MULTIPLE-POINT STATISTICAL SIMULATION IN MODELLING A STRUCTURALLY COMPLEX GEOLOGICAL ENVIRONMENT

Yelena Michailovna van der Grijp

A thesis submitted to the Faculty of Engineering and the Built Environment,

University of the Witwatersrand, Johannesburg,

in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg 2021

DECLARATION

I declare that this thesis is my own unaided work. It is being submitted for the Degree of Doctor of Philosophy to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other University. The data used in the thesis has been obtained while employed by AngloGold Ashanti.

Yelena van der Grijp

……

31 January 2021

…….. day of ……., ……

ABSTRACT

An innovative approach to geological modelling and stochastic estimation of a Mineral Resource in a structurally complex gold deposit is presented. An existing Direct Sampling (DS) multiple-point statistics (MPS) algorithm is adopted to produce stochastic models of lithology and gold grade distribution at point support, conditioned to sparse geological and grade data, following a non-parametric multi-variate framework. An ensemble of MPS realisations is upscaled to an open pit mining unit support to obtain an open pit Mineral Resource. A mineable shape optimisation algorithm is used to derive underground stopes for reporting the ore quantity and uncertainty associated with an underground mining method. Comparison is made to understand an inherent increase in uncertainty in mineralised material (tonnage, gold grade and metal quantity) when moving from the open pit to the underground highly selective environment.

The first explicit three dimensional (3D) geological model of the deposit is constructed and a workflow to do it in future proposed. A considerable effort is dedicated to understanding spatial relationships between the variables to derive the exhaustively known ‘true’ model. While modelling finely intercalated folded lithology explicitly with classical methods underperforms even in presence of dense production data, a minimal input from a geologist is envisaged in this approach for a successful model, in a form of an elementary training image and a set of structural measurements of the lithological contact to create a representative potential field of ductile deformation.

A new modelling framework is proposed where (1) a portion of the ‘true’ model is reserved to be used as a training image, while (2) the remaining part serves as a validation tool prior to arriving at a satisfactory ensemble of realisations (3) to be carried forward as a stochastic model for the poorly informed part of the deposit. This approach lends itself to the adaptation of modern developments in artificial intelligence and machine learning. The

specialist's involvement is confined to incorporating understanding of the deposit and its genesis into the explicit 'prior' model.

Validation of the algorithm performance showed that complex multivariate relationships are respected during an MPS simulation, such as honouring of the main structural controls, folded geometry of lithological contact, proximity of the mineralisation to the fold hinge zone affecting the attitude of the shears, distribution of the tenor of mineralisation in relation to the potential field, character of the soft boundary reproduction in the spatial gold grade distribution in relation to the lithological contact boundary, and localisation of ore shoots along the intersection of structural controls (bedding and shearing). The latter is considered of high importance when generating truthful underground mining designs.

Methodological contribution, among others, comprised a finding that reproduction of the gold grade patterns improves when conditioning it to a multiple indicator gold grade variable in a co-simulation mode. The potential field created from the lithological contact served as a non-stationarity descriptor to guide the gold grade simulation, improving it further. For categorical variables (the mineralised domain and lithologies), the proportions reproduction is achieved within 3% deviation from the 'true' model and this approach can be used as-is in future simulations.

The method showed robust performance with sparse data or even in absence of data, provided sufficient understanding of geology and mineralisation in a form of a multi-variate 'true' model exists and the geologist believes the 'true' model is representative of the unknown. As such, the approach yields itself well to in-depth stochastic mining engineering at early stages in a project life when only sparse drilling is available. The stochastic framework allows for an integrated approach to Mineral Resource classification and Mineral Resource and Ore Reserves reporting, with associated geological and geostatistical uncertainty tabulation.

There are other advantages of the proposed workflow such as ability to run it in a reconstructive mode with partial update of the model. Smooth and natural transition between the ‘frozen’ and new areas with honouring of the spatial structures is allowed, and the definition of non-stationarity is improved continuously.

A list of suggested future enhancements is given with regards to methodology, the improvements to the DS algorithm, and the MPS approach in general.

At the moment of conducting the research, no mentioning could be found in literature with regards to a successful MPS simulation of continuous variables in a context of hard-rock mining for large 3D models.

Dedicated to my parents, my husband and my son

To all the victims of the COVID-19, in lives and livelihoods

ACKNOWLEDGEMENTS

I would like to express gratitude to the following people:

Professor Philippe Renard and Dr. Julien Straubhaar for providing support in understanding the Direct Sampling multiple-point statistical algorithm they developed;

My two supervisors, Professor Richard Minnitt and Dr. David Rose for their continuous support, inspiration and patience reviewing the thesis;

My examiners for their valuable comments;

To AngloGold Ashanti for allowing the use of the data originating from their flagship Geita mine and a permission to publish the results.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

x	An elementary unit of MPS simulation in the simulation grid, be it a node or a patch
y	An elementary unit of MPS simulation in the training image, be it a node or a patch
d_n	Data event, a set of n nodes that are characterized by separation distances, termed lags, from a visiting node in the grid
$L(x)$	A local probability map over the simulation grid
$Z(u)$	Generic random variable at location u
$n(u)$	Number of conditioning data found in a neighbourhood centred at u
h_i	Lag vector separating a node in the neighbourhood from the node being simulated
f	Fraction of the training image to be scanned when simulating each node
n	Maximum number of conditioning nodes to be considered in the neighbourhood as stipulated in the input to simulation
t	Acceptance threshold for the similarity distance measure

PHYSICAL QUANTITIES

g/cm^3	density, grams per cubic centimetre
g/t	metal concentration, grams per metric tonne

NOMENCLATURE

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
Au	Gold
BIF	Banded iron formation
CAD	Computer-aided design
CCSIM	Cross-correlation simulation function
ccdf	Cumulative conditional density function
cdf	Cumulative distribution function
CIQ	Conditional image quilting
CPU	Central processing unit
DS	The Direct Sampling multiple-point simulation algorithm
DeeSse	A recently developed (Straubhaar, 2019) variation of the Direct Sampling algorithm with extended functionality, among which block conditioning, connectivity constraints and multiple scale simulation based on Gaussian pyramids
EDA	Exploratory data analysis
ENESIM	Extended Normal Equation Simulation
FILTERSIM	Simulation using filter scores
GHSZ	Geita Hill Shear Zone
GSLib	Geostatistical software library (Deutsch and Journel, 1998)
HOSIM	High-order simulation
IDW	Inverse Distance Weighted
Impala	An Improved Parallel Multiple-Point Algorithm

MIP	Maximum Intensity Projection
ML	Machine learning
MPS	Multiple-point statistics
MSO	Mineable Shape Optimiser
pdf	Probability density function
RAM	Rapid Access Memory
RBF	Radial Basis Function
RC	Reverse circulation
RL	Relative level
SAMREC	The South African Mineral Resource Committee
SAMREC Code	The South African Code for the Reporting of Exploration Results, Mineral Resources And Mineral Reserves
SG	Simulation grid
SGS	Sequential Gaussian simulation
SIMPAT	Simulation with patterns
SIS	Sequential indicator simulation
SMU	Selective mining unit
SNESIM	Single normal equation simulation
TI	Training image

"The inception of the multiple point concepts has brought major changes into the way one sees spatial modelling. With the remarkable processing power now available on desktop computers, there is no more reason to ignore critical prior structural information brought by human expertise under the argument that it is subjective or "messy", not delivered concisely through a few statistics. Just like the original development of geostatistics in the 1960's was made possible by the availability of digital computing (variogram calculation and solving kriging systems), now is the time to graduate into massive processing of expert structural information which will tie the discipline of geostatistics increasingly more to image and computer sciences".

— André Journel (2004)

1 RESEARCH BACKGROUND

1.1 Motivations

The global mining industry is facing considerable challenges. Growing costs to sustain generation of new Mineral Reserves and to maintain production, increasingly stringent environmental laws with a focus on renewable energy, competition for shrinking water supplies, growing social responsibility and license to operate are just a few of them. Major ore discoveries become rarer, necessitating a move into less explored regions, available deposits have increased structural complexity with orebodies at greater depths, lower mineralisation tenor and geometallurgical difficulties. Moving into hard rock accompanied by the presence of deleterious elements result in additional processing costs and lower recoveries. This leads to increased engineering complexity to establish Ore Reserves. The growing cost of borrowing combined with lower investors' appetite for risk place a bigger focus on advanced characterisation of ore bodies with a requirement for improved uncertainty assessment. The latter is becoming of paramount importance to

sound investment decision making, to prioritise investment among competing projects, and, for mines in operation, to optimally sequence the ore extraction for a faster return on investment. There is a positive side to such challenges - they speed up the development and implementation of technological innovations. With an increase in geological complexity and an aversion in risk and expenditure to define Mineral Resources, focus needs to shift to using innovative and versatile tools for an efficient description of complex orebodies while utilising all available geological knowledge and not overcapitalising on drilling.

One aspect becoming of paramount importance when modelling geologically complex orebodies is the truthful representation of high-grade continuity. In an underground environment, sufficient continuity of high-grade mineralisation is what defines and drives the design of stopes. At the same time, many shallow, but marginally mineralised orebodies, are sensitive to viable extraction if the continuity of mineralisation is not modelled correctly.

The problem of a truthful representation of continuity was realised in oil and gas reservoir engineering when facing difficulties in representing complex paths in carbon flow. This industry pioneered in developing methods to model complex structures with preserved continuity of extreme values in a stochastic framework while relying on very limited data (Strebelle & Journel, 2001; Caers & Zang, 2002; Chugunova & Hu, 2008). Twenty-seven years ago, it gave rise to multiple-point statistical simulation methods (Guardiano & Srivastava, 1993). The approach, however, has not yet found sufficient recognition in mining in 2020.

The focus of this research is on the stochastic characterisation of a complex geological and gold mineralisation environment utilising one of such multiple-point statistical simulation methods. This research is a continuation of a first attempt to model spatial connectivity by utilising the Direct Sampling multiple-point statistics algorithm for a gold deposit (van der Grijp and Minnitt, 2014). The first study focused on reproducing spatial patterns of the

lithological interface between three geostatistically different lithologies using multiple-point statistics (MPS) with further subjection of the results to Sequential Gaussian Simulation (SGS) for simulation of gold grades. A good representation of lithological patterns and, specifically, good continuity of tabular barren intrusive bodies was achieved by utilising an elementary training image (TI) with basic tabular morphologies.

The study is taken further in this research. A novel approach to geological and geostatistical stochastic modelling was adopted. A first explicit model of Geita Hill gold deposit was generated with complex multi-variate relationships and folded geometries. This ‘true’ model was used as an explicit ‘prior’ image to generate 20 realisations of lithologies and gold grade in the structurally complex gold deposit using the Direct Sampling multiple-point statistics algorithm at point-support. Validation showed good reproduction of complex multi-variate relationships, the highest significance is attributed to reproduction of categorical proportions, the soft boundary in the spatial distribution of gold grades in the vicinity of lithological contact and localisation of the high-grade ore shoots along the modelled structural control variables, important for truthful designs of underground stopes. The method performs well in presence of sparse data or even in absence of data provided exhaustively known secondary variables controlling gold mineralisation are available over the simulation volume. The new modelling framework allows for adaptation in artificial intelligence and machine learning.

Subjecting the stochastic model to a mineable stope optimisation algorithm allowed global quantification and classification of the Mineral Resource by an underground mining method. The method can be used for localised uncertainty assessment in mineralised material: tonnes, gold grade and metal content and in stochastic engineering.

1.1.1 **Challenges faced in geostatistical domaining**

The Mineral Resource evaluation process consists of estimating grade and tonnage of the material which is considered viable for eventual economic extraction. It starts with the creation of a sound genetic, geological and mineralogical model of the deposit.

In the mineral extraction process, a geological model reflects the variables that represent controls on the mineralisation, allows for the definition of geostatistical domains and forms the basis of subsequent estimation of the mineral of interest (Boisvert, Leuangthong, Ortiz & Deutsch, 2008a). In the exploitation of aquifers and carbon reservoirs, a geological model is crucial for reliable modelling of physical processes that describe flow and transport processes, or hydrocarbon production (Renard & Allard, 2013; Rossi & Deutsch, 2015).

Frequently, a complete geological model is substituted with a model of stationary geostatistical domains. The definition of such domains is based on a combination of two or more geological variables, for which a relationship with the main variable can be demonstrated (Rossi & Deutsch, 2015). In mineral deposits, it can be stratigraphic architecture, presence and frequency of discrete structures, lithological contrasts or alteration signatures. Defining the geometry of such units informs the quantification of the mineralised tonnage available in the deposit. While some domains will be mostly mineralised and non-refractory, easily yielding to an extraction process with a potential of being classified as ore, other units might have lower tenor, a patchy type of mineralisation or host deleterious elements which would make them likely to be classified as marginal grade or refractory ore (Rossi & Deutsch, 2015). This model is used for mine planning and mine design, and forecasting plant feed by material type. It also informs other mining disciplines, such as hydrogeology or geotechnical engineering, for decision making.

An accurate geological model is the most important factor in the estimation of mineralised volumes and tonnage. However, accuracy in estimating the

tonnage globally is not the only consideration. The shape and connectivity of the modelled domain morphologies are equally vital.

The correct representation of connectivity patterns becomes specifically important when engineering for a higher risk underground mining environment, be it connectivity of mineralisation driving the design of stopes or connectivity of brittle structures for estimating hydrological influx. Similarly, in reservoir flow modelling, the connectivity of extreme permeabilities is an important characteristic to be captured in a geological model (Mariethoz & Caers, 2015). Accurate reproduction of such patterns is required for successful engineering studies and is essential for the predictive ability of the model.

In the early days of industrial mining, outlines of orebodies were delineated manually on paper sections with the application of average thicknesses for volume calculations. With developments in computing technology in the 70s, computerised 2D cartographic techniques such as contouring, were developed (Glacken & Snowden, 2001; Cowan et al., 2003). With the advance of three-dimensional (3D) graphics capabilities and a new generation of geomodelling software with a graphical user interface in the 80s and 90s, geological modelling evolved into wireframing and solids creation. It became possible to represent geometric objects with curves and triangulated objects - surfaces and polygonal meshes. This process termed ‘explicit modelling’, relies on manual digitisation on several vertical, horizontal or inclined sections to be further linked and converted into triangulated solid models. The process to create a geological model can be very time consuming, specifically when dealing with large data sets.

Mineralisation of intricate geometry such as folded surfaces or volumes in the presence of dense data could not be easily described by manually digitised curves. It prompted the creation of implicit modelling routines. Implicit modelling represents a mathematical function defined continuously through space, with infinite resolution (Carr et al., 2001; Cowan et al., 2003). The function is modelled by interpolation of point values. The generation of

the wireframe representing a certain threshold is an optional step in the implicit modelling, which makes the function ‘tangible’. Categorical or continuous variables can serve as source data.

For continuous variables, like mineral grade, this function represents a continuum of the variable generated through space. For categorical variables, a signed-distance function approach (Osher & Fedkiw, 2003) is adopted: the contact points in the conditioning data are assigned values of zero, the points away from the boundary are assigned polar distance values: positive if within the category of interest, negative if outside of it or otherwise, depending on the implemented convention. The conditioning data transformed in this way is used to solve an interpolation problem (Carr et al., 2001; Cowan et al., 2003). An exact interpolator such as a Radial Basis Function (RBF), inverse distance interpolation or kriging is used in the interpolation process. The end product is a surface describing the geological contact or a triangulated wireframe, extracted from the function, representing a volume of ore. These triangulated objects are minute approximations of the function through space.

Implicit modelling has become popular in the recent decade and is being adopted extensively by the mining community, due to considerable alleviation of the workload and the auditability of the workflow. When modelling continuous variables, the implicit modelling approach performs well in the presence of dense data. When applied to categorical variables, it is invaluable in defining large-scale geological surfaces, vein-type tabular bodies and stratigraphic sequences ascribed with stationarity. The use of kriging with signed-distance functions in presence of trends, however, may lead to edge issues and biased models due to the strong non-stationarity of these functions in such conditions (Silva, 2015). Problems arise for both, categorical and continuous variables, in structurally complex situations with finely intercalated and folded geometries. Even in densely populated datasets, good performance is not ensured due to the reliance of implicit modelling on kriging which only considers linear relationships between data values, and its inability to extrapolate shapes (Silva, 2015). As mentioned

by different authors (Barnes & Gossage, 2014; Silva, 2015; Cowan, 2017) and from author's experience, uncritical use of implicit modelling tools can result in significant artefacts which require considerable intervention of explicit digitising tools. Trying to achieve better connectivity with implicit modelling, can result in exaggerated variogram ranges, morphed blown-out shapes and, as a result, overstated mineralised tonnes. A challenge of modelling patchy orebodies and finely intercalated or folded units is that quite frequently the complexity of the mineralisation does not yield itself easily to either manual digitisation, computer-aided design (CAD) or implicit modelling. Alternative modelling methods are required.

1.1.2 Advantages of stochastic modelling

Traditionally, in different branches of geosciences, linear interpolation techniques like kriging are used to produce a single value in each cell from sparse data, when local heterogeneity is replaced with a single robust average value.

A reliable reflection of the true spatial continuity cannot be established from a kriged model (Guardiano & Srivastava, 1993; Journel, 1997; Deutsch, 2013; Mariethoz & Caers, 2015) - the estimated values will have too low variance, and the variogram model, used to describe correlation between pairs of points based on sparse data will have in most cases ranges of continuity longer than those in reality. To mitigate the underestimated spatial variance, change-of-support models are applied, which have a strong reliance on assumptions of stationarity.

Having a single model as an input to engineering decisions conveys a false impression of certainty that this model is the only version of reality plausible with the available data (Caers, Strebelle & Payrazyan, 2003; Deutsch, 2013). When subjected to engineering design, such models produce a single engineering outcome or forecast, with implied high but possibly false confidence.

Stochastic modelling based on conditional simulation allows for the expression of uncertainty (Deutsch, 2013; Rossi & Deutsch, 2015). The goal is to generate an ensemble of models that form multiple representations, or realisations, of the reality. The information provided by the suite of realisations is significantly richer than a single model can allow.

In different earth science applications, stochastic modelling usually begins with the simulation of categorical variables, such as facies or rock types, and then progresses to the simulation of continuous variables, usually applying Gaussian techniques to represent such characteristics as porosity, permeability or mineral content in each of these units (Deutsch, 2013). As such, stochastic modelling can provide an integrated description of uncertainty – in the geological domaining and mineral distribution.

Implicit RBF modelling can also offer uncertainty assessment. However, multiple versions of reality produced by implicit RBF functions in the deterministic framework are based on two-point relationships and are reliant on adjustments to variogram ranges and/or orientation of anisotropy. All alternative models will produce high confidence in the vicinity of data due to two-point structures, decreasing away from the data. From the author's point of view, they are incapable of truthful representation of uncertainty in presence of structures beyond two-point statistics.

An advantage of stochastic modelling is that the change-of-support models become more intuitive. Usually performed at point support, simulated linearly averaging variables such as mineral content can be easily upscaled to different support sizes to reflect the amount of mineral resource to be recovered at a chosen Selective Mining Unit (SMU) size. Different SMU geometry scenarios can be easily benchmarked against each other. Another benefit of the stochastic change-of-support correction process is that it honours local data irrespective of conditioning data grid configuration and spacing (Deutsch, 2013; Rossi & Deutsch, 2015).

Realisations can be subjected to post-processing to quantify uncertainty. Some examples are: probability maps of geological units, maps of the most

likely to occur lithology types, E-type mean maps or maps of probability to exceed a specified cut-off grade (Deutsch & Journel, 1998; Deutsch, 2013; Rossi & Deutsch, 2015). For global content assessment, a confidence interval around the mean can be produced. The output of stochastic post-processing is easy to understand.

Multiple realisations can also serve as input to different stochastic engineering optimisation algorithms. Stochastic modelling adds significant value to high-risk projects, such as those associated with erratic grade distribution, marginal ore type orebodies, projects with multiple processing streams or stockpiling options (Deutsch, 2013; Rossi & Deutsch, 2015). A "single best" numerical model in earth sciences is no longer acceptable due to increased geological, engineering, metallurgical and remediation complexity of deposits as well as investment discretion (Caers et al., 2003).

