry deposit

A typical copper porphyry deposit was used with the following geological features to illustrate our methodology and also to compare it to the maximum upside and minimum downside approach for the open pit mine design procedure outlined above.

The mineralisation is hosted in the quartz-monzonite porphyry (QMP), which has undergone several phases of hydrothermal alteration common in porphyry systems. The economic mineralisation appears as small veins and disseminated grains, primarily in QMP. Mineralised zones (domains) have been classified by the degree of leaching and supergene enrichment of the original hypogene sulphide primary; in addition, high-grade economic mineralisation has occurred within the supergene zone. This confirms that this mineralisation is a classical porphyry copper style model, but on a small scale. A single estimation domain is focused where there are 48 million tonnes of measured and indicated geological resources. Within the

supergene domain, the copper grades are determined for a set of 414 six-meter composites from the drill holes. The mean copper grade and its standard deviation is 0.75% and 0.35, respectively. The block model has 2805 ore blocks with a size (dimensions) of $25 \times 25 \times 12.5$ (x, y and z direction) sizes. In order to assess the possible range of the response variable outcomes of mine designs, 101 realisations were generated by using the conditional sequential Gaussian simulation. Subsequently, applied pit optimisation generated the corresponding nested pits, pushbacks and scheduling. Each design maximises the NPV as a first priority in relation to the operational constraints and the financial and technical parameters, which are shown in Table 8-3. These mine planning activities for 101 realisations were the most time-consuming task of the study. We use NPV Scheduler Ver. 4 as an optimiser software for all mine planning activities in this study.

Table 8-3: The financial and technical parameters used for the mine design

Parameters	Value	Unit
Mining cost	1.25	$/tonne
Processing cost	20	$/tonne
Other costs	450	$/tonne
Mine production	2,500,000	tonne/year
Pit slope	38	degree
Copper recovery	85	%
Discount rate	16	%
Mining limit (capacity)	10,000,000	tonne/year
Grade Target	0.8	%
Min & Max allowed grade	0.6 , 1.0	%

8-5.1 Maximum upside and Minimum downside Approach

First, we applied the maximum upside and minimum downside approach which has been described in chapter 2, and NPV was selected as the key project indicator.

The NPVs were calculated by designing the optimal pit limits by selecting the corresponding three optimal pushbacks and finally by applying the yearly scheduling (using the parameters given in Table 8-3). The maximum upside and minimum downside of all calculated NPVs in each schedule are represented by red and blue points, respectively, in Figure 8-8. As can be seen, there are clear fluctuations between the different downsides and upsides of the designs.

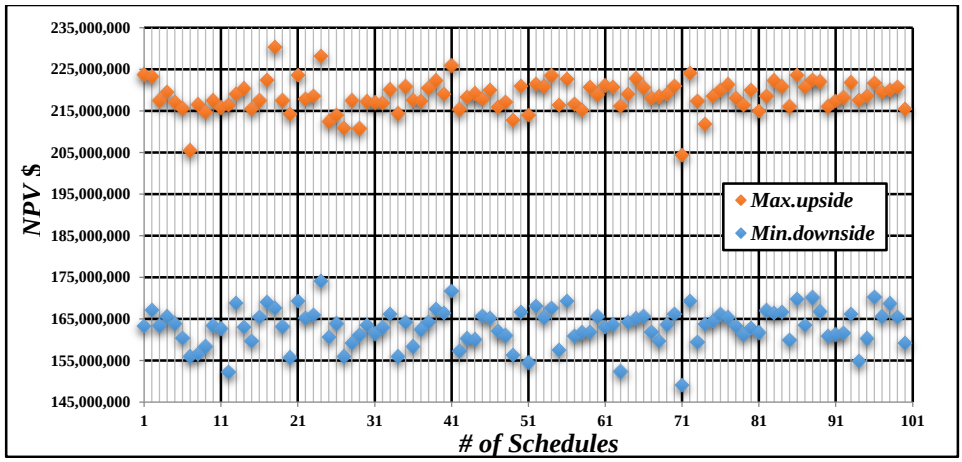


Figure 8-8: Maximum upside and minimum downside for 101 schedules

By applying an NPV of 167.3M$, as a minimum acceptable NPV (point of reference), 10 schedules are first nominated (~10 % of all schedules). Figure 8-8 shows the distribution of 101 NPVs for each nominated schedule. The amount of downside risk and the maximised reward, which was sorted by the maximum upside, is shown in Table 8-4. As can be seen in Figure 8-8, schedule no. 24 is the best in minimising downside risk and maximising reward, as 5.866M$ and 59.935 M$, respectively. The second best is 41, which is close to 24 when compared to the others. In the next section we will be able to see that these two schedules are far away from each other in the point of block extraction sequence.

As can be seen in Figure 8-9, some designs, for example, designs no. 88, 96 or 86, 56, 21 and 72, are significantly close to each other; therefore, this approach based on Type I information is unable to identify any differences between them.

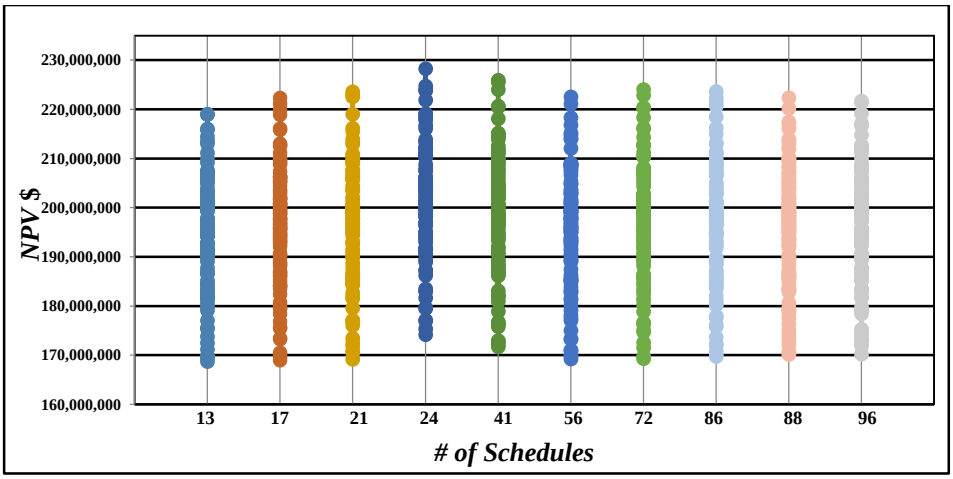


Figure 8-9: Upside and downside NPV for only 10 selected schedules

Table 8-4: Maximum upside and minimum downside for the selected schedules sorted by maximum upside

Schedule	*Min NPV $*	*Max NPV $*	*Min.downside $**	*Max.upside$**
24	174,166,232	228,235,057	5,866,232	59,935,057
41	171,734,730	225,901,357	3,434,730	57,601,357
72	169,301,915	224,010,164	1,001,915	55,710,164
86	169,737,535	223,665,355	1,437,535	55,365,355
21	169,175,104	223,581,882	875,104	55,281,882
56	169,258,947	222,556,767	958,947	54,256,767
17	168,947,698	222,356,713	647,698	54,056,713
88	170,156,877	222,316,061	1,856,877	54,016,061
96	170,181,230	221,699,368	1,881,230	53,399,368
13	168,731,643	219,022,427	431,643	50,722,427

***Please refer to chapter 2 to see how to calculate of the minimum downside and the maximum upside*

8-5.2 Distance based method approach

After illustrating some of the criteria of the risk based methods on the selection of the best design(s), we now turn to the distance based method. In section 8-3, the Kantorovich metric was presented as a robust approach to clustering and selecting representative scenarios and now we apply the approach to a realistic set of 101 schedules of the porphyry copper deposit.

Remark 8-3. In this study, we use a multidimensional scaling (MDS) algorithm, as described in chapter 5, to provide a visual representation of the space of uncertainty to get a better sense about the space, the pattern of similarity, possible clusters and their centres. Although applying MDS algorithm may take the advantage of using the Euclidean-based subsampling or clustering techniques, in this study this algorithm is only used for visualisation purposes to better explain the differences between the approaches. The optimal subsamples and clusters are based on the original distances D_{rs}, and not approximations.

8-5.2.1 Computing the distance between the schedules

Creating the space of uncertainty, the dissimilarity distance matrix can be constructed by computing the pairwise distances D_{rs} $1 < r \leq s \leq 101$ between the optimal extraction sequences of 101 schedules using equations (1)-(4), and thus a 5050×5050 dissimilarity distance matrix is constructed.

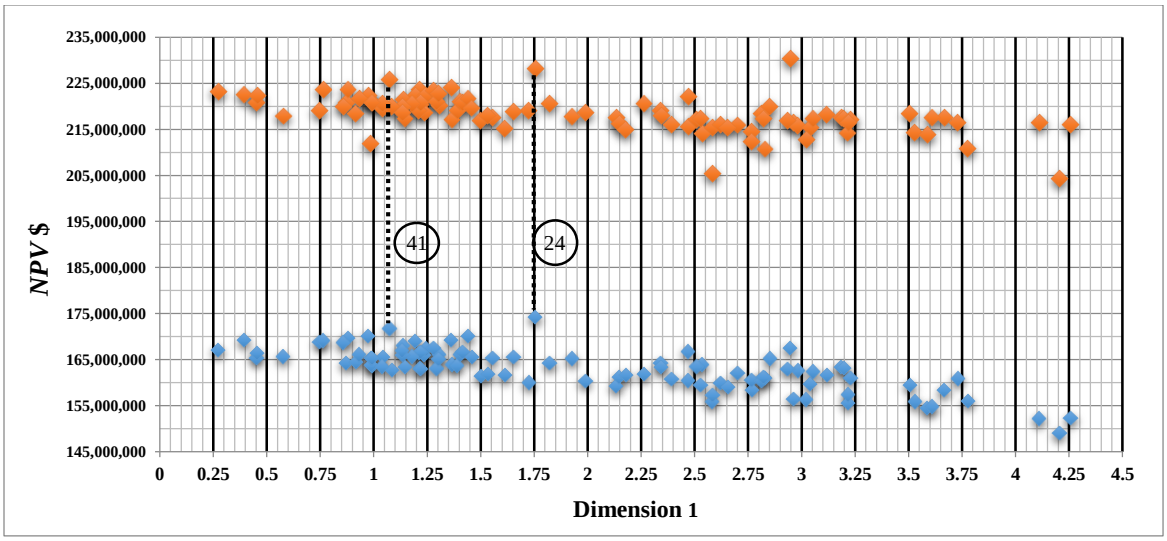


Figure 8-10: Upside and downside NPV for 101 selected schedules projected (MDS) on Euclidean space $\boldsymbol{R^1}$ ($\boldsymbol{x}$ *axis numbers in this Figures are in form of scientific format* $\times \mathbf{10^8}$ *, but for sake of brevity they are shown without* $\mathbf{10^8}$)

Figure 8-10 shows a multidimensional scaling (MDS) embedding of the given dissimilarity distance matrix in R^1, so that the interpoint distances in R^1 approximate the pairwise schedule distances D_{rs}. R^1 is used here to compare the results of two approaches, namely a comparison between Figures 8-8 and 8-10. As seen, the two graphs have the same y axis (NPV), but a different x axis. x axis in Figure 8-10 has a physical meaning and shows the projected Euclidean space coordinate, that is, it indicates how close (similar or dissimilar) the designs are to each other. Figure 8-10, representing the distance based approach, can present the combination of type I and type II response variables which give more complete information about a mine design. For example, as can easily be seen, the extraction sequence of schedule 24 is around 20% of range far from 41, while their NPVs are not so far (see Table 8-4).

8-5.2.2 Clustering and selecting representative schedules

The results of any type of clustering algorithm are the collections (clusters) where the data points inside each cluster are more similar than the data points between the clusters. We use our methodology (refer the reader to section 3 for details) to optimally subsample a large collection of schedules and to quantify how well this high-quality subsample represents the overall uncertainty of the collection. The basic issue of the subcollection selection is to identify the best number of subsets or collections in the given space. Following Dupacova et al. (2003), we report z_S relative to z_1, the latter representing the 'base' distance between the best deterministic approximation of the collection of $\boldsymbol{S}$ realisations. Thus, $z_S = z_1$ represents the distance

between the optimally subsampled S realisations and the full set of $\boldsymbol{S}$ realisations relative to the distance between the best single realisation (deterministic approximation) and the full set of S realisations. For brevity, we call the quantity $I_S = (z_i - z_{i+1})/z_1 \times 100\%$ and the relative improvement of the optimal S-subcollection. Clearly $I_1 = 0$ (S = 1 provides no relative accuracy), and $I_S = 100$ ($S = \boldsymbol{S}$ provides a 100% relative accuracy).

The number of collections has to be chosen in such a way that adding another collection does not create a better relative improvement in the data set. As seen in Figure 8-11, there is no significant change in the relative improvement after $S = 3$. The number of collections chosen can therefore be 3; consequently, depending on how many members the collections have, the weight of the collection centres ($\hat{\omega}_s$) can be determined. The number of collections in our approach indicates how much spatial grade variability can generally make a difference in mine scheduling; moreover, the weight of each collection centre clearly shows how much spatial grade variability does not significantly impact on mine scheduling. The weight of collections is also a key factor to determine the sample sizes proportion if, for any reason, using as an example what has been addressed in Leite and Dimitrakopoulos (2007), a limited number of designs have to be selected for futher processes.

Table 8-5 illustrates the best representative schedules (collections centres) of each collection and their weights. As seen, schedule 41, with 47.52 % weight ($\hat{\omega}_{41} = 0.4752$) has the highest weight, while two other collection centers have appromimatly the same weight ($\hat{\omega}_3 = 0.2673,\ \hat{\omega}_{73} = 0.2574$). This means that more than 47% of $\boldsymbol{S}$ schedules are more similar to schedule 41 compared to the other schedules. If we assume that the reality is somewhere in this space, it has a 47% chance of being similar to schedule 41 or inside the cluster of no.1. Therefore, this cluster can be the first selection for futher processing.

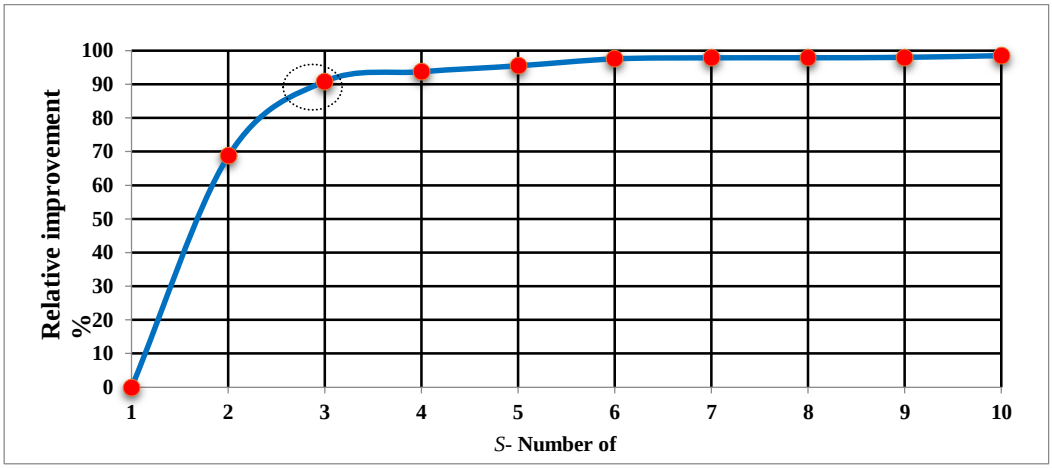


Figure 8-11: Number of collections $\boldsymbol{S}$ vs. relative improvement $\boldsymbol{I_S}$, with three collections as the optimum

Table 8-5: Representative schedules and their weights for the three collections

Collections No.	*Representative schedule*	*Weight* $(\widehat{\omega}_s)$ *%*
1	41	47.52
2	3	26.73
3	73	25.74

An MDS embedding of the all the points in R^2 (for better visualisation, $n = 2$ is chosen) can be seen in Figure 8-12. The schedules (all points) are clustered rather than being randomly distributed in the space. The result of clustering is shown as the colouring of the points into three collections and the coloured circles are the centres of each collection. As seen in Figure 8-12, schedule 24, which is the best selection of maximum upside and minimum downside approach, is in cluster no.1.

Based on what was explained in Remark 8-2, we now compare ω_{24} against $\widehat{\omega}_{41}$ to show how much weight or probability can be lost by picking up schedule 24 instead of 41 in Remark 8-2. This weight or probability (ω_{24}) is around 8.9 %, which is 4.33 times less than $\widehat{\omega}_{41}$. The main reason for this significant difference can be explained by distance D, schedule 24, in cluster no.1; it is not only far from the cluster centre point (schedule 41), but also from the other schedules, namely an outlier in cluster no.1.

Remark 8-3. Applying outlier detection techniques after clustering the data set can be helpful to have better clusters although we do not apply this technique in this chapter. For the interested reader, further background on outlier detection in clusters may be found in the reference Han et al. (2011). In the distance-based approach, schedule X in cluster Y can be an outlier (see chapter 3) if no more than n schedules in the cluster can be found at a neighbourhood radius r or less from X; that is, an outlier is a schedule(s) that is far away from the others in the point of Kantorovich distance (they are dissimilar to other schedules although they are in the same cluster).

8-5.3 Statistical analysis of type I for type II

After clustering the schedules and calculating weights or probability of the chosen schedules, we now turn to the space of uncertainty of type I (in this case NPV) to assess the NPVs of the representative schedules, also including that of the clusters. This space of uncertainty of type I is usually represented by the histogram of response values. In this case, for given NPVs in section 5.1, for each collection, two histograms (maximum upside and minimum downside) of schedules are drawn (Figure 8-13).

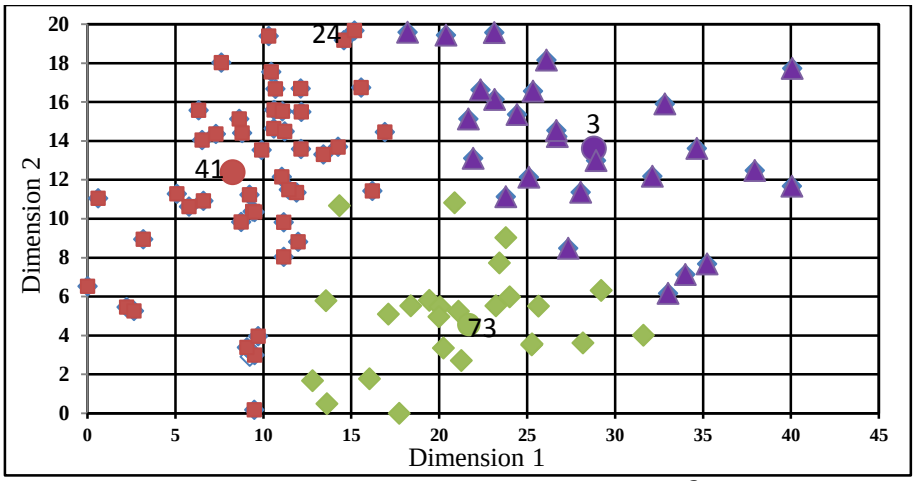


Figure 8-12: Multidimensional scaling embedding of 101 schedules in $\boldsymbol{R^2}$ classified into three clusters

Tables 8-6 and 8-7 show NPV statistics of the histogram of each cluster and the maximum upside and the minimum downside for the three representative schedules, respectively. As seen, cluster no.1 not only has the highest average NPVs and the lowest variances between the three other clusters, but it also has the highest individual NPV (schedule 41) between the other three representative schedules. That is, the tendency of individual schedules within the collection is to have approximately a similar NPV; consequently, the schedules in collection no.1 have a higher NPV and they achieve maximum upside and minimum downside. It is not unexpected that all 10 nominated schedules of the maximum upside and minimum downside approach already be in collection no.1. Therefore, the representative schedule is robust in terms of NPV as well.

Table 8-6: Mean, St. Deviation, Minimum, Maximum, of the maximum upside and minimum downside histograms for the clusters

	Maximum upside (NPV $)				*Minimum downside (NPV $)*			
Collection	*Mean*	*St. Deviation*	*Min.*	*Max.*	*Mean*	*St. Deviation*	*Min.*	*Max.*
1	220,494,007	2.74.E+06	211,908,550	228,235,057	166,299,432	2.62.E+06	161,450,217	174,166,232
2	215,921,337	3.38.E+06	204,322,242	222,075,503	159,758,897	4.28.E+06	149,049,042	166,750,458
3	217,397,927	4.30.E+06	205,397,132	230,303,072	160,627,237	3.15.E+06	155,636,339	167,583,273

Table 8-7: Maximum upside and minimum downside for the three representative schedules

Collections No.	*Representative Schedules*	*Weight ($\widehat{\omega}_s$) %*	*Min NPV $*	*Max NPV $*	*Min.downside $*	*Max.upside $*
1	41	47.52	171,734,730	225,901,357	3,434,730	57,601,357
2	3	26.73	163,348,799	217,537,810	- 4,951,201	49,237,810
3	73	25.74	159,462,138	217,287,580	-8,837,862	48,987,580

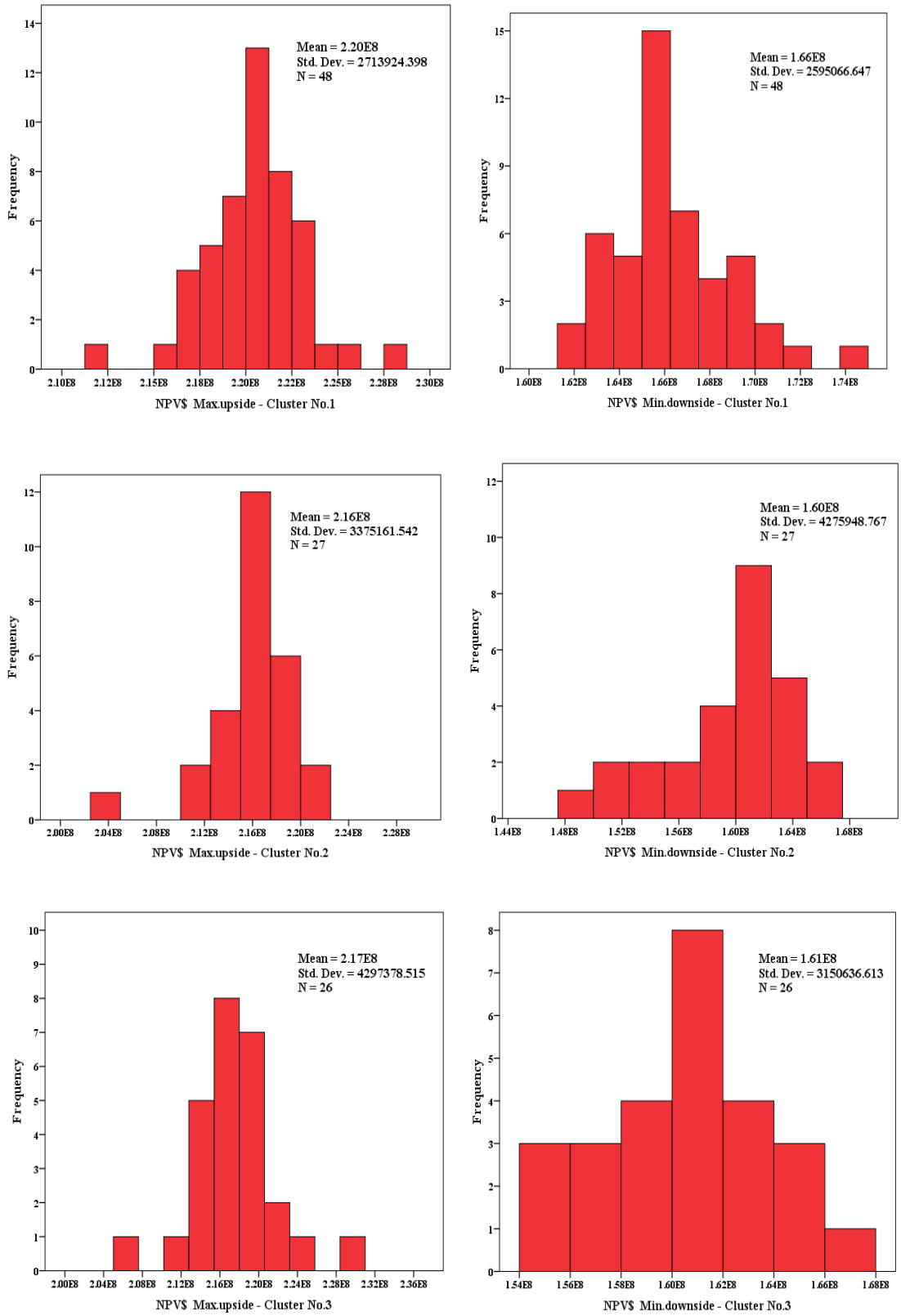


Figure 8-13: Six histograms for the NPVs of maximum upside (left sides) and minimum downside (right sides) for three clusters

If we were to choose a single cluster ($S = 1$), the representative schedule would be the most 'practically robust' in terms of meeting production targets compared to the other schedules. By increasing the number of clusters, we have more representative schedules; while each representative schedule is practically robust with respect to other realisations in its cluster, its robustness with respect to realisations outside the cluster may be poor. In an extreme case where each cluster consists of a single schedule, we have lost all the aspects of practical robustness that can be obtained via our methodology. While using a single cluster provides high practical robustness, there is no guarantee that this single representative schedule performs well in terms of type-I statistics, such as NPV-based statistics. For modest numbers of clusters, one may inspect the type-I statistics of each cluster's representative schedule and balance this against the practical robustness of the schedule, indicated by the distance of the representative schedule from all other schedules. For example, in our case study, the representative schedule 41 is relatively close to many other schedules, such as those in its cluster, but also far from some schedules in the other two clusters. Fortunately, the type-I (in this example, NPV-based) statistics of schedule 41 are superior to those of the other two representative schedules 3 and 73, so one might choose schedule 41 as a good balance of practical robustness and suggested NPV performance.