A drawback of stochastic modelling is increased processing time and storage requirements, which, fortunately, is leveraged by advances in computational techniques and cloud processing.

1.1.3 Challenges of multi-Gaussian models and two-point statistics

With advances in stochastic modelling in the 1990s, variogram-based models gained popularity. Among the most broadly used ones were SGS, sequential indicator simulation (SIS) (Deutsch & Journel, 1998), transition probability-based methods such as TProgs (Carle & Fogg, 1996), or Markovian type categorical prediction (MCP) utilising the principle of maximum entropy (Allard, D'Or & Froidevaux, 2011). Variogram-based simulation algorithms rely on geostatistical measures between two points in space and assume that the joint distribution of modelled parameters is multi-Gaussian. Multi-Gaussian models are described by two parameters only, an expected value and a covariance function. It makes these models robust in dealing with large datasets within a theoretically sound framework.

The main shortfall of the methods based on two-point statistics is a failure of modelling complex geological structures (Renard & Allard, 2013). The

methods do not ensure the reproduction of intricate non-ellipsoidal patterns of natural spatial phenomena, as most of such phenomena are not consistent with Gaussian assumptions. Univariate Gaussian distributions of geological phenomena do not guarantee a joint multi-Gaussian distribution (Gómez-Hernández & Wen, 1998).

Another disadvantage of Gaussian models is that they maximise entropy (Journel & Deutsch, 1993) - the spatial correlation of their extreme values tends toward zero. A consequence is that variogram-based Gaussian models struggle to reproduce the so important in mining and reservoir modelling applications connectivity of the geological structures (Journel & Deutsch, 1993; Zinn & Harvey, 2003; Renard & Allard, 2013; Mariethoz & Caers, 2015). Figure 1-1 demonstrates this by comparing two realisations: one produced by the Direct Sampling (DS) multiple-point statistics algorithm (image a) showing complex phenomenon with structures beyond two-points, and a realisation by SGS (image b). The histograms and variograms of the methods are almost identical (image c and image d), however, SGS fails to capture the structural complexity of the observed field (Mariethoz, 2009).

There are other problems related to two-point statistics. Obtaining reliable variograms can often be problematic. It can be due to data sparsity, presence of much higher than second-order spatial correlations, or complex cases of non-stationarity which might not lend themselves easily to estimation by variants of kriging (Mariethoz & Caers, 2015). Limitation of the variogram model approach for describing spatial continuity has been mentioned in several publications (Journel & Deutsch, 1993; Caers & Journel, 1998; Strebelle, 2002; Caers & Zhang, 2002; Mariethoz & Caers, 2015).

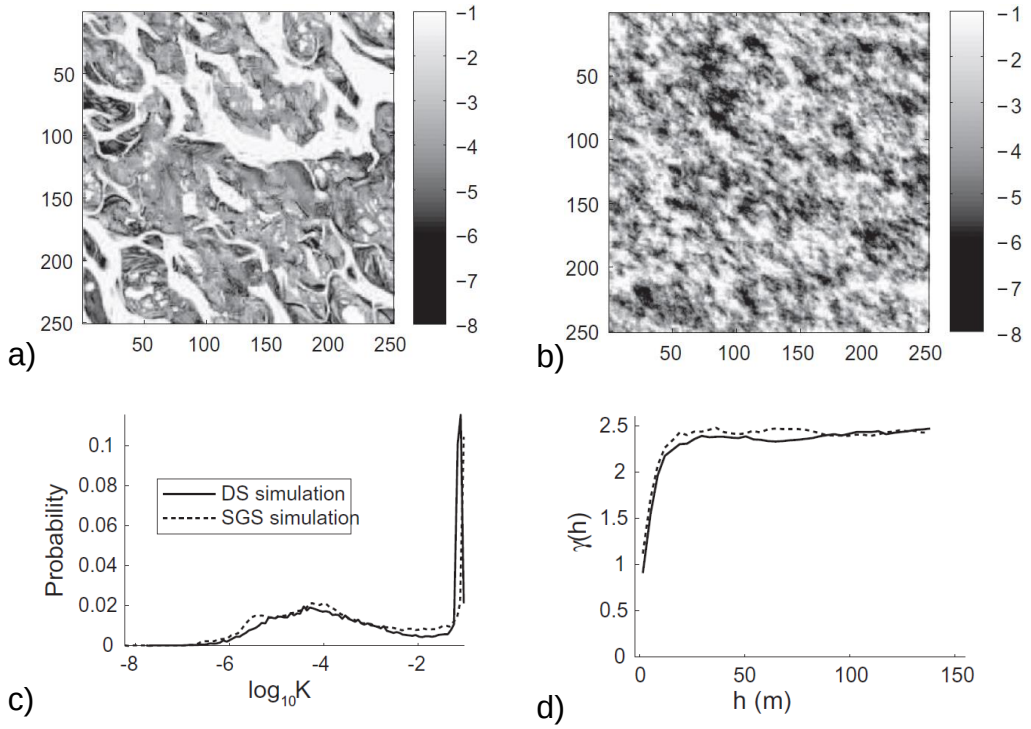


Figure 1-1. Comparison of simulated realisations obtained by the DS (a) and SGS (b). The histograms and variograms of the two fields are almost identical however the realisation by SGS is void of complex structures beyond two-point statistics. Adapted from Mariethoz, 2009

To address the limitations of multi-Gaussian models, alternative geostatistical methods started developing. Some examples include Boolean models (Lantuejoul, 2002), direct sequential simulation (DSSIM) (Soares, 2001) and pluri-Gaussian models (Armstrong, Galli, Loc'h, Geoffroy & Eschard, 2003). While being non-Gaussian, these approaches still rely on two-point parametric models (Mariethoz & Caers, 2015).

DSSIM (Soares, 2001) is an extension of SGS, where any type of distribution can be used in place of a local Gaussian conditional distribution. This allows simulating non-Gaussian variables directly without the need for a transformation of the variable. The practical advantage is that conditioning to a wide variety of data can be performed, such as block average or multi-scale data (Journel, 1999). DSSIM still relies on a variogram model, and complex spatial correlation cannot be modelled, although the connectivity may be increased (Caers, 2000; Mariethoz & Caers, 2015).

Pluri-Gaussian models (Armstrong et al., 2003) aim at performing categorical variable simulation by using multiple thresholds of a multi-Gaussian variable. The method proceeds by generating several continuous Gaussian fields using standard multi-Gaussian techniques. These fields are truncated to produce categories at different chosen thresholds. To achieve the desired spatial relationship and variability of the generated categories, the rule diagrams and the variogram models undergo modifications. The desired model is obtained by trial-and-error iteration on various combinations of parameters. Performing hard-data conditioning of pluri-Gaussian models cannot be accomplished directly and utilises a Gibbs sampler for this purpose (Armstrong et al., 2003). Such an indirect approach for modelling the desired relationships produces amorphous spatial features typical of multi-Gaussian distributions (Mariethoz & Caers, 2015).

As noted by Mariethoz and Caers (2015), even though the two-point statistics approach methodology has limitations when applied to the complex subsurface environments, these methods are robust when dealing with large amounts of conditioning data. This feature contributed to the considerable success of two-point spatial continuity models in the early stages of applications of geostatistics in carbon extraction and the environmental industry (Renard & Allard, 2013). In mining, two-point statistical algorithms are still the mainstream methods used in modelling mineralised domains and mineral content.

1.2 Research objectives

Multiple-point statistics have been gaining popularity in geoscientific numerical modelling. The application has pioneered in oil reservoir modelling 27 years ago (Guardiano & Srivastava, 1993; Strebelle, 2002), followed by applications in hydrogeological and environmental sectors (Srivastava, 2018). In these industries, realistic modelling of flow paths associated with lithologies of significant permeabilities is an important problem. It is similar to the task of truthful reproduction of high-grade

continuity in mining. Unlike reservoir modelling, precious metal mining applications do not suffer from data sparsity, making it well-suited for the MPS framework relying on a training image prior.

The purpose of this research is to explore an integrated approach to multivariate geological and geostatistical modelling of a gold deposit employing multiple-point statistics and to assess the reliability of results for quantification and classification of the Mineral Resource. The quantification of the Mineral Resource implies estimation of the mineralised tonnage, grade and mineral content, while the classification refers to establishing a degree in confidence of the estimated Mineral Resource. An emphasis in this study is placed on deriving an orebody model with well-defined structural architecture to guide the simulation.

The deposit used in this research is structurally controlled; the morphology of high-grade ore shoots is dictated by the intersection of the prevailing directions of structural geological controls. Open pit mining methods are forgiving and less sensitive to precise representation of the orebody morphology as they are less costly and usually ascribed with lower selectivity. The definition of the mineralisation shapes can be performed using classic modelling techniques. As high capital expenditures are required for underground development, a robust estimation of the true geometries of the ore is required for mine design.

The Direct Sampling MPS algorithm, developed at the University of Neuchâtel, is used to create a geological and mineralisation models. The main objective of this research is to extend the application of the Direct Sampling algorithm (van der Grijp & Minnitt, 2014) to a non-stationary geological setting, testing the appropriateness of the approach for producing a reliable and accurate orebody model and an uncertainty assessment for an underground mining method in a geologically complex, structurally folded environment.

Besides simulation of geological units, which would traditionally serve as stationary domains for further geostatistical estimation with a Gaussian-

based model, the focus of this research is to perform integrated multi-variable MPS simulation for categorical and continuous variables representing the spatial gold distribution.

The stochastic simulation framework is aimed at assessing uncertainty in the main parameters important for describing a Mineral Resource: mineralised volume, tonnage determined by the stochastically derived density, gold grade and contained metal.

This study serves as a validation of the algorithm performance in a complex non-stationary geological setting. A streamlined framework for further multivariate MPS simulation is defined with possible application in Mineral Resource classification.

1.3 Identified gaps

Despite the long history of multiple-point statistics, the method has not gained popularity in mining. Real-case studies have been mostly limited to categorical variable models such as lithology (Honarkhah & Caers, 2010; Comunian, Renard, Straubhaar & Bayer, 2011a; Rezaee, Asghari, Koneshloo & Ortiz, 2014; van der Grijp & Minnitt, 2014; Hong, Oh & Cho, 2018; Dagasan, Erten, Renard, Straubhaar & Topal, 2019). Continuous variable simulations have been implemented to oil reservoir or hydrogeological applications (Caers et al., 2003; Zhang, Bombarde, Strebelle & Oatney, 2006; Mariethoz & Kelly, 2011; Honarkhah & Caers, 2012; Straubhaar, Renard & Mariethoz, 2016), remote sensing (Mariethoz, McCabe & Renard, 2012) and topographic elevation (Pirot, Straubhaar & Renard, 2014; Dagasan, Renard, Straubhaar, Erten & Topal, 2018) to name a few. During the period of conducting the current research, no mentioning could be found with regards to the MPS simulation of continuous variables in a context of hard-rock mining for large 3D models.

A few gaps have been identified and have the potential to benefit the process of geological and mineralised content modelling, specifically in

structurally complex environments where conventional techniques often do not produce satisfactory results. Some of the gaps to be addressed in this research are:

- An ability to achieve better modelling of the continuity of the extreme values. While the continuity of mineralisation in its entirety is important for engineering applications and financial modelling, it can be adequately modelled by conventional Gaussian and two-point statistical approaches. The extreme values form a small portion of a distribution, are scattered spatially and associated with intricate networks of controlling geological structures, specifically in the case of gold deposits. Modelling of such structural complexities with classical techniques is challenging;
- Reproducing finely intercalated and folded geomorphology;
- Modelling of complex spatial structures, beyond two-point statistics, which are a result of multiple structural controls on the mineralisation;
- Designing a workflow that is insensitive to data spacing and allows robustness in the presence of sparse data or even in the absence of data;
- Ensuring reproduction of overall global statistics as indicated by drillhole proportions of categorical variables such as lithologies or mineralised material at specified cut-offs. This approach is based on an assumption that the proportions observed in the drillholes are locally representative of the orebody which is a reasonable assumption to make when the sampling grid is not biased and an application of declustering techniques is made to mitigate possible oversampling of low- or high-grade areas;
- Utilising conceptual descriptions of reality in conjunction with structural orientation readings to describe the structural architecture of the deposit in shortage of data;
- Developing a workflow which allows partial update of the model, when previously simulated areas are ‘frozen’ and used as conditioning data in the follow-up simulation. The updated areas should exhibit a smooth and natural transition between the ‘frozen’ and new areas with honouring of

the spatial structures. In cases of an advancing production interface, it should be possible to ‘freeze’ the well-known production portion of the deposit and use it as a training image, and at the same time as conditioning data for the simulation;

- Reproducing multivariate relationships;
- Considering the possibility of conditioning by secondary exhaustively known variables. Stationarity does not have to be explicitly defined but rather described by an auxiliary exhaustively known variable;
- Developing a framework which yields itself easily to machine learning (ML) with a minimal specialist’s interference limited to the definition of a prior conceptual model of reality;
- Defining a stochastic framework to conduct uncertainty assessment while allowing all of the above.

1.4 Methodology overview

The Direct Sampling algorithm is applied to a multivariate dataset, comprising of both categorical and continuous variables. A dataset has been chosen to satisfy the following requirements:

- The study area is densely informed to enable definition of a high-quality training image, which would also serve as a best possible ‘true’ model for validation of the algorithm performance;

- Several correlated variables are required to perform multivariate simulation, they have to be defined over the TI, however, not all of them need to be present as conditioning data during the simulation, e.g. co-location of the different variables in the conditioning grid is not a pre-requisite for a successful simulation of these variables, as long as a training image describing them is available. Examples in case of a mineral deposit are the mineral of interest, controlling lithofacies, a proportion of bearing quartz-veining, attitude of folding, or alteration indicators. Examples for

a geotechnical study could include rock types, structural architecture of the domains, rock mass rating and fracture frequency;

- The geological environment will have complex structural controls with multi-phase folding, affecting the spatial distribution of the variables. The units might show banding and intercalation of the magnitude coarse enough to be of significance for modelling. The size of such units must be comparable to the size of blocks on which the ore boundaries are defined, and engineering/processing decisions are made;

- The geometry, attitude, and continuity of lithologies need to display sufficient complexity to motivate the application of MPS simulation rather than manual deterministic interpretation. Two-point statistical simulation or implicit interpolation approaches would be considered as inadequate to define the distribution of the variables in space.

Such a dataset would allow for testing of the performance of the DS algorithm with rotational field variables exhaustively defined over the training image and the simulation grid, described by azimuth and dip (for linear features, such as foliation), or strike, dip and plunge (for planar features, such as orientation of ore shoots). When the character and orientation of folding differ between those observed in the training image and simulation grid, the orientation potential fields are treated as secondary variables to guide the simulation of the main ones in a multivariate framework (Mariethoz & Caers, 2015).

Briefly, the methodology is summarised in the sections below.

1.4.1 **Preparation for simulation**

- Define an exhaustively informed volume. Build a ‘true’ multivariate model over the volume. All relevant geological

information on primary and secondary/auxiliary variables should be incorporated, such as logging for lithology and structural readings of the diamond core, grade control drilling and pit mapping information if available;

- Isolate a portion of the ‘true’ model to be used as a training image in further simulations. The remaining portion of the ‘true’ model is to be used for the validation of simulations.

The importance of deriving a representative ‘true’ model cannot be over-emphasised. It has been mentioned by different authors (Chugunova & Hu, 2008; de Vries, Carrera, Falivene, Gratacos & Luit, 2009; Mariethoz, Renard & Straubhaar, 2010; Straubhaar, Renard, Mariethoz, Froidevaux & Besson, 2011; Comunian et al., 2011a; Comunian, Renard & Straubhaar, 2011b; Tahmasebi, 2018). Considerable effort has been dedicated in this research to the process of building the ‘true’ model of the deposit.

1.4.2 Simulation work utilising the Direct Sampling algorithm

A general approach to performing a simulation workflow is presented below:

- Perform unconditional simulation to validate the appropriateness of the main controlling parameters of the DS algorithm for lithology and gold grade simulation;
- Extract a subset of nodes from the ‘true’ model which will serve as conditioning data in further simulations. The nodes are located on a sparsely spaced grid, erring on the conservative side for delineation of a Mineral Resource, to arrive at an acceptable result with regards to associated uncertainty. Several such subsets can be created to test for the uncertainty associated with placement of drilling grids;
- Perform conditional simulation for lithology and gold grade, using data on the sparsely informed grid;

- Perform conditional simulation for lithology and gold with additional conditioning to block data to ensure honouring of the localised gold estimates. Kriging estimates based on sparse information could be conventionally used as a source of block data. In the current research, however, to avoid introducing another algorithm, associated with derivation of block values, as a contributor to the overall uncertainty and to limit the latter to the DS algorithm performance, the upscaled ‘true’ model is used.

1.4.3 **Validation**

The validation of the simulation results followed the steps below:

- Perform visual validation of the simulations ensuring main structural controls are honoured. The visual cortex still serves as the primary validation tool with the ability to identify closeness of reproduction of complex relationships present in the ‘true’ conceptual model. Visual validation is used as the main time-saving tool for validation in this study. Only when a simulation has passed this test it is subjected to more rigorous statistical validation routines.;

- Validate representativity of the simulated model for the reproduction of lithology and gold statistics, histograms, boundary analysis, variograms and scatterplots for the block-conditioning values;

- Perform a reconciliation study within the simulation volume between the simulation based on the sparse dataset and the ‘true’ model with the associated spread of uncertainty. The reconciliation allows a successful ensemble of realisations to be used for consideration of the Mineral Resource classification and quantification in the unknown portion of the deposit.

Tools specific to MPS simulations have been proposed by some authors (Boisvert, Pyrcz & Deutsch, 2008b; Mariethoz & Caers, 2015): validation of global and local trend statistics, multiscale histograms, fraction in tails, multiple-point density function difference, missing or zero bins for categorical variables, cluster-based histogram of patterns, continuity functions, and coherence maps to avoid producing verbatim, or pixel-for-pixel, copies of patterns. While considered of great importance, using these tools is beyond the scope of the current research. It is a belief of the author, there is still a long way to go before currently available MPS validation tools can surpass the specialist's judgement in the complexity of mathematical calculation routines and quality of validation.

1.4.4 **Uncertainty assessment in the Mineral Resource**

Public reporting of Exploration Results, Mineral Resources, and Mineral Reserves are set with minimum standards, recommendations and guidelines by reporting codes for different jurisdictions. They differ in some details. The following is an extract from the South African Code for the Reporting of Exploration Results, Mineral Resources and Mineral Reserves (SAMREC Code) defining Measured, Indicated and Inferred categories.

"A Measured Mineral Resource is that part of a Mineral Resource for which quantity, grade or quality, densities, shape, and physical characteristics are estimated with confidence sufficient to allow the application of Modifying Factors to support detailed mine planning and final evaluation of the economic viability of the deposit. Geological evidence is derived from detailed and reliable exploration, sampling and testing and is sufficient to confirm geological and grade or quality continuity between points of observation" (SAMREC, 2016:22).

"An Indicated Mineral Resource is that part of a Mineral Resource for which quantity, grade or quality, densities, shape and physical characteristics are estimated with sufficient confidence to allow the application of Modifying Factors in sufficient detail to support mine planning and evaluation of the

economic viability of the deposit. Geological evidence is derived from the adequately detailed and reliable exploration, sampling and testing and is sufficient to assume geological and grade or quality continuity between points of observation" (SAMREC, 2016:21).

"An Inferred Mineral Resource is that part of a Mineral Resource for which quantity and grade or quality are estimated on the basis of limited geological evidence and sampling. Geological evidence is sufficient to imply but not verify geological and grade or quality continuity. An Inferred (Mineral) Resource has a lower level of confidence than that applying to an Indicated Mineral Resource and must not be converted to a Mineral Reserve. It is reasonably expected that most of Inferred Mineral Resources could be upgraded to Indicated Mineral Resources with continued exploration" (SAMREC, 2016:20).