The other risk-based methods of schedule selection, including maximum upside and minimum downside, effectively consider the extreme situation where each schedule is in its own cluster and select a schedule purely on the basis of type-I (specifically NPV) statistics. This ignores the practical robustness of the selected schedule and the NPV obtained in reality may be far from the predicted NPV. Our methodology thus generalises these other risk-based methods by incorporating practical robustness into the schedule selection procedure and enables a balancing of practical robustness with type-I statistics.

8-6 Conclusion

Risk based mine design methods usually generate different mine designs based on simulated realisations; consequently, these methods deal with many mine designs so that just one of them can be, in some way, selected. Finding the best criteria for this selection can be challenging for any approach. One of the drawbacks of exiting approaches is the lack of a general classification method, namely, the clustering of the mine designs.

In this chapter, we proposed a new methodology for mapping the space of uncertainty by a distance function that is based upon a physically meaningful notion of dissimilarity between pairs of mine designs, using the

Kantorovich metric. Regarding this methodology, we are able to quantify the dissimilarity of different mine designs and use this information to quantify how well the representatives or clusters depict the overall uncertainty. In order to select the subcollection of mine design that gives the best representative design, we have again used the concept of Kantorovich distance and developed a simple optimisation model for the best samples. To choose a single representative from the total list of representatives, we suggested computing NPV statistics for schedules within each cluster and chose the representative for the cluster with the 'best' cluster statistics.

This approach relies on a reasonable clustering inside the space of uncertainty of Type II; consequently, any method(s), which can find possible spatial patterns (in this space) would get better results.

Chapter 9

9 Conclusions and future work

9-1 Overview

The application of geostatistical conditional simulations, which normally requires large numbers of generated realisations has increased enormously during the past two decades. As previously explained, there is no parameter that can provide further information about high order statistics for generated realisations; therefore, it is concluded that two realisations can be significantly different in ways that cannot be captured by descriptive geostatistics.

The main objective of this study was to model the uncertainty in the earth science; or, precisely, to quantify geological uncertainty and mine planning risk using distance based techniques. For this objective, the following steps were taken to address this issue, in the following order:

- selecting the distance function (Kantorovich distance) to explain what causes the dissimilarity between realisations;
- computing the distance between pairs of realisations for measuring the dissimilarity between them;
- optimally subsampling a large collection of realisations and quantifying how well this high-quality subsample represents the overall uncertainty of the collection;
- and presenting the applications of this approach to address controversial issues in the earth science.

We have developed a practical quantitative methodology to define and map the space of uncertainty by computing Kantorovich distance between generated realisations. This method provides a consistent and repeatable mathematical framework for modelling uncertainty, not only for geological modelling but also for mine planning. This distance approach is able to structure relations between generated realisations into the metric space. Quantifying this spatial structure, namely measuring dissimilarity, can be a powerful technique to assess and map the space of uncertainty. In addition, this space can be visualised (embedding into the Euclidean space) by using the multi-dimensional scaling algorithm.

The mentioned sub-sampling technique to take the best representatives of the space of uncertainty following the concept of Kantorovich distance provides an efficient strategy to reduce the number of realisations inside the space of uncertainty.

9-2 Summary of the study

The main conclusions observed in this thesis can be summarised, as follows:

- The notion of dissimilarity or 'distance', namely Kantorovich distance between realisations, is the key ingredient in constructing a formal measure of dissimilarity for generated realisations and for quantifying the space of uncertainty. This distance is geologically meaningful and is able to explain what causes the dissimilarity between realisations. Moreover, this distance is able to provide a structure to a collection of generated realisations establishing a meaningful relation between them.

- Multi-Dimensional Scaling (MDS) provides a Euclidean representation (R^n) of the space of uncertainty, which makes it (dissimilarity distance matrix) possible to visualise it in 2D or 3D space.

  Moreover, the approximate location of the Kriging model, the E-Type and the real model in the space and the other realisations can be easily visualised. Although this technique is able to reveal the structures of the space of uncertainty (if the stress factor is low), it has been used in order to enable the visualisation of the data.

- Determining the set of representative samples M is a challenging problem even in traditional clustering methods. In most clustering methods M has to be assumed as a known parameter, which should be chosen by users .We used our methodology to optimally subsample (M) a large collection of realisations (N) and quantify how well this high-quality subsample (N) represents the overall uncertainty of the collection. For example, our methodology can determine the smallest number of realisations that are required to cover 80% of the total geological uncertainty.

- If realisations (points) are clustered rather than randomly distributed in the space, the approximate MDS representation of the realisations in R^2 opens the possibility of using Euclidean-based subsampling or clustering techniques, such as K-means, to select S cluster centres. These clustering techniques attempt to minimise the total distance from centerpoints to other points. These cluster

centres might be used as the selected subsamples. However, the realisations are not usually clustered.

- The effects of special continuities (variogram ranges) and variability (variogram sills) have been investigated on the space of uncertainty. As has been observed, decreasing the range of variogram leads to a reduction in the relative accuracy; consequently, longer ranges of correlations (greater spatial continuity) will tend to make the collection of $\boldsymbol{S}$ realisations more structured and less random. Simulations with a smaller range will require a larger number of subsamples to achieve the same relative accuracy.

- Increasing the sill or decreasing the range leads to a reduction in the relative accuracy of the sampling method. These results are exactly what one would expect; longer ranges of correlations (greater spatial continuity) and lower sills (lower variability) will tend to make the collection of S realisations more structured and less random. This greater structure, encoded via the distances Drs, can be exploited by our optimal subsampling procedure. In other words, simulations with a larger sill and smaller range will require a larger number of subsamples to achieve the same relative accuracy.

- The comparison of distances zS between the optimal subsamples and randomly selected subsamples revealed that the variability of random sampling can be very high. That means that generating N realisations and sampling the best M ($M << N$) would achieve better results (a better relative accuracy) rather than simply generating the first M realisations of the N without pursuing it further. Therefore, it is worth investing more time to create a larger number of realisations to attempt to better represent the true resource of uncertainty and to then subsample as best possible from that larger collection of realisations.

- As was previously described and shown (the MDS's figures), in all types of simulation algorithms (SGS, TBS, SIS) and the two mentioned cases (2D and 3D), the distances D_{rs} for r = Kriged or r = E-type and $s = 1; \dots; S$ are small on average when compared with the overall average of the distances. That means that the smooth models are very close to each other rather than the other realisations.

- The sub-sampling approach normally makes the unequal weight distribution for selected realisations and the weights given to the selected realisations (by the sub-sampling approach) tend to be higher nearer to the centre of the space of uncertainty (the MDS's figures), while those

selected realisations near the periphery tend to have lower weights. These peripheral selections are effectively chosen to represent themselves only, while the selected realisations nearer to the centre of the figure represent not only themselves but those realisations nearby, which are effectively absorbing the weights from these nearby realisations.

- The condition simulation algorithms are supposed to generate equally probable realisations (fairly representing the entire space of uncertainty). However, we could verify in this study that there was a larger probability to generate realisations in the neighbourhood of the Kriging or the E-type model rather than the peripheral of the space. This means that there is a systematic tendency to configure the space of uncertainty in such a way that the probability to find or generate the realisations decreases from the centre to the edge of space. Therefore, if the realisations are randomly selected, those realisations which are around the Kriging or the E-type model have a higher chance to be collected rather than the others.

- The interpoint distances histograms (in the space of uncertainty) of all conditional simulation algorithms (SGS, TBS and SIS) follow lognormal distributions. The lognormarlity of this interpoint distance distributions can be easily observed even between a few generated realisations and would not be changed by increasing the number of realisations.

- The means and standard deviations of the interpoint distances distributions (for all conditional simulation algorithms) are stabilised by increasing the number of realisations; thus, it can be concluded with confidence that the simulation algorithms generate retaliations in such a way that the dissimilarly between them (interpoint distances), or more precisely, the structure of the space of uncertainty, ultimately follows one lognormal distribution (for each algorithm).

- We cannot mathematically explain why the similarity or interpoint distance distributions in the space of uncertainty tend to be approximately lognormally distributed. However, the main reason may be that all the lognormal algorithms are conditional to sampled points, which would not allow for the generation of realisations with a high number of differences (dissimilarity). That is, the generated realisation tend to be similar (close) to each other rather than far, causing the distributions to be positively skewed.

- By increasing the number of realisations, the average number of interpoint distances fluctuates first before remaining constant (showing very insignificant fluctuations around a fixed number). It is

noteworthy that adding more points (realisations) to the metric space would not change the average distance between points.

- By determining maximum possible dissimilarities in the space of uncertainty, we tried to answer the question whether increasing the number of realisations could make any difference to the maximum interpoint distances. The maximum interpoint distance displays the size (how big) of the space of uncertainty. As illustrated, maximum distances are approximately stabilised by increasing the number of realisations. That is, the simulation algorithm cannot generate the realisations, which are very dissimilar to each other; therefore, the size of the space remains unchanged.

- By determining the minimum possible dissimilarities in the space of uncertainty, we tried to answer the question whether increasing the number of realisations could make any difference to the minimum interpoint distances. The minimum interpoint shows how close the generated realisations can be in the space. As illustrated, minimum distances are approximately stabilised by increasing the number of realisations. That is, the simulation algorithm cannot generate the realisations, which are very similar to each other. Thus, there is a distance (neighbourhood radius $r \leq D_{Min}$) around each retaliation, where there no any realisations.

- The realisations which have larger differences (maximum distance) to the other ones are usually located on the edges of the space of uncertainty (peripheral of the space). These realisations can be classified as extreme points in the simulation process. These realisations make distributions of positively skewed interpoint distances.

- We used two methods to measure the density of the space, both of them confirming that the density of the points rises inside the space (because the space is not extended) by increasing the number of realisations. The point density (number of points around the smoothed models) is much higher than other points.

- We used two fixed points (for all simulation algorithms), namely the Kriging and the real model, to check any possible change in the location of points with respect to these two fixed points, by increasing the number of realisations.

- The variation of the average distance and its variance from these fixed points, for example Kriging, versus the number of realisations show that both of them (the average of distance and its variance)

are stabilised by increasing the number of realisations. However, they fluctuate before getting stabilised. That means that, generating more realisations does not make significant changes in the structure of the space of uncertainty; in addition, distance histograms from a fixed point (Kriging) would be unchanged.

- We have clearly shown that for all three conditional simulation algorithms the real model is closer to the edge of the spaces of uncertainty rather than the centres. This may not be a good signal for the simulation algorithms which fail to generate close realisations to the real model. Being at the edge of the pace of uncertainty means the model is quite dissimilar to the other models. As previously explained, even by increasing the number of realisations the chance of generating the realisation close to the real model is very low. As a result, if the efficiency of the simulation algorithms or generated realisations is linked to how well they can emulate the real model, it can be said that in the mentioned case (Walker Lake case, exhaustive data set) it shows very low efficiency.

- The next item demonstrates the impact of the number of realisations on the relative accuracy (sub-sampling). It was explained that after generating a certain number of realisations, the differences between the relative accuracy of the sets become insignificant. That means that the distance between the best single realisation (z_1) and the set of collected S realisations does not change by increasing the number of realisations.

- There is no significant difference in statistics (such as standard deviations, variances and minimum and maximum) of the interpoint distance distributions among the simulation algorithms (see Table 7-15). However, SGS shows a higher variation (in interpoint distances) than the other algorithms while SIS shows the lowest. This means that the ability to generate different realisations (that are not similar to each other) is the highest for SGS and in the lowest for SIS. That can be confirmed again by comparing the minimum and the maximum interpoint distances of the simulation algorithms.

- An alternative approach to compare between the spaces of uncertainty created by simulation algorithms is to check the fixed points (the Kriging and the real models). It has been confirmed that the distance between the Kriging and the real models remains constant in all the space of uncertainty. That gives us an option to insert the 2D MDS maps together to compare the spaces. As can be seen in all figures (2D MDS maps) the embedded points into R^2 show almost the same

spaces. Moreover, the average distances (original distances) between the generated realisations (by the all simulation algorithms) and these fixed points are close to each other (see Table 7-15). We believe that these results confirm that the different simulation algorithms (SGS, TBS and SIS) generate similar spaces.

- Although the quantifying of the space of uncertainty of all kinds of stochastic simulations and the sub-sampling of realisations have been well developed and addressed in this thesis, the nonlinearity of the transfer functions (such as pit optimisation, mine planning and mine scheduling) is still a big challenge when applying these transfer functions on the representative realisations. The structure of the space of uncertainty (interpoint distances) and their representatives would change by applying any nonlinear functions; consequently, the representative realisations would not be valid after that.

- As mentioned above, the output results after feeding the realisations into transfer functions is called 'response' and the space these results make is called 'the space of uncertainty of responses'. We differentiated between these two types of response to obtain a better assessment of uncertainty about the response values in this thesis. The first results are response data that can be represented by number(s) (type I), which are usually represented by the histogram of response values. The second type of response can be represented by sets/series of numbers (or even objects). For example, type II can be a series of optimal extraction sequences (OES) of blocks, or even a set of spatially connected blocks with a pit geometry, such as a series of nested pits. Measuring the distance between type II of response variables (for example, Optimal Extraction Sequences) can define a new metric space, which we called 'the space of uncertainty of responses'.

- By applying optimisation and the mine design procedure over each realisation the following series of responses of type II were obtained: Lerch-Grossman phases (Nested pits), practical pushback sequences, block extraction sequences and block extraction time. These series were used for measuring the dissimilarity between the designs. We applied the methodology on the pairs of extraction sequences of these series.

- We explained and illustrated that some these series mentioned above (nested pits, pushbacks, and OES) are more similar to each other than others. That means that some of these series are highly likely to occur, while the others would be less likely. Therefore, the assumption that chosen

design(s) are all equiprobable would not be correct and this assumption may cause misleading results in inaccurate design selection.

- We illustrated the capabilities of this methodology on mine planning or scheduling and showed how this approach can be robust for selecting the best scenario (schedule) among given schedules. To confirm the competency of the proposed methodology, we applied this method and also the maximum upside and minimum downside approach (as another risk-based method) on a typical copper porphyry deposit to make a comparison between the results (the chosen schedules) of these approaches. As the result showed, the probability of happening the selected mine schedules by our approach is significantly higher than the maximum upside and minimum downside method.

9.3 Future Work

In this section, a few ideas on the distance-based method framework are described. These ideas focus on limitations of the present study and avenues for future research. We believe that the application of the distance-based method in the mining industry is quite new and has not received much attention in earlier research, while the capabilities of this method to handle various complicated issues (in this thesis grade uncertainty) is remarkable and can present a promising performance and easy implementation.

Seeking for a better Metric

Selecting a meaningful distance function in the entire range of distance-based methods is an important stage. In this study we presented a formal measure of dissimilarity for generated realisations by adapting the Kantorovich metric which is the physically meaningful notion of dissimilarity between pairs of realisations to the geostatistics context. However, there is not a single distance measure that is able to capture or reveal the best measure of dissimilarity between pairwise data. Therefore, it is worth seeking a better dissimilarity distance function that is able to discriminate more aspects of dissimilarity between pairwise data, namely realisations.

Furthermore, the applied Kantorovich distance (in this thesis) is a complex and sophisticate metric. That has a complexity of at least $O(N^2)$ (N is the size of the data), and an optimisation problem has to be solved to measure the distance between each pair of realisations. Therefore, one of the limitation of this metric is the great number of resource consumption (CPU and memory) which is required to solve the optimisation problems. Thus, these approaches quickly become cumbersome when dealing with a large amount of data.

Moreover, this distance functions should provide a robust measure between pairwise data in such a way that the structure of the space of uncertainty (made by interpoint distances) can be effectively clustered in order to sample the best representative data points. We have to mention here that no matter how robust clustering of algorithms is, it has to be consistent with the distance function that is applied for measuring the dissimilarity.

Take advantage of Multidimensional scaling

Although, in this study, we advocate MDS for visualisation purposes only, and all calculation on the dissimilarity matrix was based on original distances and not on approximation; developing a robust MDS techniques that is able to embed the points from a metric space into the Euclidean space (R^n) in such a way that the inter-point distances after embedding are very close to what they were could take advantage of the readily-available lower-dimensional Euclidean space. That means, when the embedded points in the R^n space possess coordinates, we are able to use all Euclidean space properties such as different clustering techniques (for instance k-means clustering method), visualise all data (if $n \leq 3$) measuring the point density and having points distribution in space.

APPENDIX A

Numerical results on geostatistical simulations

A.1 Sequential Gaussian Simulation algorithm (SGS and 3D Case)

Figure A-1 shows the result of the interpoint distance calculations for 10 realisations and the Kriging model. As can be seen, the best fitted distribution is lognormal (positively skewed). The lognormal distribution is the best fit. The parameters of the histogram are shown in Table A-1.

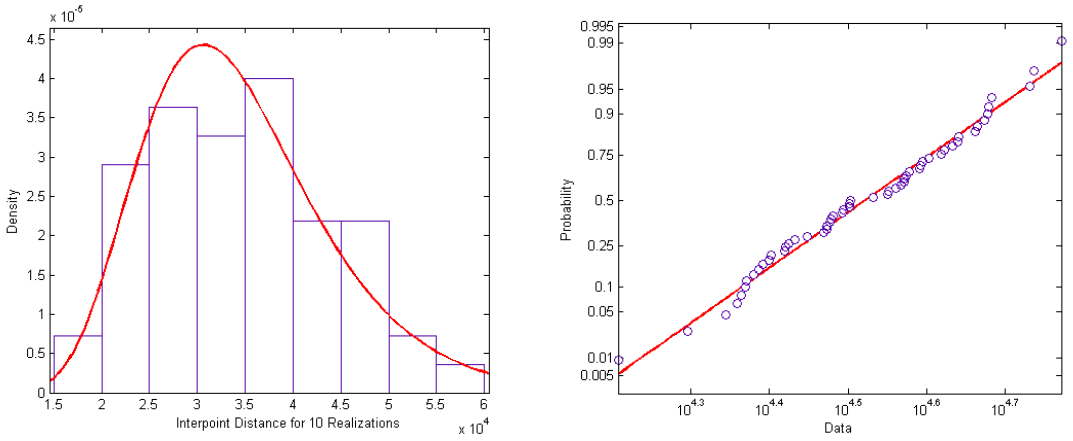


Figure A-1: The graph on the left is the histogram of 55 interpoint distances of 10 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, showing a reasonable fitting between them (SGS and 3D case)

Table A-1: Statistical parameters of the histogram of the interpoint distances of 10 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
55	34,478.66	9,652.89	93,178,264	16,099.40	59,168.50	43,069.1	0.47	25,361.7

Figure A-2 shows multidimensional scaling embedding of 10 realisations. As illustrated, the Kriging point is at the minimum distance from the others

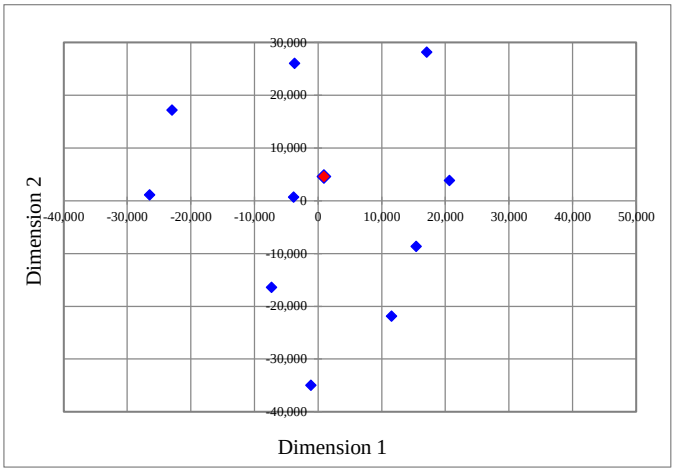


Figure A-2: Multidimensional scaling embedding of 10 realisations (blue diamonds) and the Kriging model (red diamond) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

Figure A-3 shows the result of the interpoint distance calculations for 50 realisations and Kriging model; the best fit distribution is still lognormal. The parameters of the histogram are shown in Table A-2. Figure A-4 shows multidimensional scaling embedding the realisations. The Kriging point is at the minimum distance from the others.

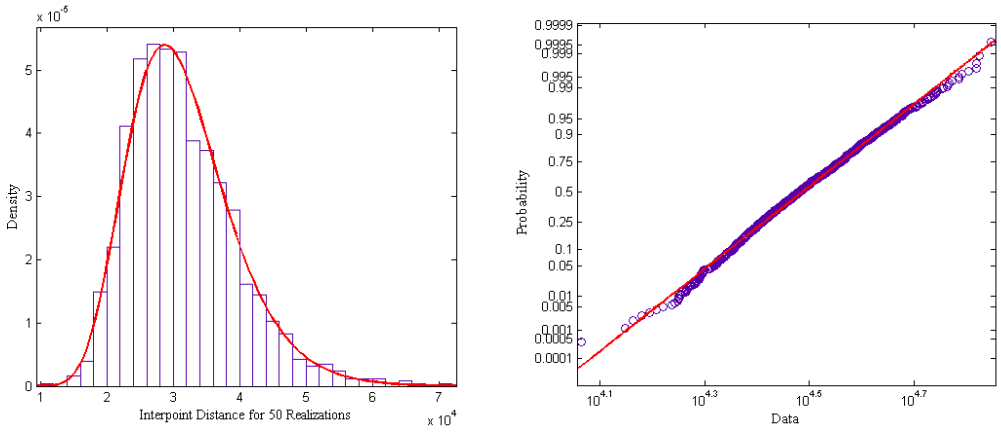


Figure A-3: The graph on the left is the histogram of 1,275 interpoint distances of 50 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, showing a reasonable fitting between them (SGS and 3D case)

Table A-2: Statistical parameters of the histogram of the interpoint distances of 50 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
1,275	31,520.48	8,166.98	66,699,619	11,582.40	59,168.30	58,729.3	0.97	22,473.9

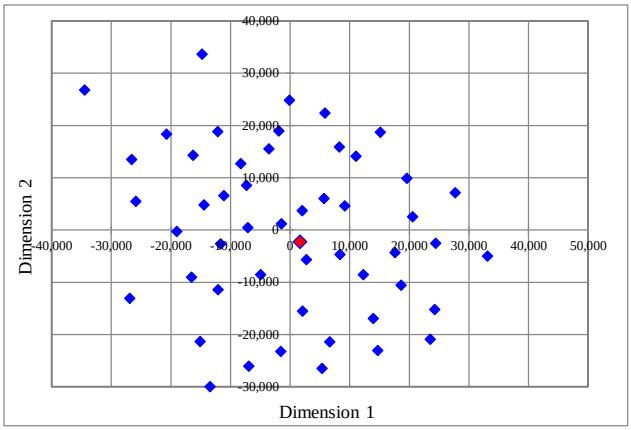


Figure A-4: Multidimensional scaling embedding of 50 realisations (blue diamonds) and the Kriging model (red diamond) in R^2. The Kriging point is at the minimum distance from the others (SGS and 3D case)