The different public reporting codes are not prescriptive in the methodology of Mineral Resource classification (Owusu & Dagdelen, 2019), a number of different approaches are being used in the industry, among them are geometric methods and methods based on calculating confidence intervals on the tonnage, grade and metal content within specified production periods either through kriging variance or geostatistical conditional simulations. An approach adopted widely in the industry is to use ±15% relative precision at a 90% confidence interval rule: the drillhole spacing used for an Indicated Mineral Resource must allow to predict tonnes, grade and metal for an annual production volume within the accepted threshold, and for a Measured Mineral Resource, the same precision is expected to be achieved on a quarterly, or sometimes, monthly production volume (Verly, Postolski & Parker, 2014; Owusu & Dagdelen, 2019). The rule is relaxed for the Inferred category, when the available data is considered inadequate for assessing confidence intervals and geological understanding and experience gained with similar deposits should be considered instead.

In the current research, the commonly used criterion for the estimated metal content to be within 15% error with 90% confidence for a volume equivalent

to an annual mining production (Verly et al., 2014) is adopted to test the suitability of classification for an Indicated or Inferred Mineral Resource. The test to pass the criterion for delineation of the Mineral Resource is performed for a drill spacing pattern of 40 m x 40 m. Appropriate mining method assumptions are taken into account. The choice to perform the study for the Inferred and Indicated over the Measured Mineral Resource category is made as classical geostatistical techniques are normally robust enough to define the Measured Mineral Resource with sufficient accuracy and precision. It becomes more challenging for the material categories of poorer confidence, which form a significant portion of publicly reported Mineral Resource of any mining company and are important for economic assessment and investor confidence in a mineral project.

Reporting in the current research is done for 15% error and 90% confidence limits, for both, open pit and underground scenarios. The extent of the simulation grid was defined to cover an annual production volume used for classification by an underground mining method. The same simulation grid extent was used in a comparative analysis for tabulating the material (tonnes, grade and metal) by an open pit mining method, to understand an increase in risk when transitioning to the deeper mining method. It is expected that the overall uncertainty is significantly reduced in a larger production volume, associated with the open pit mining (Verly et al., 2014).

The underground Mineral Resource reporting is performed after subjecting the simulations to the Mineable Shape Optimiser (MSO) algorithm, available as part of Datamine© software. The method is based on an annealing approach to arrive at an optimal stope design to meet stipulated geometric, geotechnical and economic constraints (Alford, Brazil & Lee, 2007; Datamine Product Help, 2019).

The reconciliations for each mining method are done comparing the simulation results vs the ‘true’ model with global evaluation of resulting uncertainty.

The proposed method is alike to a supervised machine learning approach. Being divided into supervised and unsupervised, both methods adhere to a three-step framework: a process of training using available data, testing of the predicted outcome, and application of successful parameters (Dey, 2016; Samson, 2019). The difference in the two methods comes with testing of the results: during a supervised approach, the solution to the problem is known and the algorithm performance is tested and validated against it. The validated results can be used to provide an answer to problems with similar phenomenon characteristics where the solution is not known. Applying this supervised approach in the study relies on a strong assumption, which is the basis of any MPS simulation, that the geological reality in the 'true' model is conceptually representative of the area to be simulated. Stationarity seldom exists in nature, more so when dealing with gold mineralisation. In practical terms, the algorithmic approach adopted in the study is well equipped to address non-stationarity in different ways which will be described later in Chapter 3.2.4.

The overall focus of the proposed methodology is placed on exploiting a full geological understanding of the deposit and making the modelling and estimation workflow inviting to geologists.

A more detailed description of the employed algorithms is presented in Chapter 3.

1.4.5 List of the algorithms and software utilised in the research

The MPS simulations were performed using the Direct Sampling multiple-point statistics algorithm developed and patented at the University of Neuchâtel, Switzerland, the patent No: US 8682624. The DS algorithm is distributed commercially by Ephesia Consult SA, Switzerland. It is also incorporated in Stanford Geostatistical Modelling Software (SGeMS) which is an open-source computer package for solving problems involving spatially related variables. In the current study, the DS was run from an executable in command prompt mode.

In addition, the following software has been utilised during this research:

- Datamine© 1983-2018 Datamine Corporate Ltd, version 1.4.132.0, used for performing kriging specific to directions and for utilising the Mineable Stope Optimisation algorithm;
- Geovariances Isatis©, used mainly in boundary analysis and calculation of experimental variograms (no variogram modelling has taken place);
- Seequent Leapfrog Geo©, for all visualisation, geological interpretation and modelling;
- Snowden Supervisor©, for statistical analysis of the composited drillhole data;

Data formatting and manipulation in preparation for simulation, as well as post-processing of results for uncertainty, were done using Python scripts in Jupiter Notebook 5.0.0 environment.

1.5 Key questions to be addressed

As the current research focuses on methodology, the questions to be addressed are of practical nature, with emphasis on ways to produce truthful and representative realisations of geological reality. An effective solution to such questions is pivotal to this research to assist in successful implementation of the MPS algorithm in the future. The key question is: Can a practical and effective methodology be developed and applied to model a structurally complex geological reality truthfully and, within that, the spatial distribution of the gold grades?

Some of the focus areas are:

- Approaches to producing an informative and representative training image/'true' model void of interpolation algorithms artefacts;

- Achieving optimum parameters of MPS simulation to enable reproduction of lithological interfaces;
- Simulation of gold grade in a multivariate framework and in a complex geological setting, aiming at an improved characterisation of the geology without explicit definition of stationarity;
- Developing a methodology aimed at having the least possible manual digitisation to lend itself easily to ML implementation in future;
- Achieving truthful definition of geostatistical variables and/or mineral content characterisation at point support which can be used as base for uncertainty assessment at larger support sizes, relevant for engineering decision making.

1.6 Research hypothesis

Thesis statement: a novice integrated approach using the Direct Sampling MPS algorithm can lead to better quantification of the Mineral Resource in a structurally complex geological environment.

For the purpose of this research, the quantification of performance and classification of the Mineral Resource is aimed to approximate the 90% confidence limits with 15% precision on estimated tonnes, grade and metal for a volume equivalent to an annual mining production tested for a widely spaced drilling pattern, by an underground mining method. A drilling pattern of 40 m x 40 m has been chosen to test the performance of the MPS for lower confidence settings, as conventional estimation techniques are normally robust in estimating the Mineral Resource of higher drilling confidence. With widely spaced Mineral Resource delineation drilling, the challenges become more pronounced when modelling thin morphologies or marginal orebodies.

It is believed that using the DS algorithm can prove valuable for simulating complex non-stationary multivariate settings, which cannot be successfully modelled by two-point stochastic approaches, or deterministically. The workflow developed is aimed at locally representative and geologically

plausible definitions of spatial variability when assisted by multivariate simulation. The reproduction of structures, which is important in many applications, is expected to improve. Examples of such structures are shapes of high-grade ore shoots along complex structural controls and thin intercalation of lithologies attributed with differing densities.

The multivariate framework is anticipated to eliminate the need for explicit definition of stationary domains.

It is envisaged that block data conditioning in the DS, not limited to regular geometric shapes, would allow more control over the simulated fields and would open an opportunity to incorporate data at different support sizes. One of the examples would be the allocation of geometallurgical properties to different geological domains or blocks of irregular shapes.

The proposed framework is aimed at limiting specialist intervention in the modelling process when translating the geological understanding into a well-informed ‘true’ prior model yielding to application in ML and ultimately a path to artificial intelligence in Mineral Resource estimation.

The visual nature of the MPS framework is expected to create scope for the involvement of specialists from different fields. In MPS, more importance is placed on a sound prior model, where rigorous rules and geological understanding are required to convey the conceptual model of natural phenomena. With this approach, the uncertainty quantification can be transferred from being exclusively in the domain of geostatisticians and mathematicians to other specialist disciplines.

1.7 Research scope limitations

The summary below represents a non-exhaustive list of limitations of the scope of this research:

- This research does not attempt to cover any issues related to logging, sampling, bulk density data collection or any data quality issues for neither estimation nor geological interpretation. The

uncertainty associated with sampling, analytical or overall data quality has not been taken cognisance of.

- No variogram modelling has been undertaken in this research. The variography was limited to the calculation of the experimental variograms.
- The dimensionality of applied quantitative validation routines has not extended beyond two-point statistics, higher dimensionality validation can be attempted in future. As the ‘true’ model is a main modelling assumption of the MPS approach, it should be used to the greatest extent possible for validation.
- No benchmarking against alternative geological or geostatistical models has taken place. No comparison to discrete Gaussian models was done in this study. Comparisons were drawn vs the ‘true’ model (its portion over the volume being simulated), the TI (representing a part of the ‘true’ model used to extract patterns) and conditioning data. It has to be noted, that the possibility of using the generated ‘true’ model beyond the extents of the orebody under consideration is limited. Moving into a neighbouring orebody, with similar host rocks, controls on mineralisation and alteration assemblages, as far as a 100 m away might necessitate a generation of another ‘true’ model, due to high non-stationarity in the overall deformational environment, intrusive network and structural architecture of the terrain.
- The author acknowledges that while the construction of a ‘true’ model should rely as little as possible on any modelling assumptions or approximation and be data-driven, the methodology followed in this research still contains approximations, such as interpolation of the gold grade values between the dense production data by kriging, or migration of the conditioning data values to the closest grid centroids (which is a practice common to all conditional simulation algorithms), and assignment of prevailing lithologies when compositing drillholes from 1 m to 2.5 m support. The latter was done as a compromise to

overcome hurdles of processing in order to simulate a grid sufficiently large for a definition of an annual production volume. While such approximations can be minimised in the future MPS workflows not to compromise the “truth” (with increased computing power it is possible to create true models at a high resolution of grid nodes, in line with a nominal sampling length), it is impossible to fully bypass modelling assumptions in any Mineral Resource estimation technique. While the ‘true’ model content will never be known at a point support, only on a global scale if the modelled block has been fed to the plant exclusively, the MPS workflow allows to incorporate the valuable dense production data as well as the specialist knowledge to the largest extent possible.

- No assessment of processing, metallurgical, legal, environmental, infrastructural, marketing or other modifying factors has taken place in the process of evaluating the prospects of eventual economic extraction. It has been limited to the application of mineable stope designs to the modelled mineralisation for generation of approximate production stopes.
- No parameter uncertainty has been attempted for the application of the MSO algorithm. The considerations deemed reasonable from the mining engineering point of view in the current geological settings have been adopted. As such, the large-scale uncertainty considered through the application of the MSO algorithm is likely understated.
- No spatial assessment of uncertainty has been attempted, only global evaluation for a chosen volume of annual production satisfying predefined mine design parameters. While the author understands that the local assessment of the method accuracy and precision can add a valuable dimension to its applicability in mining engineering, it is considered to be beyond the scope of this research.
- While the global assessment of the “Indicated” Mineral Resource tonnage, gold grade and contained metal has been done in line with the industry practice (Verly et al., 2014), at 15% error threshold and

90% confidence interval for a volume equivalent to the annual production by an underground mining method, the volume of the annual production has not been explicitly mentioned in this research due to confidentiality reasons.

- A drilling pattern of 40 m x 40 m has been chosen arbitrarily for the classification of the Mineral Resource. Drill spacing scenarios of different configurations should be attempted to shortlist the pattern most optimal for the current geological settings and the chosen confidence limits. This is considered a valuable extension to the current research, which can be attempted in future.

The following conventions are used through this research:

- Local mine grid coordinates are used for all geographic references concerning the gold deposit under consideration. In cases where a geographic coordinate system has been used, an explicit reference is made.
- The variables used in the simulation are considered at the same support.
- The gold values might have or might have not been linearly modified. The chosen mining cut-off values do not necessarily represent the cut-off values used for the deposit under consideration.
- All the references to Indicated or Inferred Mineral Resource as applicable to the current research do not bear relevance to any Mineral Resource categorisation approach adopted by AngloGold Ashanti, therefore they are given in inverted commas further in the text, e.g. “Indicated” or “Inferred” Mineral Resource.

The views and findings expressed in this research in no way represent the views or accepted estimation techniques currently in practice by AngloGold Ashanti.

1.8 Thesis structure

The research structure overview is briefly presented below:

Chapter 1 RESEARCH BACKGROUND

In the first chapter, a background of the research is presented, discussing the challenges faced when modelling categorical and continuous variables for geostatistical estimation. The advantages of stochastic approach to modelling are presented. The discussion leads to the identified gaps, research objectives, and hypothesis, with the methodology overview and research questions. The scope limit is defined.

Chapter 2 LITERATURE REVIEW

In Chapter 2, a detailed literature review is presented describing the history of multiple-point statistical simulation. This chapter starts with an explanation of basic generic steps common to the MPS framework. An overview of different MPS algorithms is presented, and challenges faced in the application of these algorithms are mentioned, presenting the reader with a background of the motives to resort to the Direct Sampling multiple-point statistics methodology.

Chapter 3 THE APPROACH TO PERFORMING STRUCTURALLY COMPLEX SIMULATION AND RISK QUANTIFICATION

In Chapter 3, the tools used to conduct the research are described. The majority of the chapter is dedicated to a detailed explanation of the Direct Sampling MPS methodology, mathematical notations used, and the advantages of the DS over other MPS algorithms in the way different challenges are handled. The parameters required as an input for the DS are explained.

The chapter closes with a description of an algorithm used for generating underground mineable stopes. Since free and independent selection is not possible during underground mining, the Mineral Resource in this mining

scenario is usually reported within mineable stope shapes which have been generated to satisfy a selection of engineering criteria.

Chapter 4 MODELLING A STRUCTURALLY COMPLEX ENVIRONMENT WITH THE DIRECT SAMPLING MULTIPLE-POINT STATISTICS

Chapter 4 deals with the practical steps adopted to test the applicability of the DS multiple-point statistical simulation to a complex environment. The process begins with a description of the geological setting, the data set used, and exploratory data analysis. Considerable attention is given to the description of the process adopted for the generation of a training image (termed ‘true’ model throughout the thesis) as the training image lies at the heart of the MPS modelling approach.

To appreciate the uncertainty in the final simulation results, understanding the inherent uncertainty associated with the placement of boreholes was considered to be of great interest to the author. To assist with this, several pseudo drillhole sets of the same grid geometry were extracted from the ‘true’ model. The resulting uncertainty in the lithology proportions, the mineralised material proportions and in the gold grade were analysed.

The chapter proceeds with the description of the simulation process and results for the lithology simulation, the mineralised zone, and the gold grade. Difficulties encountered are described, and the ways these were dealt with listed. For each of the simulated variables, the tested parameters are presented.

The chapter also provides details on the understanding of risk in the Mineral Resource: the tonnage, average grade and metal quantity. Only global uncertainty is assessed within a volume which slightly exceeds an approximate volume of annual production by an underground mining method. For the thesis, the chosen benchmark to be fulfilled for the classification of the Mineral Resource is a 15% error on the contained metal for 90% of cases at a 40 m x 40 m drillhole spacing.

Chapter 5 FINDINGS AND ANALYSIS

This chapter presents the summary of findings and contribution to new knowledge obtained in carrying out this research. Analysis of the outcomes and conclusions are presented. Recommendations for further research are given.

1.9 Chapter summary

Chapter 1 provides a perspective of challenges faced when performing geological modelling and estimation. A short review of the historical and currently used methods is given, and gaps identified.

In the next chapter, a more detailed overview of the multiple-point statistical methods is presented. Challenges common to most of these methodologies are discussed.

2 LITERATURE REVIEW

2.1 Introduction

The current chapter begins with a brief overview of the origins of the multiple-point statistics and the recent research in the field. For the reader to grasp the principles of MPS and details of this research, generic basics common to different MPS algorithms are presented, followed by a brief description thereof. Shortfalls and challenges faced in the application of the different algorithms are summarised in preparation for a discussion in the succeeding chapter on the method adopted in the current research.

2.2 Background of multiple-point statistics

A practical need to arrive at physical realism in complex 3D models and address the issue of poor conformity of natural phenomena to Gaussian models prompted interest in explicit prior model definition (Mariethoz & Caers, 2015). André Journel introduced a beginning of a new era in geostatistics. In 1985, he highlighted that a probabilistic approach to natural phenomenon modelling is just one, and "possibly not the simplest language to express very simple deterministic criteria for estimation", with stationarity being a property of the probabilistic model, not of the natural phenomenon. He emphasised that the effectiveness of a modelling choice in geostatistics must be judged by its success in solving a problem, and by practically verified facts rather than by sophisticated theoretical developments. Simplicity must be given prevalence over theoretical intricacy.

In 1993, Guardiano and Srivastava formalised the principle of multiple-point statistics. They suggested introducing conceptual changes to geostatistical modelling. The first one underscored the fact that point observations might be insufficient to infer all statistical features of a phenomena, connectivity of extreme values being one of them. Any mathematical model, irrespective of its complexity, if inferred from the analysis of point data values will fail to

reproduce that property of interest (Sanchez-Vila, Carrera & Girardi, 1996; Zinn & Harvey, 2003; Mariethoz et al., 2010; Mariethoz & Caers, 2015).

The second conceptual change was to use a nonparametric tool for describing heterogeneity, termed ‘training image’. The training image was proposed to be treated as an explicit prior model (Journel & Zhang, 2006). This provides a conceptually new approach to geostatistical modelling, when, rather than being based purely on the data, the statistical model relies to a larger extent on a geologically sound training image and on the algorithm and parameters that control its behaviour (Boucher, 2007; Mariethoz et al., 2010).

The third conceptual change was to evaluate the statistics of multiple-point data events and to proceed with simulation in a sequential framework.

The means for the MPS to reproduce the subsurface is through a training image. The training image represents a conceptual explicit depiction of the heterogeneities, a specialist believes are present in the deposit (Caers & Zhang, 2002; Caers, 2011). The structural complexities of such a geological prior often cannot be adequately described by a two-point variogram model. In the MPS framework, the training image is used as a source of multiple-point relationships to create realisations by following the sequential simulation routine.

The main methodological difference of MPS is that significant modelling effort is invested in designing an explicit geological prior, rather than focusing on mathematical parameter inference. The emphasis is placed on accurate representation of multivariate relationships and complex structures with connectivity or dysconnectivity, believed to exist in the subsurface. The difference between traditional geostatistics and MPS extends beyond a step of dimensionality of the statistical events. Rather, it is determined by the fact that the MPS simulation relies on the extraction of the multivariate statistics directly from training images instead of an explicit random function model (Mariethoz & Caers, 2015). While it can be argued that the challenge of obtaining a variogram is small relative to that of obtaining a training image,

the advantage of the MPS framework is that the training image is an explicitly visual representation of the truth, believed to exist in the area to be modelled, and places the task of the model inference squarely into geologists' hands. MPS can potentially be applied to any geological environment, provided there is a training image believed to reflect the true geological heterogeneity.

2.3 MPS algorithms: generic basics

A significant number of multiple-point statistical algorithms have been developed in the last 27 years. While these differ in the various ways of defining, storing and reconstruction of patterns, a basic framework is followed in all. The main elements of an MPS simulation and the framework common to the most MPS algorithms are described below as a reference and are as defined in the DeeSse User's Guide (Straubhaar, 2019).

2.3.1 Simulation inputs

Simulation grid. A primary shape for the MPS simulation paradigm is a simulation grid (SG). It represents a regular cuboidal shape Cartesian grid over which simulation takes place. An SG is described by its origin, the dimension expressed in terms of grid nodes, and spacing between nodes in each direction. While the grid is presumed to be three-dimensional it is possible to make it one- (1D) or two-dimensional (2D) by setting the number of nodes to one in a relevant direction. The SG is defined at point-support and upscaled linearly for linearly averaging variables to larger supports at which decision for each parcel or SMU destination is made.

Node. It represents the finest resolution of the cuboidal shape, an elementary volume $\Delta x \Delta y \Delta z$, at which geology is described in either the simulation grid or the training image. The term can be interchangeably used with the terms 'cell' or 'pixel'.

Patch. A cuboidal collection of nodes representing a portion of a simulation grid or a training image, which serves as the smallest element of simulation in patch-based MPS methods.

Training image. It represents an explicit prior geological model with at least one variable which serves as a source of patterns and multiple-point statistics for simulation. It can be generated in different ways: by using interpolation of densely spaced data, from an outcrop observation, or as a sketch of the phenomenon by an experienced geologist. Even a core photograph can be used as a source of complex patterns in cases when scale invariance is believed to exist for the phenomenon, e.g. microscopic shearing observed in a core is analogous of macro-scale structures to be simulated.