Figure A-5 shows the result of the interpoint distance calculations for 150 realisations and the Kriging model; the best fit distribution is still lognormal. The parameters of the histogram are shown in Table A-3. Figure A-6 shows multidimensional scaling embedding the realisations

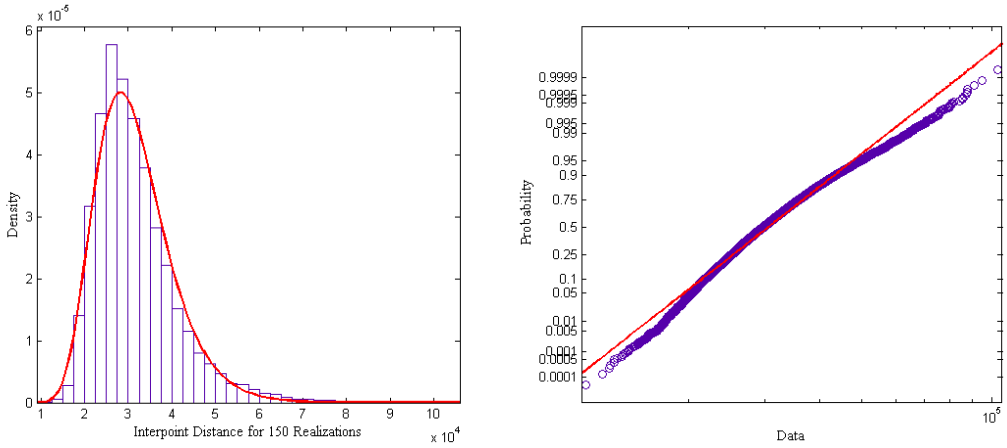


Figure A-5: The graph on the left is the histogram of 11,325 interpoint distances of 150 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them (SGS and 3D case)

Table A-3: Statistical parameters of the histogram of the interpoint distances of 150 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
11,325	31,601.94	9,408.51	88,520,153	11,582.3	103,194	91,611.70	1.50	22,331.7

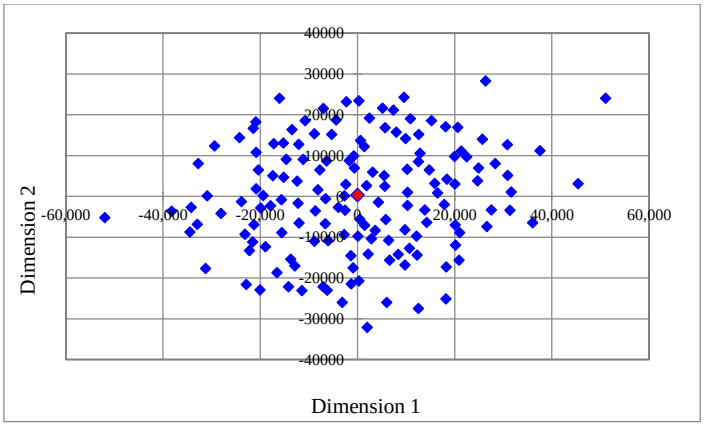


Figure A-6: Multidimensional scaling embedding of 150 realisations (blue diamonds) and the Kriging model (red diamond) in R^2. The Kriging point is at the minimum distance from the others

Figure A-7 shows the result of the interpoint distance calculations for 300 realisations and the Kriging model; the best fit distribution is still lognormal. The parameters of the histogram are shown in Table A-4. Figure A-8 shows multidimensional scaling embedding the realisations.

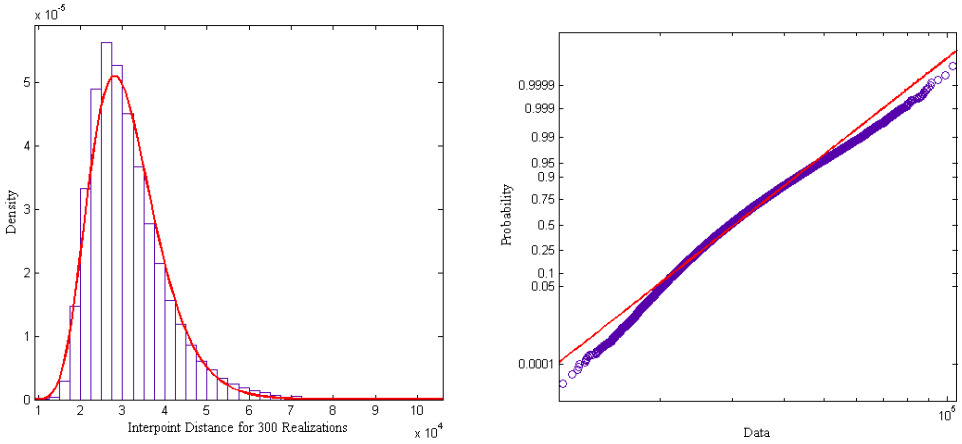


Figure A-7: The graph on the left is the histogram of 45,150 interpoint distances of 300 realisations and the Kriging model (dissimilarity distance matrix) with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution showing a reasonable fitting between them (SGS and 3D case)

Table A-4: Statistical parameters of the histogram of the interpoint distances of 300 realisations (SGS and 3D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Range	Skewness	Ave. Distance from Kriging
45,150	31,342.20	9,072.42	82,308,803	11,582.3	103,194	91,611.70	1.34	22,158.3

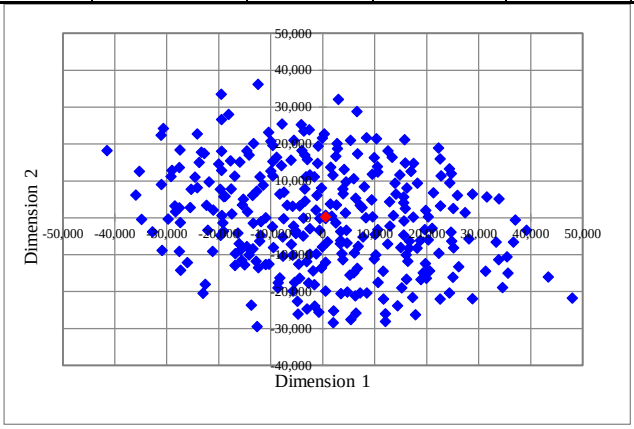


Figure A-8: Multidimensional scaling embedding of 300 realisations (blue diamonds) and the Kriging model (red diamond) in R^2. The Kriging point is at the minimum distance from the others

A.2 Sequential Gaussian Simulation algorithm (SGS and 2D Case)

Figure A-9 shows the result of the interpoint distance calculations for 50 realisations. As can be seen, the best distribution that can be fitted on output distribution is lognormal. The parameters of the histogram are shown in Table A-5.

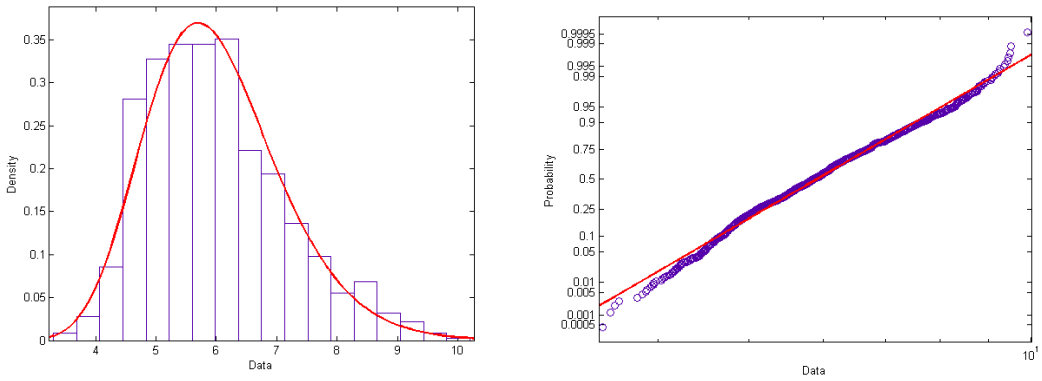


Figure A-9: The graph on the left is the histogram of 1225 interpoint distances of 50 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution which shows a reasonable fitting between them ($a \times 10^6$) (SGS and 2D case)

Table A-5: Statistical parameters of the histogram of the interpoint distances of 50 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
1225	5.998	1.144	1.309	3.492	9.905	0.651	5.996	4.901

$(a \times 10^6)$

Figure A-10 shows multidimensional scaling embedding of 50 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

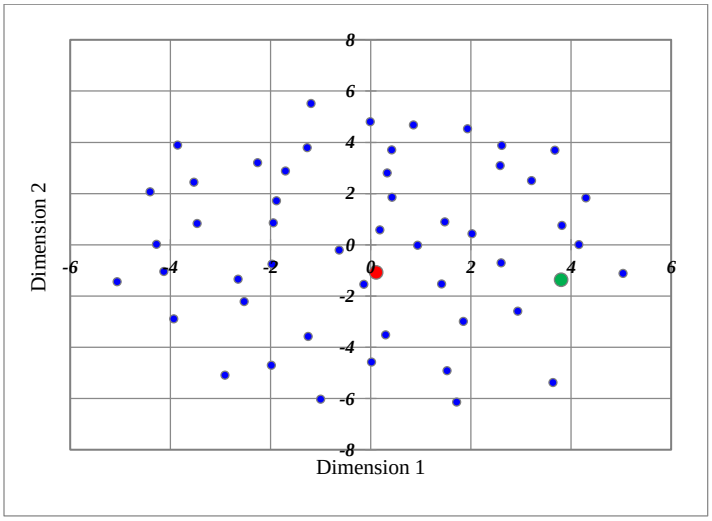


Figure A-10: Multidimensional scaling embedding of 50 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others, while the real model is quite far from the generated realisations (SGS and 2D case)

Figure A-11 shows the result of the interpoint distance calculations for 250 realisations. As can be seen, the best distribution that can be fitted on output distribution is lognormal. The parameters of the histogram are shown in Table A-6.

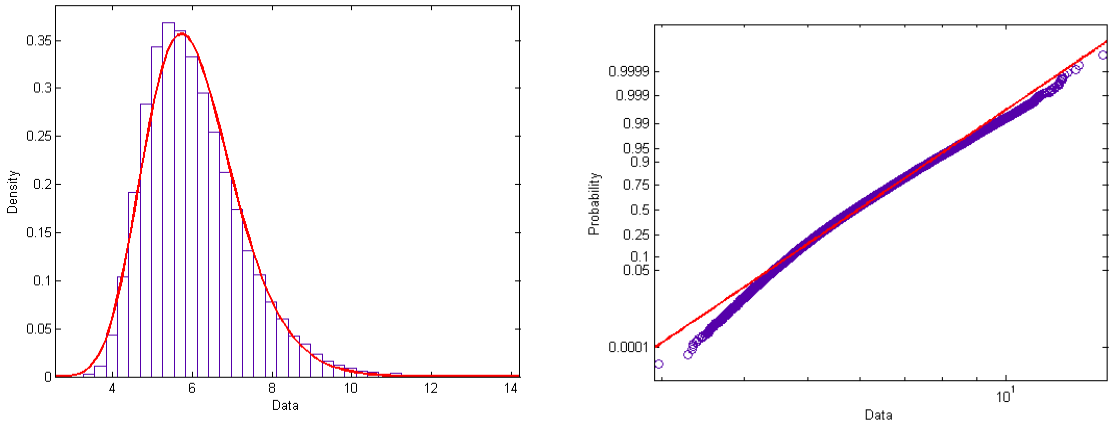


Figure A-11: The graph on the left is the histogram of 31,125 interpoint distances of 250 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution showing a reasonable fitting between them (a × 10^6) (SGS and 2D case)

Table A-6: Statistical parameters of the histogram of the interpoint distances of 250 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
31,125	6.058	1.204	1.456	3.397	14.024	0.906	6.092	4.998

($a \times 10^6$)

Figure A-12 shows multidimensional scaling embedding of 250 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

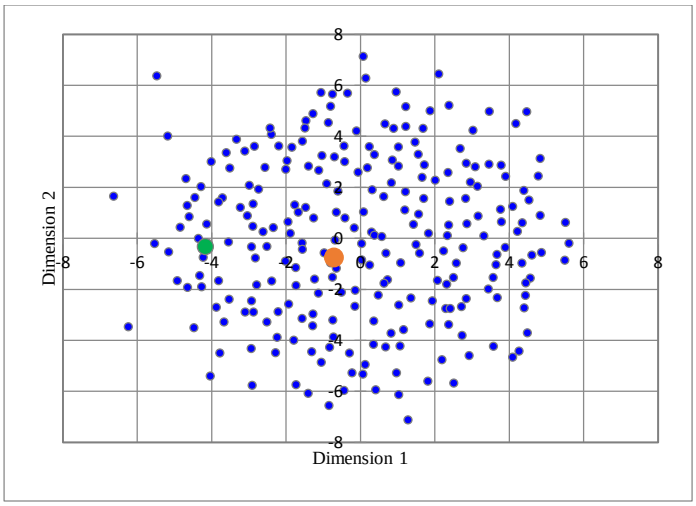


Figure A-12: Multidimensional scaling embedding of 250 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others while the real model is quite far from the generated realisations (SGS and 2D case)

Figure A-13 shows the result of the interpoint distance calculations for 600 realisations. As can be seen, the best distribution that can be fitted on output distribution is lognormal. The parameters of the histogram are shown in Table A-7.

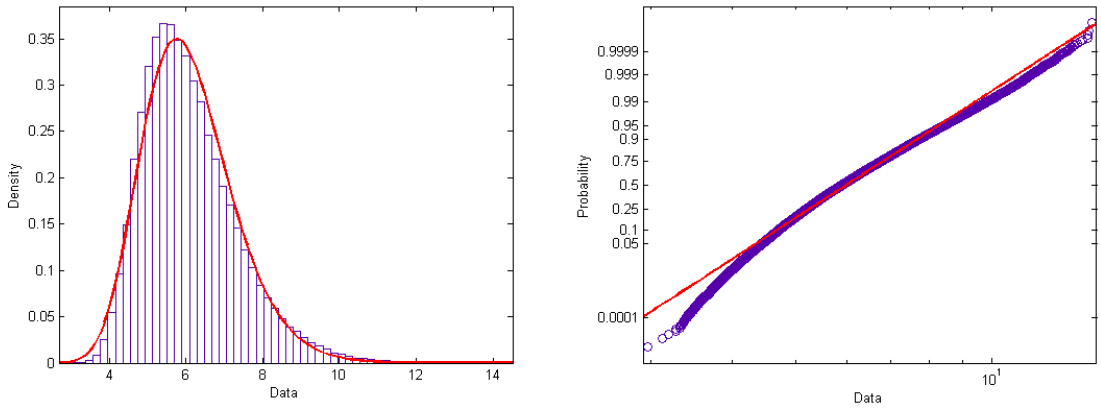


Figure A-13: The graph on the left is the histogram of 179,700 interpoint distances of 600 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution showing a reasonable fitting between them ($a \times 10^6$)- (SGS and 2D case)

Table A-7: Statistical parameters of the histogram of the interpoint distances of 600 realisations (SGS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
179,700	6.093	1.236	1.527	3.397	14.235	0.943	6.095	4.998

$(a \times 10^6)$

Figure A-14 shows multidimensional scaling embedding of 600 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

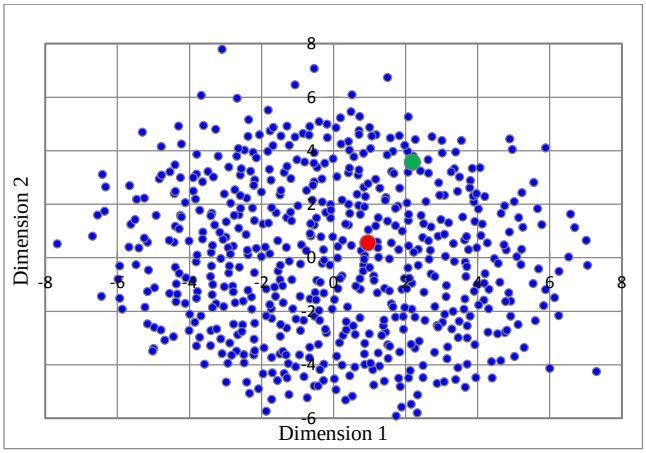


Figure A-14: Multidimensional scaling embedding of 600 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others while the real model is quite far from the generated realisations (SGS and 2D case)

A.3 Turning Bands Simulation algorithm (TBS and 2D Case)

Figure A-15 shows the result of the interpoint distance calculations for 50 realisations, and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-8

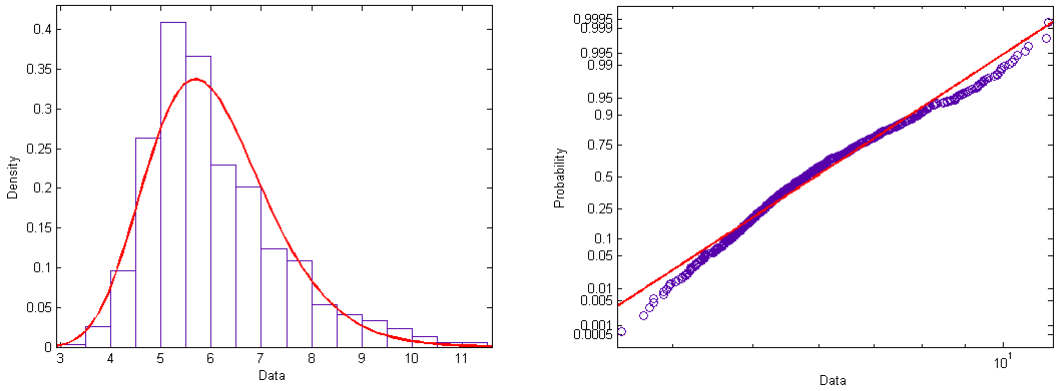


Figure A-15: The graph on the left is the histogram of 1,225 interpoint distances of 50 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (TBS and 2D case)

Table A-8: Statistical parameters of the histogram of the interpoint distances of 50 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
1225	5.975	1.290	1.660	3.472	11.343	1.031	5.845	4.643

($a \times 10^6$)

Figure A-16 shows multidimensional scaling embedding of 50 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance of the others, while the real model (green circle) is far from the generated realisations.

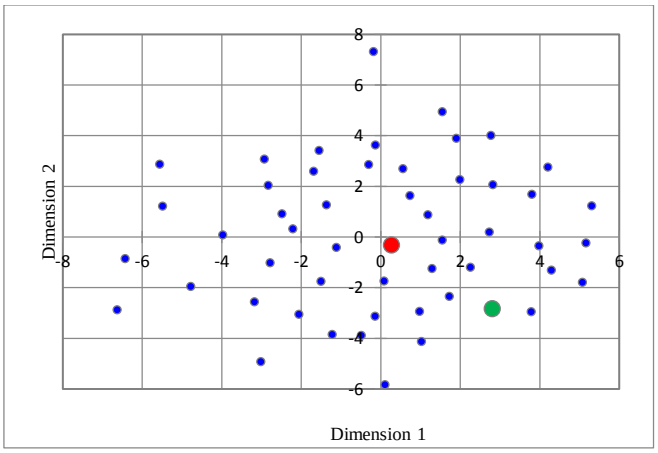


Figure A-16: Multidimensional scaling embedding of 50 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others, while the real model is quite far from the generated realisations (TBS and 2D case)

Figure A-17 shows the result of the interpoint distance calculations for 250 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-9.

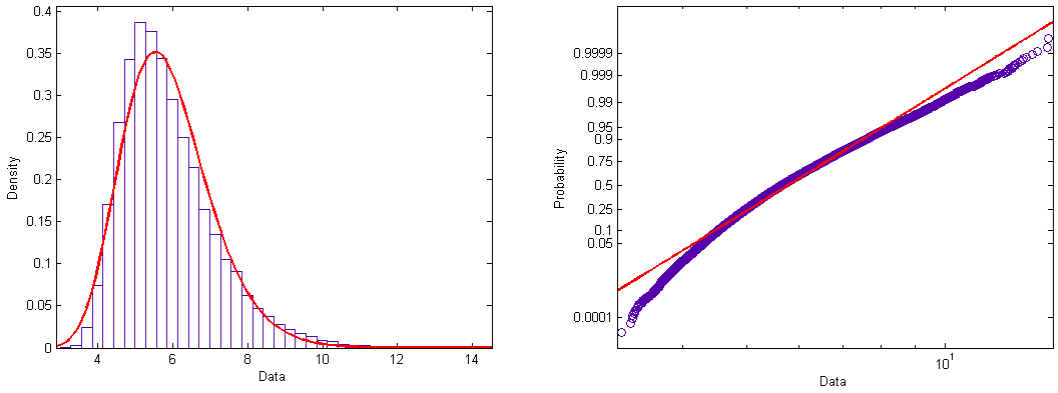


Figure A-17: The graph on the left is the histogram of 31,125 interpoint distances of 250 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (TBS and 2D case)

Table A-9: Statistical parameters of the histogram of the interpoint distances of 250 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
31,125	5.883	1.252	1.657	3.224	14.357	1.131	5.803	4.662

$(a \times 10^6)$

Figure A-18 shows multidimensional scaling embedding of 250 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

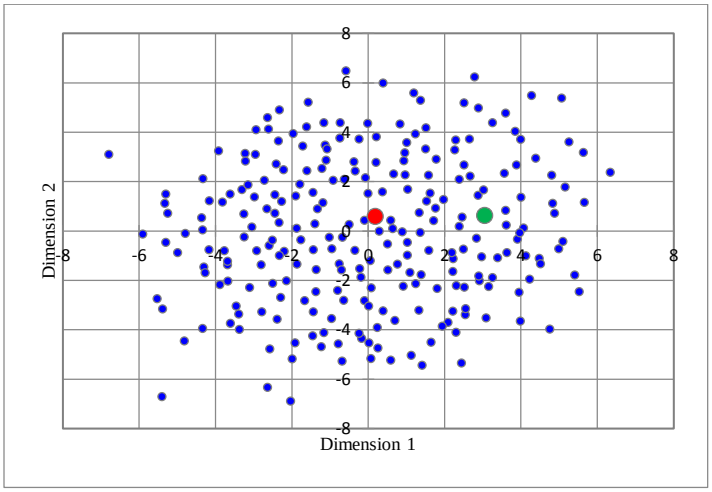


Figure A-18: Multidimensional scaling embedding of 250 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others, while the real model is quite far from the generated realisations (TBS and 2D case)

Figure A-19 shows the result of the interpoint distance calculations for 450 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-10.

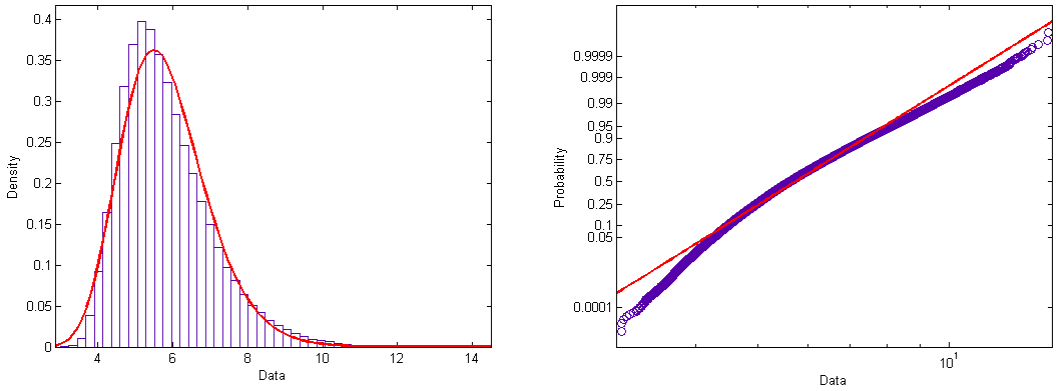


Figure A-19: The graph on the left is the histogram of 101,025 interpoint distances of 450 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circle) and the fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (TBS and 2D case)

Table A-10: Statistical parameters of the histogram of the interpoint distances of 450 realisations (TBS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Real model	Ave. Distance from Kriging
101,025	5.825	1.207	1.458	3.052	14.357	1.068	5.747	4.624

$(a \times 10^6)$

Figure A-20 shows multidimensional scaling embedding of 450 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations

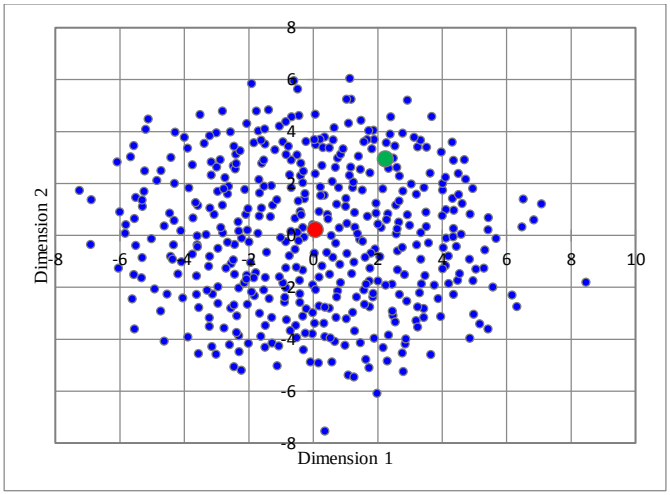


Figure A-20: Multidimensional scaling embedding of 450 realisations (blue circles), the Kriging model (red circle) and the real model (green circle) in R^2. The Kriging point is at the minimum distance from the others, while the real model is quite far from the generated realisations (TBS and 2D case)

A.4 Sequential Indicator Simulation algorithm (SIS and 2D Case)

Figure A-21 shows the result of the interpoint distance calculations for 50 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-11.