One or more training images can be used as input to a simulation, with at least one variable defined in each TI. The geometry of a training image is described in the same way as a simulation grid. Normally, the pixel resolution of the patterns observed in a training image must correspond to the resolution of the patterns to be simulated, unless scale invariance is intended.

Set of conditioning data points. Conditioning data, also termed as ‘hard data’ in MPS is defined on point support with X, Y and Z coordinates and one or more variables associated with each point. During the MPS conditional simulation, as with any other sequential simulation routine, the hard data values are assigned to the closest simulation grid nodes prior to simulation. To ensure high precision, the resolution of the grid nodes should approximate the compositing length of the drillholes.

Local probability map. If simulation is to be conditioned by local probabilities, a local probability map can be provided, with the same geometry as the SG and a probability value for each node. Probability maps are supplied per class of values, be it integers for a categorical variable, or intervals defined for a continuous variable.

Multiple grids. While MPS simulation is performed on a point scale, the idea behind the multiple grid approach (Tran, 1994) is to be able to reproduce patterns at a large scale while preserving a small size search neighbourhood. Sub-grids of different scales are considered in the simulation grid, and simulation progresses from the coarse to the fine-scale level while a search template is rescaled accordingly.

Mask object. While a simulation grid is defined over a cuboidal volume, the domain to be modelled does not have to be. A mask is used in such a case. It represents a portion of the grid that needs to be excluded from the simulation.

In the nomenclature adopted here the following notations are used:

- the elementary unit of simulation in the SG, be it a node or a patch, is denoted x;
- y denotes a location of a node or a patch in the training image;
- A data event is denoted d_n.
- A local probability map is represented by $L(x)$.

2.3.2 Algorithm steps

While MPS algorithms do differ, the following generic steps are common to most of the algorithms (Mariethoz & Caers, 2015):

1. The conditioning data are migrated to the closest grid nodes on the SG.
2. The training image is analysed to derive and, potentially, store patterns.
3. A simulation path on the SG is defined, excluding the nodes with conditioning data.
4. Looping is performed through each node of the simulation path until the stopping criterion is met:
   4.1. A node x in the SG is defined;
   4.2. The N_x neighbourhood of x is obtained. It includes the hard data and previously simulated nodes in the neighbourhood. This set of values with their spatial configuration is termed the ‘data event’;

4.3. A conditional distribution function (cdf) of the variable $Z(x)$ is defined. The cdf is conditional to N_x and the local probability map $L(x)$. The training image is scanned for replicates of this data event. The central node values corresponding to each replicate are retained. The histogram of all central values constitutes an empirical model for the conditional pdf.

4.4. A value of the variable is sampled from the distribution.

4.5. The acceptable value is pasted in the SG.

5. End.
6. Post-processing.

The MPS algorithms vary in their approaches to the following:

- Uni- or multivariate simulation;
- Node-based or patch-based simulation;
- Ways of defining neighbourhoods;
- Ways of organising the database of patterns extracted from the training image;
- Methods of computing the conditional distributions of $Z(x)$;
- The criteria for termination of the simulation process: either on completion of the sequential simulation through all nodes of the simulation grid, or upon satisfying a pre-defined convergence measure;
- Consideration of the features at different scales via multiple grids. MPS methods can utilise different types of paths for the multiple grids, or varying approaches to defining neighbourhoods when simulating a grid of a certain node density level.

2.4 Overview of MPS algorithms

A summary of different MPS algorithms is presented below. The Direct Sampling multiple-point statistical algorithm used in this research is not described in this chapter but will be covered later in the thesis.

2.4.1 Extended Normal Equation Simulation: ENESIM

The advance of MPS started with pixel-based algorithms. The first MPS algorithm, ENESIM (extended normal equation) was developed in 1993 by Guardiano and Srivastava. It was designed for categorical variables. In the implementation, the sequential simulation proceeds from calculating the cumulative conditional density function (ccdf) by sampling the SG for the available neighbourhood information (data events) using a pre-defined template window to finding replicates of the data event in the training image. The ccdf is calculated by counting the frequency of the data event occurrence in the central node of the replicates. The value of the central node with a matching neighbourhood is pasted into the node being simulated. For every node, ENESIM rescans the whole TI, resulting in high Central Processing Unit (CPU) requirements.

Pros:

- The algorithm pioneered a new conceptual approach to stochastic geostatistical modelling.

Cons:

- Like any innovation, while laying the way forward, the approach had shortcomings, the main one being high CPU requirements due to rescanning of the whole TI for each node being simulated.

2.4.2 Single Normal Equation Simulation: SNESIM

The next MPS development, SNESIM algorithm (single normal equation simulation), reduced the demand on the CPU load. A special database structure, termed ‘search tree’, is used to store TI events, and retrieve them

during simulation (Strebelle, 2002). The TI is scanned with a search template of predefined geometry only once, and the resulting multiple-point statistics are stored in a tree-structure. In a sequential simulation framework, MPS data events are retrieved from the catalogue and the central point values are used to populate the nodes in the SG. The multigrid approach in SNESIM allows the use of small templates while ensuring good reproduction of patterns at different scales. It became the first practically usable MPS algorithm which could handle large 3D training images and large simulation grids (Caers et al., 2003; Liu, Harding, Gilbert & Journel, 2005; Harding, Strebelle & Levy, 2005; Straubhaar et al., 2011; Mariethoz & Caers, 2015).

There are a few limitations to the application of SNESIM, one being the implicit assumption of stationarity. To address non-stationary cases, either the field to be simulated has to be subjected to a traditional decomposition into a known trend and unknown residuals, or the TI has to be large enough and contain replicas of the trend (Mariethoz & Caers, 2015). Another shortcoming is that, for reproducing large-scale features large size templates are required, and the size of the search tree would grow considerably, increasing memory requirements (Strebelle, 2003; Arpat & Caers, 2007). The number of rock types and the degree of entropy (Mariethoz et al., 2010) of the training image would also aggravate the matter. The applicability of the algorithm is limited to univariate simulation of categorical variables in structurally simple cases.

Pros:

- The ‘tree’ structure for storage/retrieval of data events lightened the CPU requirements and enabled simulation over large volumes while referencing large 3D training images.
- The multigrid framework ensured improved reproduction of structures at different scales.

Cons:

- The ‘tree’ search database structure is practically usable for structurally simple cases when small templates are sufficient to capture geological complexity with only a few rock types present. With increased complexity, it becomes Random Access Memory (RAM) demanding.
- Strong stationarity assumptions of the geological settings. For non-stationary simulations, a large TI is required with repetition of patterns.
- It can only accommodate categorical, univariate cases.

2.4.3 **An Improved Parallel Multiple-Point Algorithm: Impala**

Some of the issues in SNESIM were addressed with the development of the Impala algorithm (Straubhaar et al., 2011). The search tree database for the ccdf storage was replaced with a list structure. This allowed for a reduction in the RAM requirements, and, as a result, being able to simulate using large 3D training images with multiple variables, continuous or categorical. Many different lists can be used during simulation, as well as additional information within the lists can be stored. Simulation in non-stationary settings could be tackled by using a number of training images in conjunction with a continuous auxiliary variable describing the influence of each training image over the simulation grid.

Parallelisation was achieved by calculating in parallel the ccdf at each node (Straubhaar et al., 2011), rather than simultaneously simulating multiple nodes in a realisation or several realisations at the same time (Vargas, Caetano & Mata-Lima, 2008; Mariethoz et al., 2010). The drawback of the algorithm is an increased CPU demand.

Further improvements to Impala allowed to increase acceleration by combining the tree and list structures (Straubhaar, Walgenwitz & Renard, 2013).

Both, SNESIM and Impala, employ a multiple grid approach (Tran, 1994) to enable reproduction of large-scale features while resorting to small data events.

Pros:

- List database structure lightened the burden on RAM allowing one to simulate complex multi-variate non-stationary cases, while utilising multiple training images.
- Multigrid approach.
- Acceleration is achieved by search parallelisation through the list database.
- Further acceleration is obtained in implementation by Straubhaar et al (2013) by combining ‘tree’ and list structures.

Cons:

- Increased CPU requirements.
- Simulation of categorical variables only.

2.4.4 **High-Order Simulation: HOSIM**

The HOSIM (high-order simulation) algorithm has similarities with SNESIM, the difference comes in the way it stores TI patterns. Rather than frequencies, it uses spatial cumulants in a tree structure (Dimitrakopoulos, Mustapha & Gloaguen, 2010; Dimitrakopoulos & Mustapha, 2011). The spatial cumulants represent combinations of statistical moments that allow characterisation of non-linear and non-Gaussian random variables in a tree structure. The cumulants are calculated from template data obtained by scanning the TI and a conditioning dataset. The simulation proceeds using high-order Legendre polynomials with coefficients established on the cumulants database. The Legendre polynomials are solutions to Legendre ordinary differential equations frequently used in technical fields (Legendre polynomials, 2017). The advantage of HOSIM is that it does not rely on any

distribution-related assumptions and does not require pre- or post-processing steps (Mustapha & Dimitrakopoulos, 2010).

The approach is data-driven: the conditioning data has a strong effect on the simulated realisations, while the training image is only used to fill in high-order data relations that are not possible to infer from the data. This can be regarded as an advantage while at the same time challenges a strong explicit prior model influence of the multiple-point statistical simulation.

Pros:

- Free from distribution assumptions, provide tools to characterise non-linear and non-Gaussian stationarity and ergodic spatial random fields (Dimitrakopoulos & Mustapha, 2011).

Cons:

- The algorithm performance has been reported to be deterred when complex sinusoidal patterns are found in a training image.

2.4.5 Simulation with Patterns: SIMPAT

In patch-based algorithms, the probabilistic framework is abandoned. The simulation proceeds by generating the realisation pattern by pattern rather than pixel by pixel. The focus of the simulation shifts from the reproduction of statistics to reproduction of patterns. Rather than drawing from cdf in pixel-based methods, a patch-based simulation is based on finding patterns of acceptable similarity. Satisfying patterns are retrieved from the TI and pasted into the SG while being conditioned to the hard data.

While there have been a number of patch-based algorithms developed, two examples are described below: SIMPAT (simulation with patterns) and FILTERSIM (simulation using filter scores).

SIMPAT was developed for simulating categorical and continuous variables (Arpat, 2005; Arpat & Caers, 2007). It was inspired by pattern recognition and texture synthesis literature, albeit a more complex 3D paradigm. It uses

a rectangular or cuboidal template to search for all the patterns in the training image and builds a pattern database at different geological scales. The same template is used to scan the SG for the conditioning data event during simulation. The pattern most similar to the conditioning data event is retrieved from the pattern database according to some similarity criteria and pasted onto the simulation grid. Undesirable patterns can be discarded with the use of filters.

Instead of simulating one node at a time, SIMPAT pastes a patch at each grid location. This allows for the reproduction of fine-scale structures within patterns.

Multigrid simulation is possible in SIMPAT. Unlike the classical multigrid approach (Tran, 1994), SIMPAT does not 'freeze' the coarse grid values when they are transferred to a finer grid, but rather updates them in subsequent finer grid simulations. The previously simulated value is considered when minimising the similarity distance (Arpat, 2005). Soft data conditioning, also termed secondary, indirect or auxiliary, is also possible in SIMPAT.

Pros:

- No probabilistic assumptions;
- Improved continuity of simulated patterns;
- Less computationally intensive than pixel-based algorithms, mostly in 2D.

Cons:

- Difficulty in obtaining a suitable pattern in the presence of dense conditioning data.
- Highly CPU-intensive for 3D simulations.

2.4.6 **Simulation using Filter Scores: FILTERSIM**

FILTERSIM is a pattern-based simulation algorithm (Zhang, Switzer & Journel, 2006) that allows deviation from the patterns found in the TI to

ensure the preservation of conditioning data in a case of event mismatch. During the preparation step, the TI is scanned with a template of fixed size. For each template location, several local statistics are recorded. These statistics result from specific linear filters applied to the template data values. They describe local and directional mean levels, gradients, and curvatures present in the template. The filters provide a low-dimensional representation of the template multiple-point statistics and carry out a task of dimensionality reduction (Tahmasebi, 2018). Data events of high dimensionality are summarised into small vectors, which are classified into groups each represented by an average value.

During simulation, a predefined similarity distance is used to find a training class closest to the data event. A training pattern is retrieved from the class and patched in the simulation grid, honouring the conditioning data nodes. The inner part of the pattern is frozen and added to the conditioning data in the subsequent stages of simulation. The outer part of the simulated patch is added to the soft data conditioning allowing the corresponding grid points to be re-simulated. The algorithm is applicable to categorical and continuous variables. The filter-based approach results in more efficient and fast processing.

A problem with filter-based algorithms is that they may result in realisations devoid of rich heterogeneous patterns expected. It can be because these algorithms prompt the most similar but not necessarily the most suitable pattern to be selected randomly from its prototype. Filters obscure the influence of extreme conditioning values. The realisations can be affected by clustered data.

In patch-based algorithms, unmodified patterns, as stored in the template database, are used during the simulation. Any non-stationary elements present in the pattern will manifest themselves in the final simulation (Honarkhah & Caers, 2010).

Pros:

- In comparison to SIMPAT, improved speed as a result of using filters.
- Ability to reproduce complex structures and connectivity.

Cons:

- Deteriorated quality of complex pattern reproduction in comparison to SIMPAT. The suitability of a chosen pattern in each simulation iteration is not guaranteed, as the pattern is selected randomly from the respective filter class (Tahmasebi, 2018).
- Realisations may lack heterogeneity and complexity of structures due to the reduced impact of extreme values and clustering.
- Problems with conditioning to dense hard data.

The framework of FILTERSIM experienced a few improvements (Honarkhah & Caers, 2010). For example, to account for the scale dependencies of the geological processes, Dimitrakopoulos et al. (2010) use a discrete wavelet transform of a training image instead of using filter scores.

In general, patch-based algorithms are efficient when dealing with continuous properties. They improve the simulation efficiency and quality of pattern reproduction. However, the efficiency in replicating the spatial continuity is often offset by deterioration in hard data conditioning (Hu & Chugunova, 2008), specifically in cases of densely populated data grids, be it hard or soft data, such as geophysics. A task of calculating similarity distance between the patterns of heterogeneities in the TI and the model and between all the conditioning data variables becomes a hurdle (Hashemi, Javaherian, Ataee-Pour, Tahmasebi & Khoshdel, 2014; Rezaee & Marcotte, 2016). Besides using a pattern database built from the TI, the SIMPAT and FILTERSIM algorithms also incorporate the multiple grid approach.

Among other patch-based MPS are cross-correlation simulation function (CCSIM) and conditional image quilting (CIQ). These algorithms do not use multiple grids or databases.

2.4.7 Cross-Correlation Simulation: CCSIM

An improvement to address high RAM- and CPU-requirements was proposed in the CCSIM, cross-correlation simulation (Tahmasebi, Hezarkhani & Sahimi, 2012). This simulation algorithm consists of pasting overlapping groupings of pixels of cuboidal shape along a raster path while minimising a cross-correlation function with the region of overlap.

During conditioning, an approach based on recursive template splitting is applied in CCSIM. It allows for improving the accuracy of conditioning while maintaining continuity of the features. The method makes use of the acceptance threshold, which defines an ensemble of patterns that preserve the structures observed in the TI.

The CCSIM algorithm offers significant improvement in speed of simulation and quality of pattern reproduction. Even small data events can be used to identify the most similar pattern and produce good results, as the pattern selection is based on the region of overlap. The overlaps are usually smaller than data events, resulting in CPU reduction (Tahmasebi et al., 2012). The CPU demand is about 20-30 times lower in comparison to FILTERSIM. A number of input parameters are available to balance the quality of simulation and CPU usage: template size, the size of TI, size of the overlap region and acceptance threshold.

Pros:

- Computationally, the similarity distance calculations based on a cross-covariance function are more effective than those based on Euclidean similarity distance. Together with the use of overlap regions which introduce only a small portion of the data event into the similarity measure computing, it results in reduced CPU in comparison to other pattern- or pixel-based MPS algorithms.

- Being a raster-based algorithm, CCSIM yields itself easier to parallelisation than the algorithms based on random paths.
- Good reproduction of complex curvilinear features.

Cons:

- The algorithm is less sensitive to the patterns observed in the hard data and more reliant on the patterns present in the training image. This can be regarded as an advantage (Tahmasebi et al., 2012), but, might need to be approached with care in non-stationary cases, which are more usual than not, specifically, in structurally controlled precious metal mineralisation environments.
- Like other raster-based algorithms, it is susceptible to simulation artefacts, which can be more pronounced in 3D. Large patches (over 50 nodes in 2D) tend to cause an identical copy (Mahmud, Mariethoz, Caers, Tahmasebi & Baker, 2014).

2.4.8 **Conditional Image Quilting: CIQ**

Conditional Image Quilting (CIQ) is an example of a patch-based simulation method inspired by developments in computer graphics, where the objective is to generate large amounts of different textures (Mahmud et al., 2014). The process, termed ‘texture synthesis’, refers to an algorithmic approach to construct a large image from a small example utilising its textural content. The generated images appear natural, they are stochastic and non-repetitive.

Previously described patch-based MPS algorithms search for patches that fit together with minimal error. The CIQ has been inspired by another texture synthesis algorithm, Image Quilting (Efros & Freeman, 2001), which adjusts the geometry of the patches such that the overlap error is reduced. The main shortcoming, however, is its inability to condition to hard data and applicability to 2D cases only.

CIQ aims at minimising an error between the common area of simulated patches while pasting the overlapping boxes of pixels, subjected to an optimal cut through the area to ensure good quality of conditioning. The main improvement in comparison to other patch-based MPS algorithms is maintained continuity of patterns and reduction in artefacts.

Pros:

- Like other patch-based MPS simulation algorithms CIQ ensures better connectivity of the structures than pixel-based algorithms.
- Improved computational performance compared to other MPS methods, with a 50-fold reduction in computational cost for large 3D cases (Mahmud et al., 2014).

Cons:

- Less efficient when honouring dense conditioning data in comparison to pixel-based methods.

As highlighted by Mariethoz and Lefebvre (2014), the texture synthesis methods employed in computer graphics share a purpose similar to MPS – to produce realisations with pattern morphologies similar to a reference model within a stochastic framework. Alike the MPS, texture-synthesis methods accommodate user-defined non-stationarity supplied as control maps when simulating the underlying texture exemplars. The two fields can benefit each other: the texture synthesis algorithms ensure improved reproduction of patterns from the reference image while allowing much wanted user control. They are computationally highly effective. MPS methods have better ways to condition and are better fitted to 3D cases.

Several other methods have been developed for generating conditional realisations reproducing high-order spatial statistics of the training image. These include Markov random field-based multiple-point type approaches (Daly, 2004; Tjelmeland & Eidsvik, 2004), kernel approaches (Scheidt & Caers, 2009), simulated annealing (Peredo & Ortiz, 2011). Some authors proposed to integrate indicator probabilities with multiple-point probabilities.

Incorporation of the multiple-point statistics in a Bayesian framework under the assumption of conditional independence between the sources of information by an indicator technique was proposed by Ortiz and Deutsch (2004). Texture synthesis concepts were borrowed from the computer graphics field (Honarkhah & Caers, 2010), where Markov models are combined with sequential simulations (Daly & Knudby, 2007; Parra & Ortiz, 2011).

2.5 Stumbling blocks in the application of MPS algorithms

There are several limitations, common to all MPS algorithms, as has been identified by different authors (Arpat, 2005; Mariethoz & Caers, 2015; Tahmasebi, 2018). The main one is the dependence of the method on the stationarity of the training image. The problem is common to any simulation or geostatistical estimation technique relying on the Probability Theory. As the TI serves as a source of the conditional probabilities, it needs to be stationary (Arpat, 2005). To account for non-stationarity in the training image or in the simulation, Chugunova and Hu (2008) proposed to include additional variables in the co-simulation framework.

Derivation of a sufficiently large training image is another challenge. Using a small TI can often produce discontinuous or unrealistic structures in the simulation (Tahmasebi, 2018). A practical balance between the size and richness of the training image is required.