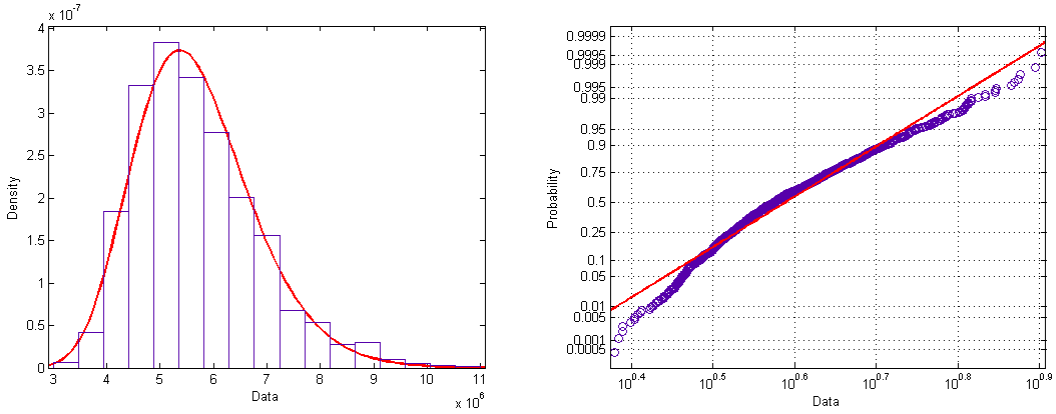


Figure A-21: The graph on the left is the histogram of 1,225 interpoint distances of 50 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (SIS and 2D case)

Table A-11: Statistical parameters of the histogram of the interpoint distances of 50 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
1225	5.682	1.156	1.336	3.482	10.64	0.893	4.03	6.37	4.84

$(a \times 10^6)$

Figure A-22 shows multidimensional scaling embedding of 50 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations. The results are similar to the SGS algorithm.

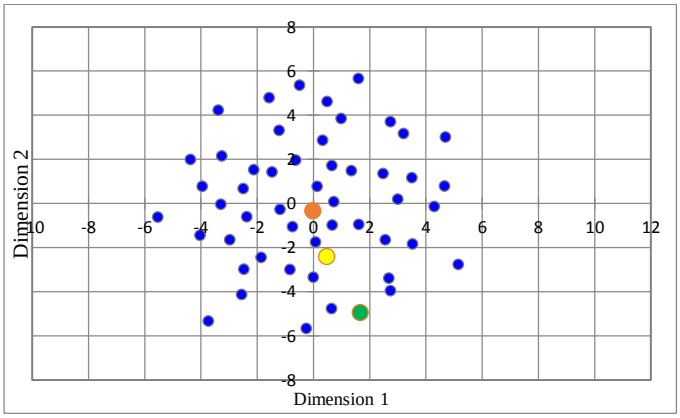


Figure A-22: Multidimensional scaling embedding of 50 realisations (blue circles), Average of 30 realisations (red circle), the Kriging model (yellow circle) and the real model (green circle) in R^2. The average model is at the minimum distance from the others, while the real model is quite far from the generated realisations (SIS and 2D case)

Figure A-23 shows the result of the interpoint distance calculations for 250 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-12.

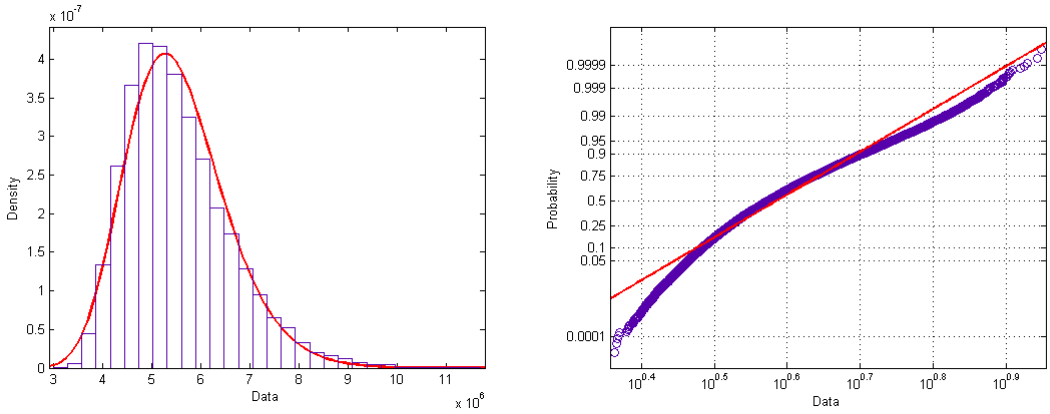


Figure A-23: The graph on the left is the histogram of 31,125 interpoint distances of 250 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (SIS and 2D case)

Table A-12: Statistical parameters of the histogram of the interpoint distances of 250 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
311,125	5.547	1.056	1.115	3.145	11.42	0.898	3.950	6.275	4.711

($a \times 10^6$)

Figure A-24 shows multidimensional scaling embedding of 250 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

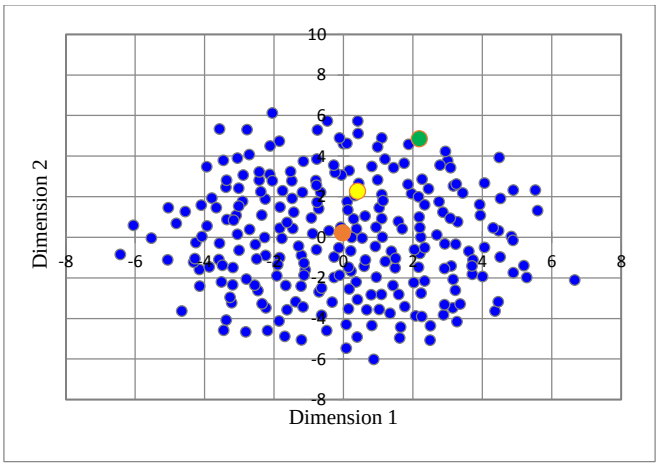


Figure A-24: Multidimensional scaling embedding of 250 realisations (blue circles), the average of 250 realisations (red circle), the Kriging model (yellow circle) and the real model (green circle) in R^2. The average model is at the minimum distance of the others, while the real model is quite far from the generated realisations (SIS and 2D case)

Figure A-25 shows the result of the interpoint distance calculations for 450 realisations and the fitted lognormal distribution. The parameters of the histogram are shown in Table A-13.

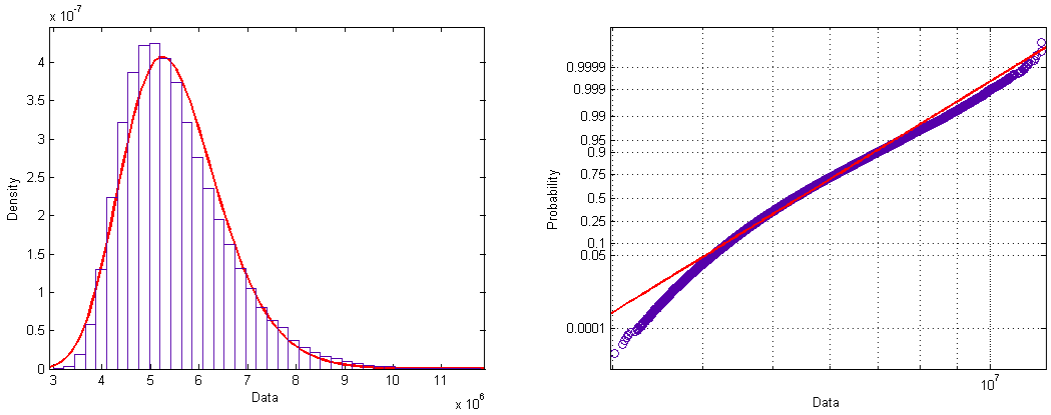


Figure A-25: The graph on the left is the histogram of 101,025 interpoint distances of 450 realisations with the best fitted lognormal distribution (red line). The graph on the right shows the probability plot of the histogram (blue circles) and the fitted lognormal distribution, showing a reasonable fitting between them ($a \times 10^6$)- (SIS and 2D case)

Table A-13: Statistical parameters of the histogram of the interpoint distances of 450 realisations (SIS and 2D case)

No. Interpoint Distance	Mean	Std. Deviation	Variance	Minimum	Maximum	Skewness	Ave. Distance from Average model	Ave. Distance from Real model	Ave. Distance from Kriging
101,025	5.524	1.059	1.121	3.015	11.78	0.922	3.929	6.278	4.70

$(a \times 10^6)$

Figure A-26 shows multidimensional scaling embedding of 450 realisations. As illustrated, the Kriging point (red circle) is at the minimum distance from the others, while the real model (green circle) is far from the generated realisations.

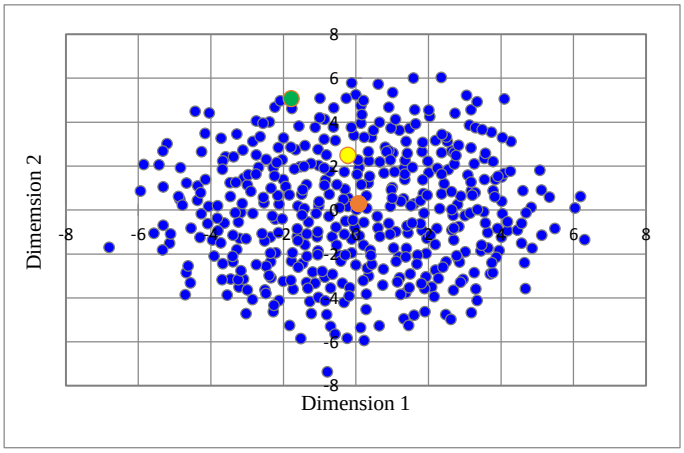


Figure A-26: Multidimensional scaling embedding of 450 realisations (blue circles), the average of 450 realisations (red circle), the Kriging model (yellow circle) and the real model (green circle) in R^2. The average model is at the minimum distance from the others, while the real model is quite far from the generated realisations (SIS and 2D case)

Bibliography

Akaike, A. , Dagdelen, K. A, 1999. Strategic production scheduling method for an open pit mine. The 28th Application of Computers and Operations Research in the Mineral Industry (APCOM). pp.729-738.

Arpat, B. , Caers, J. 2007. Stochastic simulation with patterns. Math Geol 39, pp. 177–203.

Armstrong M. , Ndiaye A. , Razanatsimba R. and Galli A. Scenario Reduction Applied to Geostatistical Simulations. Math Geosci (2013) 45:165–182

Assent I., Wichterich M., Seidl T. 2006. Adaptable Distance Functions for Similarity-based Multimedia Retrieval. Datenbank-Spektrum Nr. 19 p. 23-31

Boland, N., Dumitrescu, I., Froyland G. and Gleixner, 2009. A.M. LP-based disaggregation approaches to solving the open pit mining production scheduling problem with block processing selectivity.

Bellman, R. E. 1957. Dynamic Programming, Princeton University. Press: Princeton.

Boucher, A. , Dimitrakopoulos, R. 2009. Block Simulation of Multiple Correlated Variables. Math Geosci, 41, pp. 215–237.

Caccetta, L. , Hill, S. P. 2003. An application of branch and cut to open pitmine scheduling. Journal of Global Optimization, pp. 349 – 365.