Another issue, encountered not only in MPS algorithms but in classical simulation tools such as SGS and SIS, is honouring the unique neighbourhood in large grid simulation. With the progression of the sequential simulation, the last nodes to be simulated have to be conditioned to all previously simulated nodes (Deutsch & Journel, 1998). Two solutions have been proposed to address the problem, namely, restricting the template size, and using multigrids. While the inference of the ccdf does become less computationally intensive when constraining the size of the

search neighbourhood, the reproduction of long-range correlations can deteriorate (Tran, 1994).

The multigrid method allows for alleviating the effect. Instead of a single resolution grid and one size template, a set of different support grids and templates are used. The simulation proceeds from the coarsest to the finest grid size. During the process, the values simulated at coarse resolution are frozen and added to the conditioning nodes in simulation of fine resolution levels. The multigrid method has shown shortcomings in reproducing complex geological continuities, due to its one-way coarse-to-fine conditioning hierarchy (Arpat, 2005).

Currently available similarity functions for algorithm performance validation are mostly based on variants of two-point criteria. Such distance functions are limited in their ability to convey the spatial information and frequently insufficient to identify the optimal patterns (Tahmasebi, 2018).

Many MPS algorithms allow incorporation of multiple training images in a simulation. Currently, professional expertise and knowledge of the deposit are required to decide on the suitability of each training image across the volume to be modelled. It will be beneficial to design methods that can assist in quantifying such a decision in a geological environment.

Other challenges encountered in MPS are simulation of continuous variables, performing multivariate co-simulation, and simulation of real-life complex 3D cases.

The MPS framework poses a paradigm shift to a classical way of modelling geology and quantifying the distribution of mineral content. It can be useful when describing a complex non-stationary environment that is difficult to formulate parametrically. MPS does not yield itself easy to classical validation tools. The prime validation criterion used to assess geostatistical simulations is visual examination. The visual cortex is still the best image recognition and processing tool. Performance and benchmarking of different MPS algorithms against each other require the development of new

validation methods, which would allow for the quantification of the differences between the realisations and the training image.

A topic related to the one above and not mentioned frequently is a lack of easily visualisable ‘wireframe’ objects to assess the validity of geological forms in 3D. Most geologists are used to visualising the geological phenomenon as deterministic surfaces and interfaces rather than clouds of pixels. A certain paradigm shift is required to accept the idea of the inherent uncertainty in a geological model that cannot always be expressed in the form of a wireframe.

Issues faced in the MPS framework are aggravated when modelling in non-stationary settings, dealing with nuggety phenomena, complex multivariate relationships, presence of non-linearly averaging variables, or simulating a 3D environment.

2.6 Summary of the MPS algorithms functionality

A comparison of the different algorithms’ performance is not a straight-forward task as it would require benchmarking using the same dataset. A high-level comparison is presented in Table 2-1, which is an extension to a summary by Mariethoz and Caers (2015) and by Renard and Straubhaar (2018). A short recap of the main features of the Direct Sampling algorithm used in the current research is also presented in the table.

While the described algorithms all have their shortcomings and advantages, the Direct Sampling shows a broad range of tools to make it quite flexible and versatile. It is suitable for simulating categorical and continuous variables, but its real value comes in simulations of multivariate cases. Most of the algorithms described above can handle univariate cases, however, the DS is the only simulation methodology at the time of writing this thesis which can be used in multivariate settings.

Non-stationarity can be modelled by subdividing the domain into multiple stationary zones. This approach is broadly used in other MPS algorithms. A

more efficient approach in the DS consists of introducing auxiliary exhaustively informed variables describing the non-stationary setting. This route makes it a very powerful tool for application in natural Earth sciences.

Besides conditioning to the point-support hard data, it is possible to provide additional conditioning constraints: in the form of global proportions or by local proportion maps - for categorical and continuous variables, or, in addition, by blocks with average proportions - for categorical variables.

The patterns present in the training image can be subjected to rescaling or rotation. While the former does not prompt for wide applicability in mining geostatistics, rotation can add significant value to the process of modelling in structurally deformed environments.

The DS can be run in reconstructive mode when parts of the simulation grid are informed, and simulation is only required over the remaining portions of the volume while honouring of the known parts is ensured (Mariethoz, 2009).

All the features described above serve in support of choosing the DS over other MPS algorithms for the purpose of this research. It is believed to be well-suited for the kind of the environment envisaged for the test work - with high non-stationarity, where the morphology of the ore shoots is a product of an interaction of a few geological controls, and geostatistical domaining in presence of sparse data cannot be easily defined by CAD-based methods or variogram-based approaches.

A more detailed summary of the Direct Sampling is provided in the next chapter.

Table 2-1. Expanded summary of MPS algorithms proposed by Mariethoz and Caers (2015) and by Renard and Straubhaar (2018)

Algorithm	Smallest simulation element	Simulation variables			Conditioning				Rotation		Non-stationarity		Repro stru differ
		Categorical	Continuous	Multi-variate	By global proportions	By local proportions maps	By block values	By connectivity path	Uniform by zone	Continuously defined over SG	Uniform by zone	By soft probabilities	Multi-grids
ENESIM	pixel	+	-	-	+	+	-	-	-	-	-	-	-
SNESIM	pixel	+	-	-	+	+	-	-	-	-	+	+	+
IMPALA	pixel	+	-	-	+	+	-	+	+	-	+	+	+
FILTERSIM	patch	+	+	-	+	+	-	-	+	-	-	-	+
CCSIM	patch	+	+	-	-	-	-	-	-	-	+	-	-
CIQ	patch	+	+	-	-	-	-	-	-	-	-	-	-
DS	pixel	+	+	+	+	+	+	+	+	+	+	+	-

2.7 Chapter summary

In this chapter, an overview of different MPS algorithms was presented. The evolvement of the MPS was traced down from its inception in 1993 by Guardiano and Srivastava to the modern-day. It is by no means an exhaustive account.

General features, advantages and shortcomings of the main-stream algorithms were given. It allowed for establishing a basis for benchmarking with the Direct Sampling algorithm, to present a motivation for using the DS and to prepare the reader for a more detailed description of the algorithm in the following chapter.

Chapter 3 will describe in detail the functionality of the DS.

3 THE APPROACH TO PERFORMING MULTIPLE-POINT STATISTICS SIMULATION OF COMPLEX GEOLOGY AND STOCHASTIC DEFINITION OF AN UNDERGROUND MINERAL RESOURCE

3.1 Introduction

The purpose of this study is to investigate a new integrated approach to multivariate geological and geostatistical modelling of an orebody employing multiple-point statistics and to test the reliability of results for classifying and quantifying the Mineral Resource. The focus of the current chapter is on the description of the tools used in the study. The research design has been covered in Chapter 1.

Two algorithms are used: the Direct Sampling multiple-point statistics algorithm and the Mineable Shape Optimiser. The DS algorithm is adopted to generate multiple realisations of geology and gold grade distribution. A detailed description is presented in this chapter, deemed appropriate to explain the algorithm usefulness in comparison to the other methods described previously. The most important features are the adoption of the concept of similarity distance, dealing with multigrid scales, with non-stationarity and conditioning to the block data. MSO is used for the purpose to report the mineralised material contained within reasonable underground mineable shapes.

3.2 The DS multiple-point statistics algorithm description

3.2.1 An overview of the Direct Sampling algorithm

The Direct Sampling adopts a principle of sequential simulation. The algorithm does not rely on any strong parametric assumptions about a natural phenomenon, but the multivariate simulation environment is anticipated to provide a geologically sound outcome. Unlike most MPS algorithms, the DS does not store the data events present in the training image in a catalogue of any form prior to simulation (Mariethoz et al., 2010).

It samples the training image directly in a random fashion during the simulation. The fact that the explicit calculation and storage of the ccdf is avoided makes the DS much faster than other MPS algorithms. Since the TI is scanned randomly, this strategy is equivalent to drawing a random value from the cumulative distribution function (Mariethoz et al., 2010).

The DS uses an expression of similarity between a data event d_n and the TI patterns, termed distance. When simulating a node x in the SG, the TI is randomly scanned for matching patterns of nodes y. A measure of matching between the two data events is determined by a distance threshold, which can be defined in the range between 0 and 1. As soon as a TI pattern that matches the data event d_n within the defined threshold is found, the value of its central node is pasted into the node being simulated. Fine-tuning the distance threshold makes it possible to deal with categorical, continuous variables and multivariate co-simulation in geological settings with complex relationships between variables. The method is appealing due to its simplicity and flexibility. Running it in multivariate mode provides an intuitive tool to manage various types of non-stationarities (Mariethoz et al., 2010).

The following notation is utilised further in the description of the DS simulation process:

- x – the current node being simulated in the simulation grid
- y – the current node in the training image which is being checked for a match;
- $z(x_i)$ – a conditioning dataset, where $i \in [1, .., N]$;
- n – the maximum number of conditioning nodes to be included in the neighbourhood as stipulated in the input to simulation;
- t - acceptance threshold for the distance measure (degree of mismatch allowed in the patterns);
- f – a fraction of the training image to be scanned when simulating each node.

The parameters n, t and f have the most significance during the DS simulation in terms of the quality of reproduced patterns and memory requirements.

During the simulation of a node x, a data event $d(x)$ centred at x is compared with patterns of the same geometry $d(y)$ found at random in the TI, until the two patterns $d(x)$ and $d(y)$ are acceptably similar.

The DS proceeds as follows (Mariethoz et al., 2010; Straubhaar et al., 2016):

1. Each conditioning data, if available in the input, is assigned to the closest node in the simulation grid.
2. A random path through the remaining nodes of the SG is defined. The path contains all the indices of the grid nodes to be simulated in a random (Strebelle, 2002) or unilateral (Daly, 2004) sequence. In a random path, the simulation sequence is determined by a series of random numbers based on a random seed. In a unilateral path, the nodes are visited starting from one side of the grid, in a raster-scan fashion.
3. For simulation of each node in the path, the following is performed:
   3.1. Determine the data event on the SG:
      3.1.1. Find the closest conditioning neighbour nodes of x consisting of the maximum n closest informed grid nodes $\{x_1, x_2, .., x_n\}$, that have been already assigned or simulated.
      3.1.2. Compute the lag vectors $L = \{h_1, .., h_n\} = \{x_1 - x, .., x_n - x\}$ defining the neighbourhood $N(x, L) = \{x + h_1, .., x + h_n\}$ of the node being simulated. The separation distances are expressed in terms of nodes. An example is shown in Figure 3-1a, where the neighbourhood of the node to be simulated consists of three lag vectors between it and the conditioning nodes, $L = \{(1, 2), (2, 1), (-1, -1)\}$.

3.1.3. Define the data event, a vector containing the values of the variable of interest made up of the maximal n nodes closest to the candidate node:

$$d_n(x, L) = \{Z(x + h_1), .., Z(x + h_n)\}$$

Equation 1

In the categorical variable example of Figure 3-1a, the data event is $d_n(x, L) = \{0, 0, 1\}$, with category 0 nodes shown in white and category 1 in black.

3.1.4. Set:

best current error $E_{cur} = \infty$,
best current candidate $y_{cur} = null$,

currently scanned fraction of the TI $f_{cur} = 0$.

3.1.5. At the commencement of an unconditional simulation, when no neighbours of x exist yet, randomly locate a node y in the TI and assign its value $Z(y)$ to $Z(x)$ in the SG.

3.2. Find a suitable pattern in the TI. While $E_{cur} > 0$, and $f_{cur} < f$ do:

3.2.1. Define the search window. It represents the internal portion of the TI, with the masked zone on its periphery equivalent to the dimensions of the conditioning event lags, as shown in Figure 3-1b.

3.2.2. Randomly draw a location y within the search window, for which:

3.2.2.1. The data event $d_n(y, L)$ is calculated. In the example of Figure 3-1c, the data event is $d_n(y, L) = \{1, 0, 1\}$.

3.2.2.2. A comparison is made between the data event in the TI and the target data event in the SG. This similarity distance is expressed as $\boldsymbol{d}\{d_n(x, L), d_n(y, L)\}$.

The error between the TI and SG data events is calculated as:

$$E = max\ (0, \frac{\boldsymbol{d}\big(d(x), d(y)\big) - t}{t})$$

Equation 2

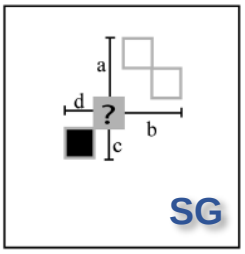
3.2.2.3. The obtained TI event $Z(y)$ as well as its distance $\boldsymbol{d}\{d_n(x, L), d_n(y, L)\}$ is stored if it is the lowest distance obtained so far for the target data event.

3.2.2.4. The distance $\boldsymbol{d}\{d_n(x, L), d_n(y, L)\}$ is compared to the acceptance threshold t, specified in the input of the algorithm.

If $E_{cur} = 0$, assign the candidate node value and exit the scan;

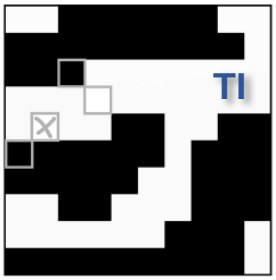
If $E < E_{cur}$, set $E_{cur} = E$ and $y_{cur} = y$.

3.2.2.5. The count of iterations in the steps 3.2.2.1 - 3.2.2.4 is compared to the maximum allowed as per fraction of the TI to be scanned f, specified in the input. The currently achieved scanned fraction f_{cur} is updated by adding the inverse of the number of nodes scanned so far in the TI. If the maximum allowed fraction has been reached, the value $Z(y)$ of the node with the lowest mismatch distance found so far y_{cur} is pasted to the $Z(x)$ (Figure 3-1e). The algorithm proceeds to the next node in the simulation path.

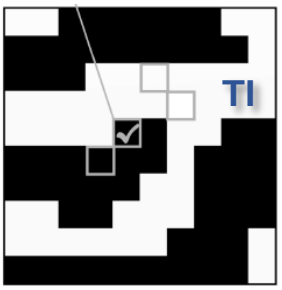


a) The data event is defined in the simulation grid. The block with a question mark represents a node to be simulated, the two white and one black block are previously simulated nodes or have been assigned values from conditioning data

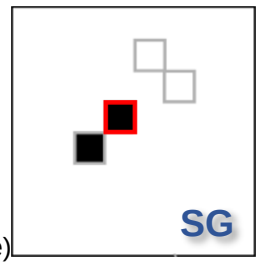
b) A search window is defined in the TI grid by using the lag dimensions a, b, c, d of the data event



c) The search window is scanned starting from a random location



d) The search is terminated once a satisfactory match for the simulation data event is found, indicated by the tick mark



e) The value of the central node of the matching data event is pasted to the node being simulated

Figure 3-1. Schematic showing neighbourhood search during the DS simulation. Adapted from Mariethoz et al., 2010

According to the definition (Equation 2), $E = 0$ only if $\boldsymbol{d}(d_n(x), d_n(y)) \leq t$. If it is satisfied, the scan is interrupted, if not then the candidate node with the smallest error achieved after scanning the maximum allowed fraction f of the TI is pasted into the SG.

3.2.2 **Similarity distance**

The main advantage which opened up the possibility of the application of the DS for simulation of categorical and continuous variables is the concept of similarity distance (Mariethoz et al., 2010). In complex 3D or continuous

cases, the data event $d_n(y)$ matching $d_n(x)$ perfectly is often not found. In such situations, traditional MPS algorithms usually drop successive data locations from the data event until a perfect match is encountered. In the DS, this practice is avoided by using the acceptable error, referred to as similarity distance $\boldsymbol{d}\{d_n(x, L), d_n(y, L)\}$. It represents the difference between two data events, one - in the simulation grid, and the other one - in the training image. When set above 0, deviations from the perfect match are allowed. In complex cases, a slight deviation from 0 is preferred, e.g. 0.05.

For categorical variables, the distance is expressed as a fraction of mismatching nodes in the data event being simulated (Mariethoz et al., 2010). This fraction is denoted a:

$$\boldsymbol{d}\{d_n(x), d_n(y)\} = \frac{1}{n}\sum_{i=1}^{n} a_i, \qquad \in [0,1]$$

Where

$$a_i = \begin{cases} 0 \; if \; Z(x_i) = Z(y) \\ 1 \; if \; Z(x_i) \neq Z(y) \end{cases}$$

Equation 3

In Equation 3, all the nodes in the data event are allocated the same weight. In Equation 4, the contribution is adjusted by the norm of the lag vector h_i from the central node, to the power of $-\delta$ to which the similarity distance is raised when computing the weight of each conditioning node. The value of δ is provided in the input to the DS algorithm.

$$\boldsymbol{d}\{d_n(x), d_n(y)\} = \frac{\sum_{i-1}^{n} a_i \|h_i\|^{-\delta}}{\sum_{i=1}^{n} \|h_i\|^{-\delta}} \quad \in [0,1]$$

Where

$$a_i = \begin{cases} 0 \; if \; Z(x_i) = Z(y) \\ 1 \; if \; Z(x_i) \neq Z(y) \end{cases}$$

Equation 4

For continuous variables, a weighted Euclidean distance is used in the DS algorithm (Mariethoz et al., 2010), which is expressed as the square root of the weighted mean square differences between the two data events:

$$\boldsymbol{d}\{d_n(x), d_n(y)\} = \sqrt{\sum_{i=1}^{n} a_i [Z(x_i) - Z(y_i)]^2} \qquad \in [0,1]$$

Where

$$a_i = \frac{\|h_i\|^{-\delta}}{d_{max}^2 \sum_{i=1}^{n} \|h_i\|^{-\delta}}, \qquad d_{max} = \max_{y \in TI} Z(y) - \min_{y \in TI} Z(y)$$

Equation 5

An example of a simulation based on the Euclidean distance is shown in Figure 3-2 (Mariethoz et al., 2010). The training image is borrowed from Zhang et al. (2006). A 100 conditioning data points have been used in the simulation (shown as circles in Figure 3-2b). Maximal number of 80 nodes and a tolerance value of 0.01 were specified as input controlling parameters.

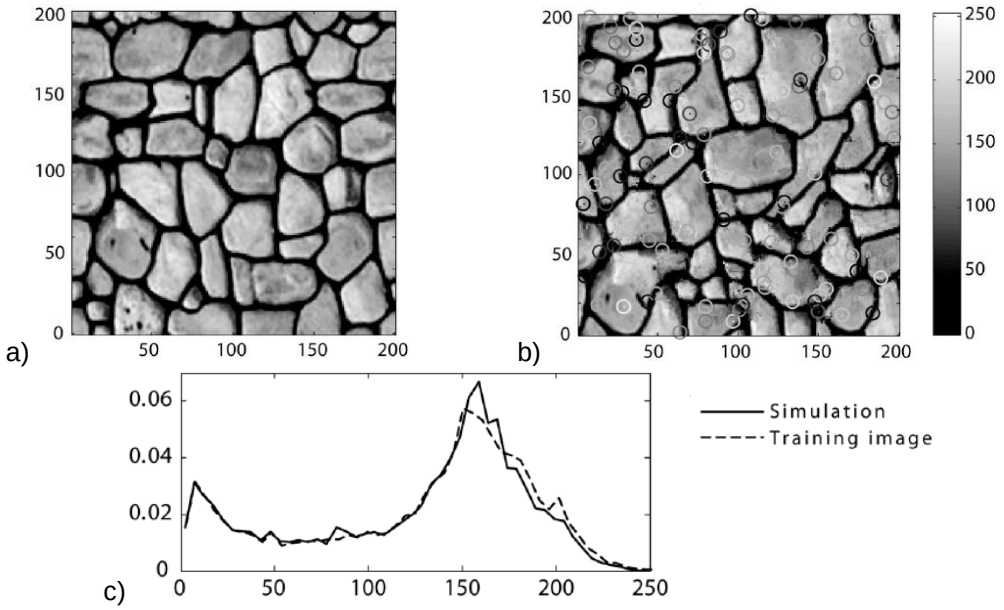


Figure 3-2. An example of continuous variable simulation using the DS algorithm: a) training image of a continuous variable, b) a realisation of the continuous variable using unilateral path. The 100 conditioning data are shown as circles with the hue reflecting the values of the variable according to the scale, c) histograms comparison of the training variable values in the TI and one realisation. Adapted from Mariethoz et al., 2010

The training image and realisation histograms of the continuous variable are compared in Figure 3-2c. The reproduction of the complex structures observed in the training image as well as the histogram reproduction is good. The simulation honours the hard data values well.

For multivariate simulation with m variables Z_k, where $k=1, .., m$, the ccdf of the variable $Z(x)$ is expressed as (Mariethoz et al., 2010):

$$F_k\left(z, x, d_{n\ 1,}^{1} \ldots, d_{n\ m}^{m}\right) = Prob\{Z_k(x) < z | d_{n\ 1}^{1}(x, L^1), \ldots, d_{n\ m}^{m}(x, L^m\}$$

Equation 6

Each variable Z_k can have its own data event $d_{n\ 1}^{k}\ (x, L^k) = \{Z_k(x + h_1^k), .., Z_k(x + h_n^k\ )\}$, making co-locality redundant. The distance between the data event x in the simulation grid and y in the training image represents the weighted average distances for each variable:

$$\boldsymbol{d}(d_n(x), d_n(y) = \sum_{k=1}^{m} w_k\, d\{d_{n\ k}^{k}(x, L^k), d_{n\ k}^{k}(y, L^k)\}$$

Where

$$\sum_{k=1}^{m} w_k = 1$$

$and\ w_k \geq 0$

Equation 7

This approach to distance calculation makes it possible to simulate several variables jointly or for multivariate conditioning. The variables do not have to be co-located. The multivariate simulation can benefit from soft conditioning by densely-informed secondary variables (Mariethoz et al., 2010). When multiple-point statistics and complex spatial patterns are considered for both, exhaustively known and sparsely informed variables, the net result is a better reproduction of the complex spatial patterns of a poorly informed variable.

A weight w_k can be defined per variable, to introduce user-control over the contribution of each variable in the final simulation. This approach can be used to produce a joint simulation, which reflects the quality of measurements of different variables and strength of the correlation.

When one or several of the joint variables are exhaustively known, the resultant simulation of the main variable is associated with less uncertainty, even in absence of primary variable conditioning data. Multiple-point dependencies and cross-correlations between all combinations of nodes within multivariate data events are respected. The variables do not have to average linearly. An appropriate measure of distance is defined for each variable, and a single random path can be used.

Other approaches to distance calculations are available in the algorithm, such as Manhattan distance (Arpat & Caers, 2007; Mariethoz et al., 2010; Straubhaar et al., 2016). The DS also allows for the utilisation of custom-defined similarity distances.

3.2.3 **Dealing with multigrid scales**

The multigrid approach, implemented in different MPS algorithms, is aimed at improving the reproduction of subsurface morphologies at different scales (Strebelle, 2002). It has been made redundant in the DS with the definition of the neighbourhood via the number of conditioning nodes n. This definition allows the radius of the data events to decrease dynamically as the simulation progresses and the density of populated grid nodes increases. The advantage of the DS implementation is redundancy of the migration of the conditioning data between the multigrid levels. With this method, conditioning to large datasets is highly efficient, while the reproduction of the different scale structures is improved (Mariethoz & Caers, 2015). The schematic of the inherent reduction of neighbourhood radius with the progression of simulation is shown in Figure 3-3.

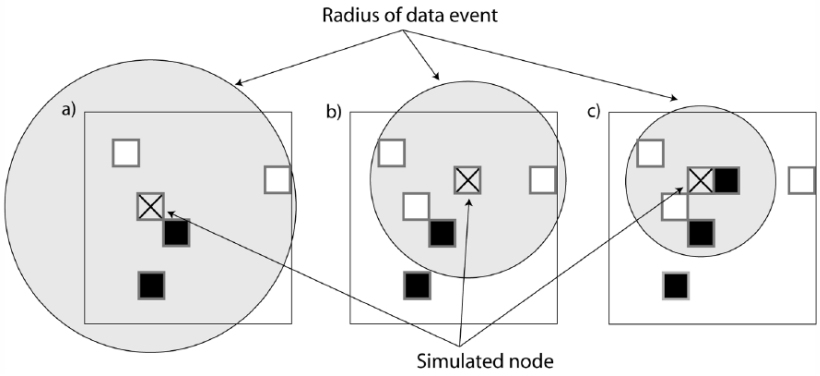


Figure 3-3. Schematic showing the process of reduction of the data events size during the DS simulation.
The maximum of four conditioning nodes has been specified in the input to the simulation. As the grid becomes more densely informed through stages in (a), (b), and (c), the data event radii reduce. Adapted from Mariethoz et al., 2010

It is important to note that the neighbourhood radius can also be fixed in the input to the simulation. An example of the applicability of this approach is when using an exhaustively known auxiliary variable with macro-scale patterns which describes stationarity of the main variable of the simulation. Providing certain size radius of the neighbourhood will ensure the stationarity zone depictured in the auxiliary variable is adequately captured when retrieving the data event for the primary variable while being stringent with memory usage during the pattern search. A further example will be given in Chapter 4 when using the potential field of lithological contact attitude for conditioning gold grade variable simulations.

3.2.4 **Non-stationarity**

The DS allows different approaches for dealing with non-stationarity, whether observed in a TI or a simulation grid (Chugunova & Hu, 2008; De Vries et al., 2009; Journel, 2002; Strebelle, 2002).

Explicitly defined zonation. The first approach, common to several MPS methods, is to divide a non-stationary TI into stationary zones, each considered as a separate stationary TI (Boucher, 2009; De Vries et al., 2009), and to apply zonation to appropriate parts of the simulation grid. No limitation to the number of zones exists in the DS with this approach.

Variation-based distance. Other methods to deal with non-stationarity in the DS are more straight-forward to implement and require less interpretive work from the modeller. One of the methods is to apply variation-based distance to non-stationarity cases (Mariethoz et al., 2010). This method is useful when simulating first-order non-stationarity, seen as a locally varying mean. It happens when the conceptual model provided as an input shows different ranges of values than the conditioning data. In this implementation, the data events are compared by their relative variations rather than actual values. An example of a similarity distance can be a pair-wise Euclidean distance:

$$\boldsymbol{d}\{d_n(x), d_n(y)\} = \sqrt{\left(\sum_{i=1}^{n} a_i[(Z(x_i) - \bar{Z}(x) - (Z(y_i - \bar{Z}(y))]^2\right)}, \in [0,1]$$

Equation 8

This approach to distance calculation allows for finding matches between the data events in the SG and the TI despite the difference in range and the non-stationarity. The resulting simulations (Mariethoz et al., 2010) have the same range and the non-stationary behaviour as the data while reconstructing the spatial structures found in the TI.

Figure 3-4 demonstrates the application example (Mariethoz et al., 2010). While the range of values in the TI is in the interval [-3.52; 3.99], the values in the realisation, [98.13; 111.72], reproduce those in the conditioning dataset: [99.55; 110.92].

This approach to describing non-stationarity can be useful in mining scenarios when the continuity and morphology of mineralisation or distribution of deleterious elements are exhaustively known in the mined-out portion of the deposit and need to be modelled in an environment with similar controls and multivariate relationships but a different tenor of mineralisation.

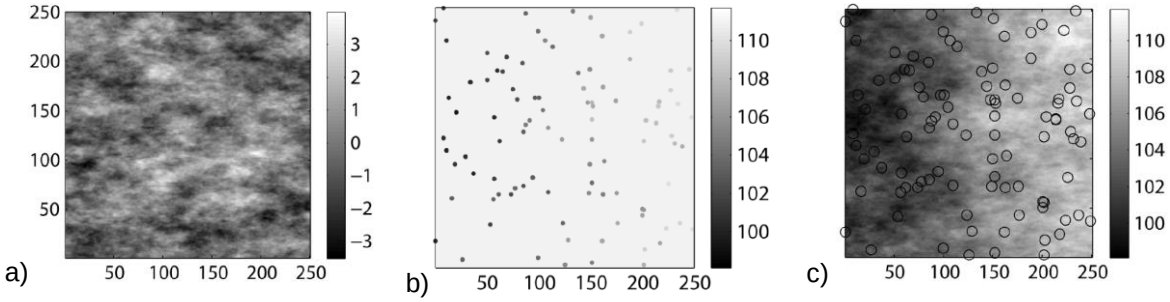


Figure 3-4. An example of the DS simulation with variation-based distance: a) multi-Gaussian stationary training image with a range of values in the interval [-3.52; 3.99]; b) non-stationary conditioning dataset with range of values in the interval [99.55; 110.92]; c) one realisation produced with variation-based distance, with a range of values in the interval [98.13; 111.72]. Circles represent the location of the conditioning data. Adapted from Mariethoz et al., 2010

Transformation of data events by rotation and homothety. A special case of non-stationarity modelling in the DS deals with rotation and changing the size of the patterns. An example is shown in Figure 3-5.

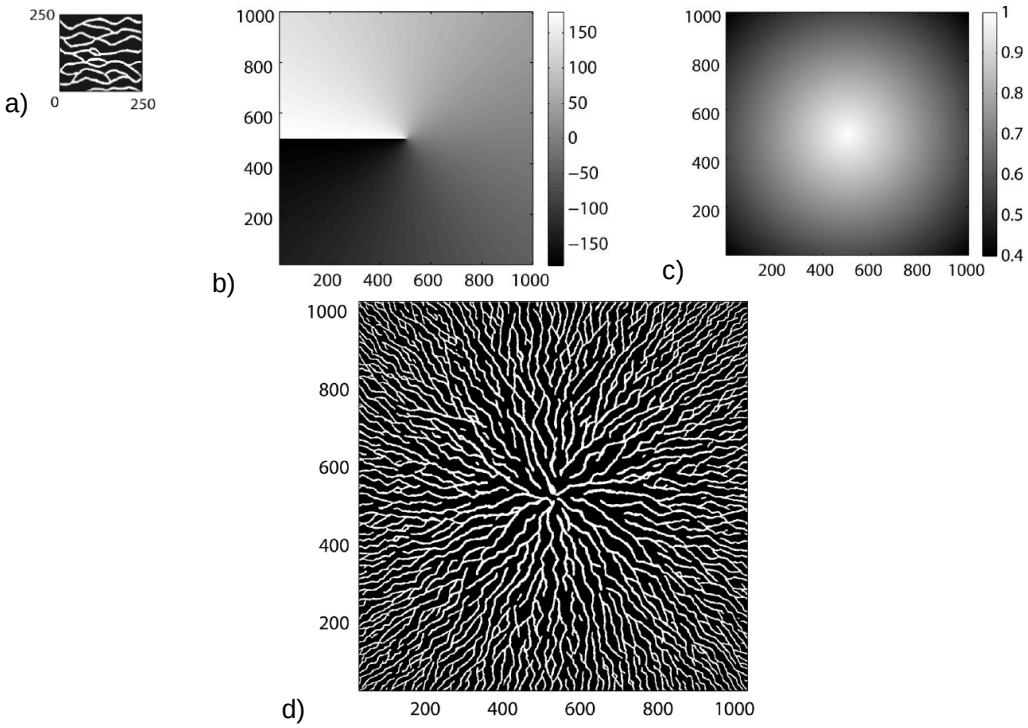


Figure 3-5. An example of the DS simulation with rotational and scaling transformations of the data events.
a) elementary training image; b) rotation azimuth map exhaustively informing the simulation grid; c) map representing the size of the patterns exhaustively informing the simulation grid; d) a realisation based on the transformed data events. Adapted from Mariethoz et al., 2010

In this case, a continuous field describing the non-stationarity over the simulation volume in relation to the stationary field observed in the TI needs to be introduced into the simulation (Mariethoz et al., 2010). In traditional MPS algorithms, this implies first constructing a data event catalogue using a transformed template (Strebelle, 2002). The simulation that follows applies a non-transformed template. In the DS, the TI is scanned directly with the transformed data event (Mariethoz et al., 2010).

Non-stationarity by auxiliary variables. Another example is encountered when either the training image or the simulation grid, or both, exhibit strong non-stationarity (Mariethoz et al., 2010). In practical situations, when the TI is constructed from dense observations of natural phenomena, it often displays strong non-stationarity. The DS allows using one or more secondary variables, which need to be available over the TI as well as over the simulation volume and to exhibit dependency with the main variable. The conditional probabilities for the main and secondary variables are directly inferred from the training image (Chugunova & Hu, 2008). It has been implemented in the DS as multivariate co-located cosimulation.

Non-stationarity by local probability constraints. To deal with categorical variable non-stationarity, a method utilising local probability constraints has been adopted in the DS (Straubhaar, 2019). The local probability constraints are input as grids of the same geometry as the simulation grid, where for each node of the grid a probability of a given category in a certain neighbourhood is provided. One grid needs to be supplied per category. The information in the local probability grid describes moving local probabilities over a fixed support size.

A variety of methods is implemented in the DS to deal with the main corner-stone of geostatistics – non-stationarity. It allows to account for complex types of non-stationarity and can prove invaluable in precious metal mining when abundance of information is gathered through the life of a deposit. Complex multivariate co-dependencies between mineralisation assemblages, alteration, geometallurgical variables can be challenging to

subject to parametric descriptions. Such cases yield themselves as good examples for MPS simulation with the DS.

3.2.5 Conditioning to block data

In the pre-production stage of geostatistical evaluation of mineral deposits, it is a common practice to assess the mineral content at large scale support, defined by wide drilling patterns. Blocks of a size approximating an average drillhole spacing provide a robust estimate by classical techniques such as ordinary kriging. Conventionally, data of different support sizes for additive variables are integrated by using parametric covariance-based models (Liu & Journel, 2009; Journel, 1999; Straubhaar, 2016). These methods are based on cokriging theory with Gaussian assumptions and point-to-point, point-to-block and block-to-block covariance models.

Traditionally in MPS, irrespective of the algorithm used, the simulation domain is defined at the same resolution as the training image and incorporating information at coarser support is challenging. Classical techniques to deal with such cases employ co-kriging framework with parametric definition via covariance models with Gaussian assumptions (Liu & Journel, 2009; Journel, 1999).

In the case of MPS application, Straubhaar et al (2016) proposed an extension to the DS, called DeeSse, to account for block support data. Simulation with block conditioning aims at reproducing the spatial structures present in the TI, and, at the same time, respect the block data. The input is provided as target mean values per block. Each block represents a group of nodes of any size or shape over the simulation grid. An overlap or gaps are allowed between the blocks. This block data is used in addition to the distance-based pattern acceptance.

Each block is described by the parameters $B, \mu_B,$ and t_B where:

- B - a set of nodes defining the block in the simulation grid,
- μ_B - a target value of the variable Z for the block mean,

- t_B - tolerance expressed as half-length of a target interval $[\mu_B - t_{B\,left}, \mu_B + t_{B\,right}]$ containing μ_B.

Each conditioning block data is considered satisfied if the mean of node values in the block is within the target interval.

The new algorithm proceeds as follows (Straubhaar et al., 2016): during the scan of the TI for the simulation of a node x, the blocks containing x are identified and the current mean over each of these blocks is computed. The values at the informed nodes within each block and of the candidate value at x are taken account of. For each block containing x (in case of overlapping blocks), the current block mean is compared to the target mean value specified in the input, and an error is calculated. The error introduced in Equation 2 is updated by adding each of these errors. After the step 3.2.2.2 the following steps are added:

Each block containing the node x in SG is subjected to the following (Straubhaar et al., 2016):

i) The current mean of the block B with the candidate node y in the TI is calculated:

$$\mu_B^*(y) = \frac{1}{k}\left(\sum_{i=1}^{k-1} Z(x_B^{(i)}) + Z(y)\right)$$

Equation 9

where $x_B^{(1)}, .., x_B^{(k-1)}$ are informed nodes in the block B in the neighbourhood of k nodes.

ii) An error on the constraint for the block is calculated based on the target mean value μ_B and the tolerances:

$$E_B = E_B(\mu_B, \mu_B^*(y), t_{B\ left}, t_{B\ right})$$

Equation 10

iii) The total error is updated with the error on the block constraint:

$$E = E + E_B$$

Equation 11

If the distance between the data events $d(d_n(x), d_n(y)) \leq t$ and $E_B = 0$ for each block containing x, the scan is stopped, and the value is pasted for the candidate node. The computational time is very fast as the means on every block are stored and updated after the simulation of each successive node.

This implementation of block conditioning in the DeeSse algorithm can accommodate variables which add linearly, such as mineral grade distribution, when the value of a variable at the block support is expressed as arithmetic mean of the simulated variable of the nodes at point support. The block values can be defined in a complex manner, for example, within irregular shapes delineating domains with similar properties. With this approach, the method can be extended to and be specifically useful for dealing with non-additive variables, which is often the case in geometallurgy.

The block conditioning routine can be used for either categorical or continuous variables. It can also be adopted for multivariate problems when each variable is constrained by its own block conditioning data (Straubhaar et al., 2016). An example of it would be using geophysical data with geological controls and grade to simulate a multivariate mineralisation setting. This new development has the advantage of generating complex spatial features while accounting for data collected at different scales, which is a common practical problem in geosciences.

3.2.6 **Conditioning to connectivity data**

Connectivity reproduction has been topical to groundwater, oil reservoir and multi-phase flow applications. It has proven to be challenging with Gaussian-based models, which result in renditions of a variable where intermediate values of a continuous variable are ensured to be connected

rather than extreme values (Journel & Alabert, 1990; Journel & Deutsch, 1993; Renard, Straubhaar, Caers & Mariethoz, 2011). Truthful reproduction of connectivity is just as important when modelling structurally controlled mineralisation. It can be practically achievable and meaningful when unique structural features can be identified in the field or from diamond drillholes. Selective mining calls for a good definition of high-grade values.

As mentioned before in the motivation to this research, traditional statistical measures for describing the spatial distribution of a field struggle to define connectivity. They are usually based on two-point statistics and are only sufficient to understand the probability of occurrence of a certain value or a range of values at a location knowing the value at another location (Renard & Allard, 2013). No structural considerations beyond two-point statistics are considered.

The method of conditioning to connectivity data implemented in the DS requires spatial identification of connected points in the input to the simulation. The geometry of connectivity must be provided explicitly: nodes connected by face, by edge or by vertex. Categorical or continuous variables can be used for continuity constrains. In the latter case, the connectivity is provided for a range of values falling within a connected class.

The base method employing connectivity between two points proceeds by borrowing connected paths from a training image prior to commencement of the main variables' simulation (Renard et al., 2011). After having pasted the connectivity pattern of the two points, the conditioning data of the main variable is assigned to the simulation grid. The values of the connected groups of pixels belonging to the same class, termed 'geobodies', will be modified locally to honour conditioning data, before proceeding with random path simulation. The simulation will ensure that both constraints, the connectivity and the local conditioning, are honoured.

In the case of n-point connectivity, where $x_1, \ldots, x_n$ are n points that must be connected, first a path connecting x_1 and x_2 is built. For $i = 3, \ldots, n$, the

algorithm proceeds by constructing a path connecting x_i to a point from the previous path until all n points provided in the input are connected. The method is computationally efficient while ensuring consistency between the geometry of the connectivity paths and the random function model (Renard et al., 2011).

3.2.7 **Reconstructive mode**

Another advantage of the DS over other MPS methods is its applicability in reconstructive mode (Mariethoz, 2009) when an explicitly defined training image is not required but exhaustively informed portions of the simulation volume can serve as a source of patterns. Configuration of any informed geometry can be used, applied to continuous, categorical or multivariate cases. The variables do not have to be exhaustively known over the same set of nodes. Stationarity does not need to be ensured over the simulation volume. In non-stationary cases, an auxiliary variable can serve as a guide for extracting patterns from a statistically similar portion of the simulation volume.

This functionality makes it appealing to mining applications when with the progression of the mining interface, exhaustively informed production portions of the orebody can serve as a source of multivariate relationships to simulate the unknown. This is only valid under the assumption that the mined-out proportion is representative of the unknown portion, which may not always be the case. It is, however, presumed that the geological knowledge gained from the mined-out part will allow to identify such a case timeously, and a geologist's expertise is still a prerequisite.

Figure 3-6 represents an example of the DS application for reconstructing a borehole microresistivity image. A device used to obtain the image in Figure 3-6a is composed of four pads carrying two sensors each. The latter serves as the source of current which, when modulated in amplitude, provides information about structural properties of the rock. The information is not obtainable outside of the pads' trajectory.