Caers, J. 2011. Modeling Uncertainty in the Earth Sciences, John. Wiley & Sons.

Coen M.H. 2007. A Similarity Metric for Spatial Probability Distributions. in Proc. Natl. Conf. on Artificial Intelligence.

Cox, T. F. , Cox, M. A. A. 2000. Multidimensional scaling. Chapman and Hall.

Dagdelen, K. , Johnson, T. B. 1986. Optimum open pit mine production scheduling by Lagrangian parameterization. The 19th International Symposium on the Application of Computers and Operations Research in the Mineral Industry (APCOM). pp. 127-142.

Deng, Y. and W. Du. 2009. The Kantorovich metric in computer science: A brief survey. Electronic Notes in Theoretical Computer Science (253). pp. 73-82.

Duch, W. , Grudzinski, K. 1998. A framework for similarity-based methods. Theory and Applications of Artificial Intelligence: pp. 33-60.

Deutsch, C. V. , Journel, A. G. 1998. GSLIB Geostatistical Software Library and user's guide, ,New York, NY, , Oxford University Press.

Dimitrakopoulos, C. T., Farrelly, et al. 2002. Moving forward from traditional optimization: grade uncertainty and risk effects in open-pit design. Trans. Instn Min. Metall (Sect. A: Min. technol.)

Dimitrakopoulos, R. 1998. Conditional simulation algorithm for modelling orebody uncertainty in open pit optimisation. international of surface mining, reclamation and environment, 12. pp. 173-178.

Dimitrakopoulos, R. 2011. Stochastic Optimization for Strategic Mine Planning: A Decade of Developments. Journal of Mining Science, 47.

Dimitrakopoulos, R., L., Martinez, et al. 2007. A maximum upside / minimum downside approach to the traditional optimization of open pit mine design. Journal of Mining Science 43(1): pp. 73-82.

Dimitrakopoulos, R. , Ramazan, S. 2004. Uncertainty-based production scheduling in open pit mining. Society for mining, metallurgy, and exploration (SME), 136.

Dowd, P. A. 1994. Risk assessment in reserve estimation and open-pit planning. Trans. Instn. Min. Metall, 103, A148-A154.

Dowd, P. A. , David, 1976. M. Planning from estimates: sensitivity of mine production schedules to estimation methods. In Advanced Geostatistics in the Mining Industry, Italy. D. Reidel Publishing Company.pp. 163-183.

Dowd, P. A. , Onur, A. H. 1992. Optimizing open pit design and sequencing. The 23rd International Symposium on the Application of Computers and Operations Research in The Mineral Industries (APCOM). pp. 411-422.

Dubuisson, M. P. , Jain, A. 1994. A modified Hausdorff distance for object matching. International conference on pattern recognition. Jerusalem, Israel. pp. 566-568.

Dupacova, J., Growe-Kuska, N., Romisch, W., 2003. Scenario reduction in stochastic programming an approach using probability metrics, Math. Program. 95, pp. 493-511.

Growe-Kuska, N., Heitsch N., Romisch, W., 2003 Scenario reduction and scenario tree construction for power management problems. IEEEE Bologna Power Tech Proceedings.

Franti, P. , Kivijarvi, J. 2000. Randomised Local Search Algorithm for the Clustering Problem. Pattern Analysis & Applications 3, pp. 358–369.

Gershon, M. E. 1983. Optimal mine production scheduling: evaluation of large scale mathematical programming approaches. International Journal Mining Engineering, pp. 315-329.

Glacken, I. M. , Snowden, D. V. 2001. Mineral Resource Estimation Mineral Resource and Ore Reserve Estimation-The AusIMM Guide to Good Practice. Australia-Melbourne.

Goovaerts, P. 1997. Geostatistics for Natural Resource Evaluation, Oxford Universty Press.

Goovaerts, P. 1999. Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. Stochastic Environmental Research and Risk Assessment 13, pp. 161-182.

Gotway, C. A. , Rutherford, B. M. 1993. Stochastic simulation for imaging spatial uncertainty comparison and evaluation of available algorithms. In Geostatistical simulations, Kluwer Academic Publishers, pp. 1-21.

Han, J., M. Kamber, Pei, J. 2011. Data Mining: Concepts and Techniques. Morgan Kaufmann.

Honarkhah, M. , Caers, J. 2010. stochastic Simulation of Patterns Using Distance-Based Pattern Modeling. Mathematical Geosciences, 42, 487–517.

Honarkhah, M. 2011. Stochastic Simulation of Patterns using Distance-Based Pattern Modelling. The department of energy resources engineering. California, Stanford University. Ph.D.

HU, L. Y. , Ravalec-dupin, M. L. 2004. On some controversial issues of geostatistical simulations. Geostatistics Banff. Springers.

Isaaks, E. H. , Srivastava, R. M. 1989. An Introduction to Applied Geostatistics, Oxford University Press.

Johnson, T. B. 1969. Optimum production scheduling. the 8th International Symposium on Computers and Operations Research. pp. 539 – 562.

Journel, A. 1997. The abuse of principles in model building and the quest for objectivity. In: Academic, K., ed. Geostatistics Wollongong '96, pp. 3-14.

Journel, A., Kyriakidis, P. C. , Mao, S. 2000. Correcting the smoothing effect of estimators: A spectral postprocessor. Mathematical Geology, 32, pp. 787-812.

Journel, A. G. 1982. The indicator approach to estimation of spatial data. The 17th International Symposium on the Application of Computers and Operations Research in The Mineral Industries (APCOM). Port City Press. pp. 793-806.

Jović M., Hatakeyama Y., Stejić Z., Stopić S., Seidl T., Oljača D. and Hirota K. 2007. Earth Mover's Distance Region-Based Image Similarity Modeling Applied To Mineral Image Retrieval. MJoM. vol. 13, br. 2, str. 117-130

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29(1). pp. 1-27.

Lapworth, A. 2011. Datamine strategic planning solution, NPVscheduler, Editor. Datamine Corporated Limited.

Leite, A. , Dimitrakopoulos, R. 2007. Stochastic optimisation model for open pit mine planning: application and risk analysis at copper deposit. Mining Technology, 116, pp. 109-118.

Lerchs, H. , Grossman, F. 1965. Optimum design of open-pit mines. Trans. CIM, Vol. 58 (1), pp. 47-54.

Malvin K., H. , Whitlock, P. A. 1986. Monte Carlo Methods, John Wiley & Sons.

Martinez, L. A. 2009. Why Accounting for Uncertainty and Risk can Improve Final Decision-Making in Strategic Open Pit Mine Evaluation. Project Evaluation Conference. Melbourne, Australia.

Matheron, G. 1965. Les variables régionalisées et leur estimation. Masson, Paris.

Mukherjee. M. N. 2006. Elements of metric spaces, Dhar Academic Publishers.

Myers, D. 1994. Geostatistical simulation thoughts and questions. Geostatistics, 7(1).

Onur, A. H. , Dowd, P. A. 1993. Open pit optimization-part 2: production scheduling and inclusion of roadways. Min. Metall. (Sec. A: Min. Industry), 102, A105 – A113.

Osanloo, M., Gholamnejad, J. , Karimi, B. 2008. Long-term open pit mine production planning: a review of models and algorithms. International Journal of Mining, Reclamation and Environment, 22, pp. 3-35.

O'Searoid. M. 2006. Metric spaces. Springer.

Pele O. and Werman M. 2009. Fast and robust earth mover's distances. In Proceedings of the 12th International Conference on Computer Vision (ICCV '09), pp. 460–467.

Qureshi, S. E. 2002. Comparative study of simulation algorithms in mapping spaces of uncertainty. MPhil University of Queensland.

Qureshi, S. E. , Dimitrakopoulos, R. 2007. Comparison of stochastic simulation algorithms in mapping spaces of uncertainty of non-linear transfer functions. Springer, pp. 959-968.

Ramazan, S., Dagdelen, K. , Johnson, T. B. 2005. Fundamental tree algorithm in optimizing production scheduling for open pit mine design. Trans. Inst. Min. Metall. (Sec. A: Mining Technol), 114, A45-A114.

Ramazan, S. , Dimitrakopoulos, R. 2004. Traditional and new MIP models for production scheduling with in-situ grade variability. International Journal of Mining, Reclamation and Environment, 18, pp. 85-98.

Roman, R. J. 1974 . The role of time value of money in determining an open pit mining sequence and pit limits. The 12th Symposium on the Application of Computers and Operation Research in the Mineral Industries (APCOM). pp. 72-85.

Shirali, S. , Vasudeva, H. L. 2006. Metric Spaces. Springer.

Shishibori M. and et al., 2011. Fast Retrieval Algorithm for Earth Mover's Distance Using EMD Lower Bounds and a Skipping Algorithm. Hindawi Publishing Corporation, Advances in Multimedia.

Rubner, Y., Tomasi, C. , Guibas L. J. 1998, Metric for Distributions with Applications to Image Databases. Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India.

Rossi, M. E. , Parker, H. M. 1994. Estimating recoverable reserves: Is it hopeless? In: Dimitrakopoulos, R., ed. Geostatistics for the Next Century, Kluwer Academic Publishers, pp. 259-276.

Suzuki, S. , Caers, J. 2008. A distance based prior model parameterization for constraining solution of spatial inverse problems. Math Geosci, 40, pp. 445-469.

Sarma, P. 2006. Efficient closed-loop optimal control of petroleum reservoirs under uncertainty. Stanford University, USA.

Savage, S. L. , Danzi, J. 2009. The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty, John Wiley & Sons.

Scheidt, C. , Caers, J. 2009. Representing spatial uncertainty using distances and kernels. Mathematical Geosciences, 41, pp. 397-419.

Srivastava, R. 1994. Thoughts and comments on conditional simulation algorithms. Geostatistics, 7(2).

Srivastava, R. Matheronian geostatistics: where is it going? 1997. Geostatistics Wollongong '96. Kluwer Academic.

Strebelle, S. , Journel, A. G. 2001. Reservoir modeling using multiple-point statistics. Annual Technical conference and Exhibition. New Orleans.

Suzuki, S. , Caers, J. 2006. History matching with an uncertain geological scenario. SPE Annual Technical Conference and Exhibition.

Tang, Y.; U, L. H.; Cai, Y.; Mamoulis, N. & Cheng, R. (2013), 'Earth Mover's Distance based Similarity Search at Scale.', PVLDB 7 (4) , 313-324

Tebboth, J. Richard , 2001. A computational study of Dantzig-Wolfe decomposition, PhD dissertation . University of Buckingham, United Kingdom.

Tolwinski, B. , Underwood, R. 1992. An algorithm to estimate the optimal evolution of an open pit mine. The 23rd International Symposium on the Application of Computers and Operations Research in The Mineral Industries (APCOM), pp. 399-409.

Tolwinski, B. 1998. Pit optimization and mine production scheduling—the way ahead. The 27th International Symposium Application of Computers and Mathematics in The Mineral Industries (APCOM), pp. 19-23.

Tolwinski, B , Golosinski, T. S. 1995. Long term open pit scheduler. The International Symposium on Mine Planning and Equipment Selection, pp. 256-270.

Vallee, M. 2000. Mineral resource + engineering, economic and legal feasibility = Ore reserve. . CIM Bulletin, pp. 53-61.

Villani, C. 2003. Topics in optimal transportation. USA, American Mathematical Society.

Vershik, A. M. 2005. Kantorovich Metric: Initial History and Little-Known Applications. Journal of Mathematical Sciences. pp. 1-8.

Webb, A. R., Copsey K. D. 2011. Statistical Pattern Recognition, John Wiley and sons.

Webster, R. , Oliver, M. A. 2007. Geostatistics for Environmental Scientists, England, John Wiley & Sons.

Whittle J. 1998, Whittle Programmings's Present and Future Tools for the Strategic Planning of Mines, The 3rd International Symposium on the Application of Computers and Operations Research in The Mineral Industries (APCOM), pp. 27-34.

Yamamoto, J. K. 2005. Correcting the Smoothing Effect of OrdinaryKriging Estimates. Mathematical Geology, 37, pp. 69-95.

Yamamoto, J. K. 2008. Estimation or simulation? That is the question. Computer Geosciences 12, 273-592.

Yoram, R. 2003. Applied Stochastic Hydrogeology, New York, Oxford University Press.

Zhang, T., Switzer, P. , Journel, A. G. 2006. Filter-based classification of training image patterns for spatial simulation. Math. Geosci, 38, pp. 63-80.