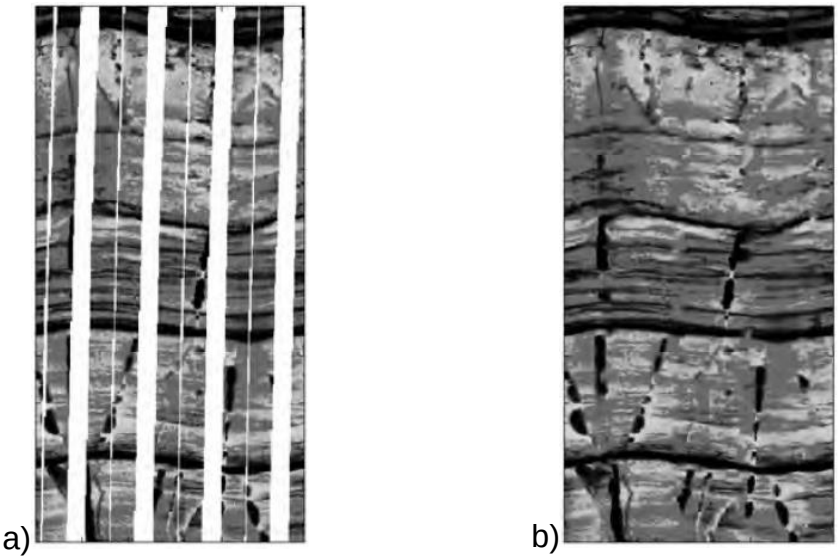


Figure 3-6. An example of the DS application in reconstructive mode for restoring a complete image of downhole microresistivity measurements. a) the original image taken by 8 sensors of a downhole microresistivity device; b) one reconstructed realisation. Adapted from Mariethoz, 2009

The image in Figure 3-6b represents one realisation by the DS which allowed to reconstruct missing zones. The complete image shows meaningful reproduction of fine-scale features, with continuity maintained, and non-stationarity of fractures and lamination preserved.

3.2.8 **Summary of the DS advantages**

In summary, the main advantage of the DS is related to the direct nature of sampling for pattern extraction and in bypassing the traditional steps of first counting and storing the ccdf obtained from a TI. As such, the size of conditioning data events does not pose a problem for the DS and high CPU requirements associated with complex natural phenomena are avoided. The more efficient usage of memory allows for the simulation of multivariate and continuous variables.

The trade-off between CPU and the resulting quality of the simulation is controlled by several parameters: the size of the neighbourhoods, the value of the acceptance threshold and the fraction of the TI to be scanned for the simulation of each node.

By not storing the templates of the TI, the requirement of fixed geometry of the data event is not needed, as the templates are retrieved dynamically and can change at each simulation node. The need for an explicit definition of multigrid approach is made redundant.

The DS has computational advantages. It can be used in parallel core processing. In this case, each CPU performs scanning of a limited portion of the TI. The parallelisation strategies can employ Open Multi-Processing (OpenMP) libraries, or parallelisation strategies using Graphics Processing Units (GPU) (Mariethoz et al., 2010).

Other advantages of DeeSse are:

- Conditioning to connectivity data;
- Dealing with incomplete TIs and training data sets;
- Using multiple TIs;
- Multiple scale simulation based on Gaussian pyramids.

The modified form of the algorithm introduced by Rezaee, Mariethoz, Koneshloo and Asghari (2013) allowed for patch-pasting of groups of nodes rather than a single-pixel pasting approach, ensuring processing efficiency of patch-based methods.

The amount of data available for mapping physical parameters in geosciences is becoming enormous. In mining projects, it includes dense drillhole information, data from geophysical surveys, surface mapping, hyperspectral imaging acquired via satellites or with drones, to name just a few. With rich information, significant nonparametric spatial statistics can be derived from the data directly (Mariethoz et al., 2010). All the above make the DS a powerful tool for real geological scenarios, to exploit complex nonparametric relationships between variables.

3.2.9 **Input into the Direct Sampling algorithm**

Input to the DS algorithm includes several parameters. Some of them have already been described in Section 2.3.1 Simulation Inputs and are briefly mentioned here as applicable to the DS (Straubhaar, 2019).

Simulation grid. In the DS, the simulation grid can be empty or have variables attached to all nodes or a subset of nodes. A partially informed SG can be useful when performing hierarchical simulation, or when simulating a portion of a deposit adjacent to an exhaustively known part of the orebody.

Training image. Each node of the training image has at least one or a few variables of interest. The undefined cells will carry a default value of -9999999. If required, several training images can be provided as an input to the algorithm. The simulation grid itself can serve as a training image. This can be useful when conducting simulation in reconstructive mode.

Conditioning data. Can be in a form of a regularly gridded file with the same geometry as the simulation grid, or a point set file representing sample centroids with X, Y, Z coordinates and associated variables. Multiple conditioning data files can be provided in the input.

Mask image. A mask is another type of a grid object useful when only a portion of the volume is to be simulated. It has the same geometry as the simulation grid with nodes to be simulated flagged as 1, and those to be omitted from the simulation as 0. Mask files can be supplied per variable if necessary.

Search neighbourhood parameters. Search neighbourhood represents a 3D ellipsoid describing anisotropy. It is defined in terms of:

- Angles of anisotropy direction (azimuth, dip and plunge), according to GSLib convention (Deutsch & Journel, 1998);
- Search radii in each direction expressed in number of grid nodes;
- Anisotropy ratios in each direction expressed as a number of nodes representing a geological distance of 1. The ratios are used for

computing the distance from each node in the search neighbourhood to the central node;

- Power of computing weights according to distance of each node found in the search neighbourhood to the central node of the data event.

The data event nodes inside the search neighbourhood are sorted according to their distance to the central node, from the closest to the furthest one. The anisotropy ratios are used when computing the anisotropic distance of each node to the central node in the search neighbourhood. Usage of anisotropy is optional when performing the DS simulation. For a default setting with isotropic neighbourhood definition, an unlimited search radius is used in each direction which equivalent to the multigrid framework.

Rotation. This parameter is useful when the spatial arrangement of the elements in the simulation grid is rotated in relation to the elements present in the training image (Mariethoz, 2009).

The rotation input can be in the form of constant parameters for each angle (azimuth, dip, plunge) for the whole simulation grid, or alternatively, as an input grid image file representing locally varying field of azimuth, dip and plunge. Tolerance threshold angles can be supplied in each case.

Simulation path. A variety of simulation path types can be used: unilateral, random or random with priority allocated by weight inversely proportional to the distance between the conditioning nodes and the node to be simulated.

Consistency of conditioning data. This parameter, used in simulation of continuous variables, represents the maximum accepted expansion in both extremities of the range of values in the training image for covering the conditioning data values when calculating similarity distance (Mariethoz, 2009). For example, if set to 0.05, the conditioning data values are allowed to be 5% beyond the range of the values in the training image in both extremities.

Maximum number of neighbouring nodes for each variable. The parameter describes the conditioning dataset (Mariethoz, 2009) with the maximum number of nodes n in the simulation grid closest to the node x being simulated, which are to be used in the search neighbourhood. This value is defined per variable.

Similarity distance threshold. One of the most significant parameters of the algorithm (Mariethoz, 2009), the distance threshold t defines the allowed variation in the similarity distance being searched for in the training image. It needs to be specified because a TI pattern matching a data event d_n exactly is often not found, specifically for continuous variables. When the distance between the TI pattern and d_n is smaller than t, the central node of the TI pattern is pasted at location x. The distance is measured as the fraction of non-matching nodes for categorical simulations. The distances are normalised. It ensures the minimum to be zero with exact match and the maximum to be 1 without a match. This parameter needs to be provided per variable to be simulated.

Probability constraints. Probability constraints can add extra conditioning to the categorical variable simulations (Mariethoz, 2009). Global or local probability constraints can be used.

Connectivity constraints. This feature allows for ensuring connectivity by either setting connecting paths prior to simulation or ensuring connectivity during pattern simulation. The sequencing of the connectivity can be defined either in random order or in order of increasing distance of separation. The types of connectivity available are by cell face, by face and edge, or by face, edge and corners. In this context, a face represents a side of a cell.

Maximum scan fraction for a training image. This parameter represents the volumetric fraction of the TI to be scanned when searching the TI for suitable matches to the data events (Mariethoz, 2009). It limits the number of TI patterns that are scanned for similarity match. The parameter ranges from 0 (no scan) to 1 (scan full TI). If the maximum fraction of the TI is scanned

and still no TI pattern with acceptable distance is found, the central node of the TI pattern with the lowest distance is pasted at the location x.

3.3 The Mineable Stope Optimiser to delineate the Mineral Resource

The Mineable Shape Optimiser (MSO), available as part of Datamine© software, enables the generation of underground mineable stopes of optimal size, shape and locality (Alford et al., 2007; Datamine Product Help, 2019). It uses as an input a grid or a block model containing the properties of the mineralised material such as density, grade of the mineral of interest and, optionally, of deleterious elements. It allows for maximising the recovered value while accounting for mining dilution and taking cognisance of the orebody attitude, width of mineralisation and engineering constraints. The algorithm is used for reporting a Mineral Resource to be mined by underground mining methods. It is frequently employed for delineation of mineable shapes in an open pit scenario.

The necessity of using the MSO algorithm stemmed from the research objective - to test the performance of a stochastic geological model created by the DS algorithm for the Mineral Resource reporting purposes. The latter necessitates a sufficient confidence in the Mineral Resource characteristics to allow for the application of modifying factors in support of a chosen mine planning method.

MSO proceeds by slicing sections through the block model at specified regular or variable intervals and generating strings around the mineralised material satisfying a sequence of constraints. The strings are subsequently linked to produce a wireframed shape. The latter is evaluated against a block model to give a stope that maximises a recovered resource value above a given grade cut-off value. Engineering parameters such as anticipated wall dilution, minimum separation distance between sub-parallel stopes, minimum and maximum values for mining width, wall angles, and stope height and width are provided in the input.

When using a block model as input to the algorithm, a convention similar to the one with a simulation grid is used. Unlike an SG grid, a block model's elementary unit, a panel or block, is defined at a larger scale. The block size approximating average spacing of data normally produces robust interpolation estimates. Finer resolution is achievable by subcelling the parent block to allow capturing of domain boundaries. In the current work, the input to MSO is provided in the form of a simulation grid at point support.

MSO terminology is explained using an example in Figure 3-7 (Datamine Product Help, 2019). A narrow orebody consisting of subvertical tabular splays striking along the block model X-axis is shown. Stoping design to extract the ore would consist of individual panels. The spacing between the panels along the strike (in this case, along the X-axis) is termed section spacing. The vertical size of the panels is termed level spacing.

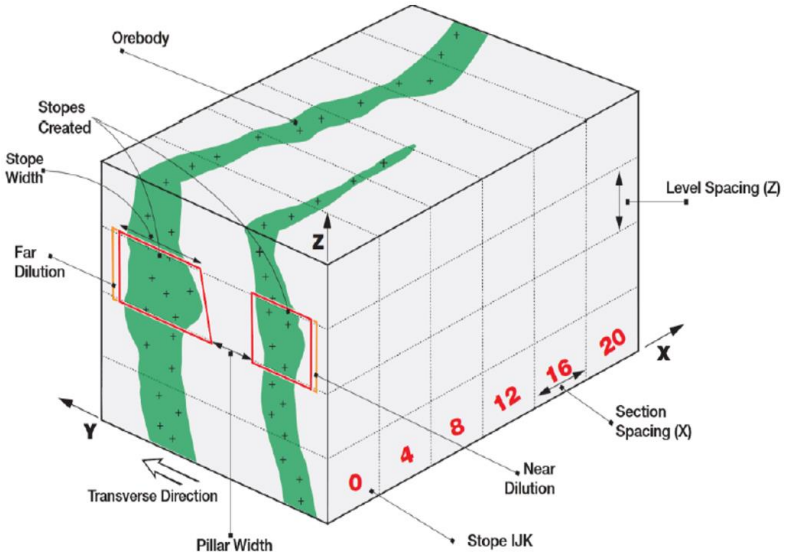


Figure 3-7. Schematic explaining terminology and parameters of Mineable Stope Optimise parameters. Adapted from Datamine Product Help, 2019

The method is based on annealing approach with the objective to optimise the stope design to meet stipulated geometric, geotechnical and economic constraints (Datamine Product Help, 2019). The process of optimisation takes place in the transverse direction along the axis which is perpendicular to the strike of the orebody. The size of individual stope shapes

perpendicular to the strike is termed stope width. Stope shapes are trapezoidal prisms, with parallel sides along the YZ plane in the example above. The base of each prism is a trapezoid using four points connected by straight lines between levels. More control over the shape is allowed by incorporating a few more points.

The optimisation algorithm ‘floats’ the stoping shape along the orebody transgressing the individual section boundaries at a specified section spacing, hence a fine resolution of section spacing is preferred with further post-processing to aggregate stopes and discard isolated ones.

Extensive functionality is present in MSO, such as the incorporation of dilution or making stopes honour brittle structures. Describing the full MSO functionality is beyond the scope of the current work. Only basic parameters for stope definition sufficient for the accuracy of reporting have been used in this research.

3.4 Chapter summary

The emphasis of the chapter was on introducing the reader to workings of the algorithms used in this research, namely the Direct Sampling MPS method and the Mineable Shape Optimiser. The description focused on the basics and advantages of the DS over other MPS algorithms. Different parameters have been explained in detail.

Together, the two methodologies are providing tools for performing stochastic simulation and generating mineable shapes for an underground mining scenario to allow quantification of geological uncertainty in the estimated Mineral Resource. The next chapter will focus on the practical application of the methodology to a complex gold deposit.

4 MODELLING A STRUCTURALLY COMPLEX ENVIRONMENT WITH THE DS MULTIPLE-POINT STATISTICS ALGORITHM

4.1 Introduction

An emphasis in this chapter is on describing a practical workflow to derive a stochastic model of an orebody with complex structural architecture using the DS.

The discussion in the current chapter details the methodological facets of the practical DS implementation: study strategy, geological settings, data description, exploratory data analysis, the process adopted for generation of a training image, the simulation workflow and the global uncertainty assessment. The progression of generating the different variables in the training image includes the primary variables – those having a direct impact on the calculation of the tonnage, grade and contained metal, and the secondary variables assisting in the modelling process of the primary variables. Input parameters which produced satisfactory realisations are presented for each simulated variable. For the methodology overview, the reader is referred to Chapter 1, Section 1.4 Methodology Overview.

4.2 Study strategy

The specific orebody chosen for the research application is characterised by compound controls on mineralisation and non-stationarity in the distribution of mineralisation. It is believed to be a good test for the DS method performance in the environment where stationarity can be described by a secondary variable rather than explicitly modelled.

In the execution of this research, a significant amount of time and effort have been invested in the understanding of the controls on mineralisation and incorporating them into a representative model. This model, based on dense data, will be referred to in this thesis as the ‘true’ model. The term is adopted in place of the ‘training image’ due to the multi-purpose of the model: only a

part of it is used as a training image for simulation, the rest, alike to a machine learning framework, is utilised for benchmarking the simulation performance.

Multiple realisations of categorical variables representing geostatistical domains and a continuous variable for gold mineralisation are generated. They are analysed for deviation from the 'true' model over the mined-out area of the chosen gold deposit. A spread of uncertainty of the stochastic model from the 'true' model is expected to be within acceptable limits. A criterion adopted to classify the "Indicated" Mineral Resource category in the current research is for the estimated tonnes, grade and metal content to be within 15% error with 90% confidence for a volume equivalent to annual mining production, and for the "Inferred" Mineral Resource the error can exceed this limit. The test to pass the criterion for classification of the Mineral Resource is performed for a drill spacing pattern of 40 m x 40 m. Appropriate mining method assumptions are taken cognisance of.

A dataset for the study originated from the Geita Gold Mine, a subsidiary of AngloGold Ashanti in Tanzania. One of the many orebodies, namely a mined-out portion of Geita Hill West has been chosen which at the time of conducting the research, was being mined as an open pit operation. Historically, some portions of the deposit have been mined by underground excavations. Extensive areas of the deposit have been drilled at dense spacing by production drilling, with lithological description of the chips and gold assays available providing a good geological dataset for a training image.

The orebody has a distinct definition of geological and geostatistical domains with pronounced folding architecture. The intercalation of two main lithologies does not present an easy case for the application of any conventional geological modelling tools, such as CAD-aided methods, implicit variogram-based functions or variogram-based simulation algorithms. An MPS approach appears to be a feasible choice. Modelling the two lithologies is deemed to be of high importance due to the difference

in densities and pronounced control on mineralisation their interaction exhibits. Besides the lithologies, modelling of gold mineralisation controlled by multivariate geological dependencies provides a good example for MPS simulation of continuous variables.

The orebody is known exhaustively. A fine resolution model with node spacing 2.5 m x 2.5 m x 2.5 m was created using the abundantly available production and exploration drilling. All available raw drillhole data, consisting of lithological logging and gold grade assays in the exploration and production drillholes, was used for creating the ‘true’ model. Thereafter this raw data was discarded from the project.

Further on, an approach akin to supervised machine learning is adopted. The ML consists of three phases: 1) a process of training using all available data, 2) testing of the predicted outcome, 3) application of successful parameters (Dey, 2016; Samson, 2019). In the proposed workflow, the ‘true’ model is subdivided into two portions emulating a mining scenario with the transition from open pit to underground mining, or from the ‘known’ into the ‘unknown’. The top part is presumed to be mined-out and exhaustively known. This portion is used as a training image for simulating over the bottom sparsely conditioned part. This step is equivalent to the phase one in ML. The results are compared to the ‘true’ model available over the bottom part (phase two). Once they are deemed representative, they can be extended into the unknown (phase 3).

During the MPS simulation, only sparsely spaced conditioning data is used. The hard data is presented as a pseudo drillhole set. It is created by sampling the bottom half of the ‘true’ model along synthetic drillhole traces on a grid equivalent to the lower confidence grid of the Mineral Resource delineation spacing, namely at 40 m x 40 m in the plane of mineralisation continuity. The sparse spacing was adopted to yield higher confidence in the conclusions drawn on the performance of the algorithm.

All further references to conditioning hard data implies the pseudo drillhole set. This approach is used to avoid introducing bias and irregularities of real

exploration grids which are never uniformly spaced, with drillholes often fanning off from available drill pads. As such, no declustering for statistical inference was deemed necessary in this research.

The exhaustively defined model of the deposit has been termed the ‘true’ model, the top portion of it – the training image. The simulated realisations are validated vs the ‘true’ model available over the bottom part, the training image and the input conditioning data. The data inputs for simulation are an elementary training image for lithology simulation, a ‘true’ model extract as a TI for the mineralised zone and Au simulation, and a conditioning data extract on a grid of 40 mE x 40 mN x 2.5 mRL node spacing.

A general workflow diagram is presented in Figure 4-1.

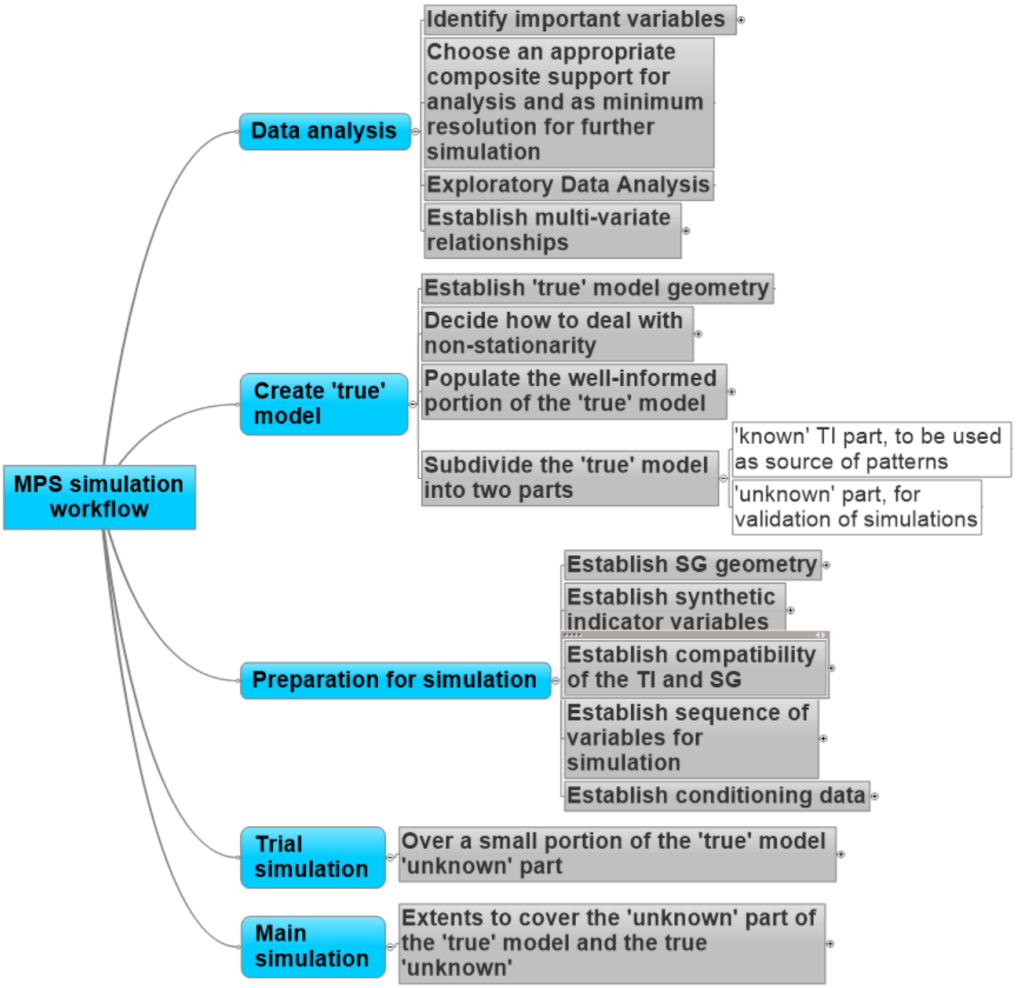


Figure 4-1. General workflow diagram for performing MPS simulation

As mentioned in Section 1.4.5, the data analysis was performed using 3rd party proprietary software packages: Datamine©, Geovariances Isatis©, Snowden Supervisor© and Seequent Leapfrog Geo©. Data formatting and manipulation in preparation for simulation and post-processing was coded in Python scripts utilising a Jupiter Notebook environment. Processing was done using an i7 8-core solid-state laptop with 16GB RAM. Further references to processing time are made in relation to these specifications. The detailed modelling process is described in the pages below.

4.3 Geological Setting

The Geita Hill deposit forms one of the deposits of the Geita Gold Mine. It is situated in the northwest of Tanzania, to the south of Lake Victoria. It is located within a fragment of Sukumaland Greenstone Belt, termed Geita Greenstone Belt. The deposit is hosted by deformed supracrustal rocks intruded by diorite dykes and sills originating from large diorite bodies at depth (Sanislav et al., 2017).

The supracrustal package is comprised of sandstone, siltstone, shale and felsic volcanoclastics metamorphosed to greenschist facies (Sanislav et al., 2017). The sequence is generally referred to as banded-iron formation (BIF) due to its well-bedded nature, widespread silicification and the presence of extensive magnetite alteration (Kabete, Groves, Mcnaughton & Mruma, 2012; Sanislav et al., 2017). The sedimentary pile is affected by magmatic intrusions of different age and composition. Dioritic intrusions termed further diorites are, volumetrically, the most significant. They occur as dykes and sills up to 10m thick and are mostly sub-parallel to bedding.

Mineralisation at Geita Hill is a function of deformational history and is determined by two groups of structures. The earlier group resulted from folding and shearing events of ductile rocks. The later group of structures, responsible for the main phase of mineralisation, formed during brittle-ductile faulting and shearing (Sanislav et al., 2017). The shears are preferentially developed in sedimentary units, deflecting around intrusive

diorite margins. A few geologists spoke about strong lithological controls on mineralisation at the larger Geita mine deposits (Paranhos, 2008; Cowan, 2009), which hold true for Geita Hill orebody. The competency contrast between the diorites and BIFs and the folding geometries play an important role in the morphology of the ore shoots.

Gold mineralisation is predominantly shear-controlled and is localised along a distinct moderately dipping brittle-ductile deformation zone, termed Geita Hill Shear Zone (GHSZ) which extends over 180 m. It is shown in semi-transparent yellow on the geological map and cross-section in Figures 4-2a and 4-2b. Often in the literature, the zone has been described as “braided” shear zone (Brayshaw, Nugus & Robins, 2010). The anastomosing structures are narrow, less than 10 cm, mostly parallel to bedding, and have no significant displacements (Sanislav et al., 2017).

The GHSZ is sub-parallel to the axial trace of a large S-fold in the pit. In the central portion of Geita Hill, the individual shears are subparallel to bedding. The shearing and, consequentially, mineralisation are preferentially developed in the central part of the S-fold hinge zone, related to high strain fracturing, and along diorite-ironstone contacts. The intersection of the GHSZ with different structural elements and diorite-ironstone contacts provided ideal mineralisation conduits and the best depositional sites for gold (Sanislav et al., 2017). The high grades occur in shallowly dipping sigmoidal lenses. They are localised along the contact between the BIF and diorite, the highest grade in BIF. Some high-grade mineralisation occurs within faults and has a steeper dip.

The transition from high gold values and intense sulphide alteration to barren unaltered ironstone takes place over a few meters (Sanislav et al., 2017). The gold mineralisation is genetically related to sulfidation reactions of sulphur-rich auriferous fluid in chemically reactive iron-rich host rocks (Borg & Rittenauer, 2000).

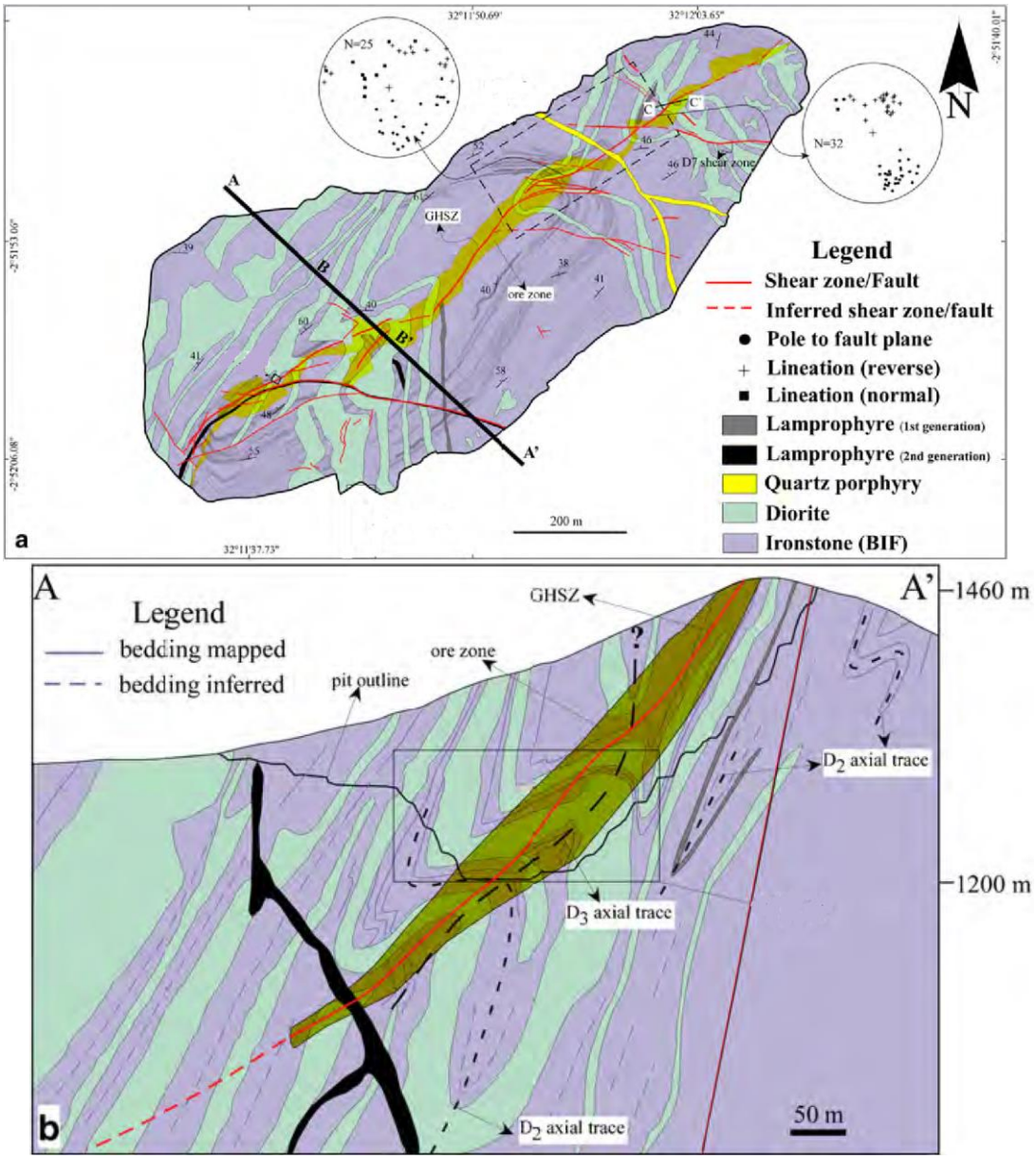


Figure 4-2. Geita Hill geological map (a) and geological cross-section (b) along line A–A′. The deposit geology is dominated by ironstones (purple) intruded at various times mainly by diorite sills and dykes (green), late lamprophyres (black) and minor quartz-porphyries (yellow). GHSZ ore zone is shaded in semi-transparent yellow. It is sub-parallel to the axial trace of the large S-fold. The map is presented in geographic coordinate system. Adapted from Sanislav et al., 2017

Of relevance to further description of the modelling process are the controls on geostatistical stationarity of the deposit. At Geita Hill, it is defined by the prevailing lithologies, their folding attitude, and the occurrence of the GHSZ. As mentioned above, the latter is described by the folding architecture and is located in the hinge damage zone of the folded lithologies (Paranhos,

2008; Cowan, 2009; Brayshaw et al., 2010; Sanislav et al., 2017). As such, the main mineralisation controls influencing the morphology of ore shoots, and considered during the modelling process are:

- Structural elements (bedding attitude, fold hinges) pre-dating emplacement of mineralised fluid flow;
- Brittle-ductile shear zones – conduits of gold-bearing fluids;
- Proximity to contacts between the two main host rock lithologies: BIF and folded diorite intrusions. Gold deposition is related to competency contrast in BIF and diorite, and sulphidation reaction in iron-rich BIF.

Modelling of late lamprophyre and quartz-porphyritic intrusive bodies is considered immaterial for the current research.

4.4 Generating a representative ‘true’ model

The simulation was undertaken over the mined-out portion of Geita Hill West orebody.

4.4.1 Drillhole summary

The drillhole dataset comprised 19,978 drillholes: 763 exploration drillholes and 19,215 production drillholes. Of the total, 126 holes were by diamond core (DC) and 19,852 by reverse circulation (RC) drilling. The majority of the drillholes were sampled at 1m intervals.

The exploration drilling, available at 40 m (along strike) by 20 m (across strike) to 40 mE x 40 mN grid, together with the production drilling on a grid of 10 mE x 5 mN was used to construct a well-informed ‘true’ model. The drillholes with lithological logging of the two main units are shown in Figure 4-3.

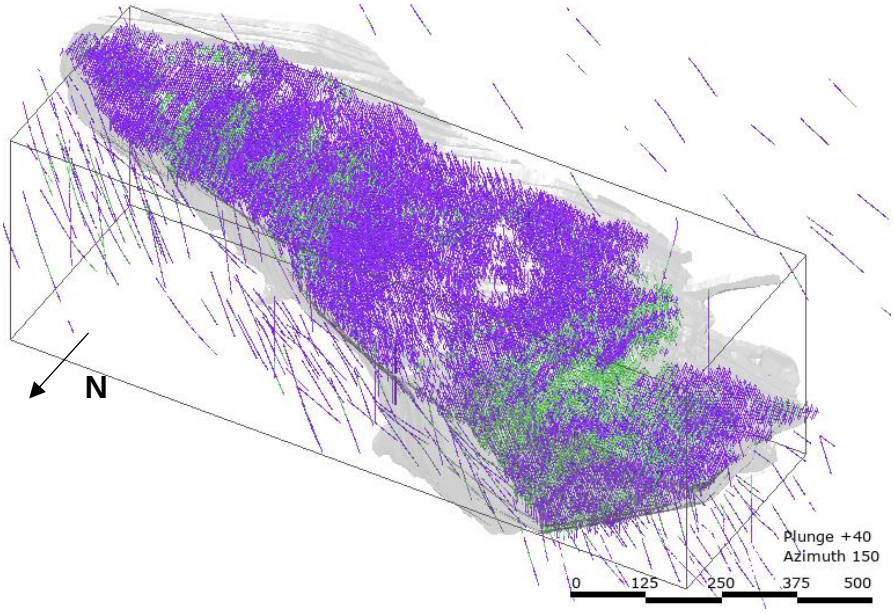


Figure 4-3. Isoclinal view of the study area with exploration and production drillholes coloured by lithology (purple – BIF, green – diorite)

4.4.2 **Drillhole gold grade statistics and analysis**

Exploratory gold grade data analysis

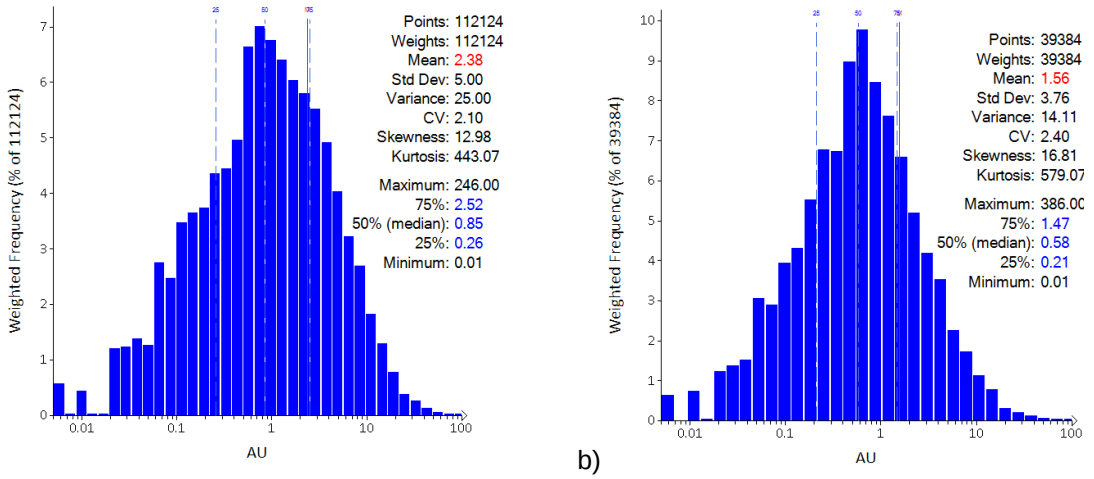
Multiple-point statistical simulation as any other classical geostatistical approach relies on a reasonable definition of stationarity. Ideally, clear geological characteristics should be used in arriving at the domain boundaries. When multiple and complex geological controls exist, there is often no distinct geological boundaries present which could be used for explicit geologically supported delineation of the orebody.

For this purpose, it is a common practice to define a cut-off grade for the variable of interest at which the mineralisation morphology “holds together” for viable economic extraction. Justification for the selection of the cut-off to define the boundary between waste and ore is an important decision. In the current case, such a threshold for mineralisation was chosen at a cut-off gold grade of 0.5 g/t suitable for the open pit mining method. The gold grades greater than or equal to the threshold were considered to belong to the mineralised zone. Further in the text, when reference is made to the

‘shear zone’ or ‘mineralised zone’, the material with the grade equal to or above the specified threshold is implied.

Since lithology imposes an important control on the tenor of mineralisation, the exploratory data analysis was conducted by considering the lithological domaining and the mineralisation threshold. As the sampling was done to 1 m intervals it was considered appropriate to use the same composite length for exploratory data analysis (EDA).

The histograms of logarithmically transformed gold values together with the declustered gold statistics within the shear zone per the main lithologies are shown in Figure 4-4. Gold mineralisation within the two lithologies in the mineralised shear zone displays strongly positively skewed lognormal distributions. Higher grades are associated with BIF, a mean of 2.38 g/t in BIF vs 1.56 g/t in diorite. The mean value in both lithologies is much higher than the corresponding median value. The interquartile range is higher in BIF, showing broader spread of values. The coefficient of variation is quite high in each lithology.



a) b)

Figure 4-4. Log histograms and statistics of declustered 1 m gold composites per lithology within the mineralised shear zone:
a) in BIF, and b) in diorite

The shape of the BIF distribution at the lower end of values displays some deviation from an ideal lognormal shape indicating that a mixture of distributions is present, possibly attributable to presence of geostatistically

different lithologies combined under the BIF group. While of significance to the understanding of the deposit, a further investigation into this topic is considered to be outside the scope of this research.

The gold distributions outside the mineralised zone are also highly positively skewed with gold statistics in the two lithologies being quite similar. There are marginally higher grades and broader spread in BIF (Figure 4-5).

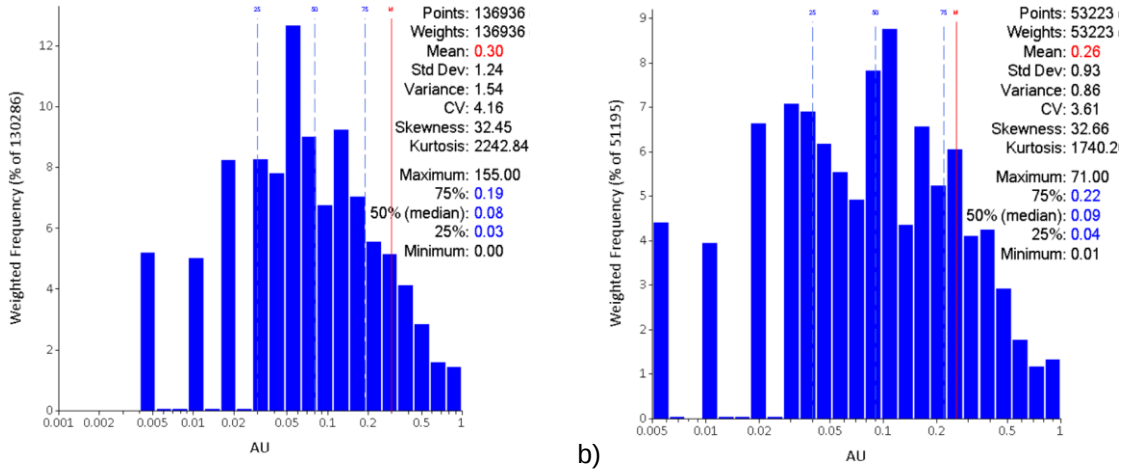


Figure 4-5. Log histograms and statistics of declustered 1 m gold grade composites per lithology outside the mineralised shear zone:
a) in BIF, and b) in diorite

Boundary analysis

Boundary analysis is considered important in the current research for understanding the character of mineralisation as a function of distance from the contact between the two main lithologies. As noted by different authors, BIF has higher tenor of mineralisation due to sulphidation reaction in iron-rich rock. At the same time, competency contrast between the two main rock types prompted preferential shearing of sedimentary units with deflection around diorite intrusives (Sanislav et al., 2017).

Boundary analysis is applied to the gold simulation results as one of the validation criteria vs the 'true' model later in this research.

Boundary analysis for 1 m composites. To demonstrate the significance of the lithological contrast on mineralisation, boundary analysis was first

performed on 1 m composites, resolution finer than the one used for point-support MPS simulation. Exploration and production holes were treated separately. Figure 4-6 shows the results within the mineralised shear zone. For drillholes of both types (Figure 4-6a – exploration drilling and Figure 4-6b – production drilling), the higher grades are observed in BIF while the diorites exhibit a gradual drop in grade away from the lithological contact.

The production drillholes exhibit lower grades within BIF: 3.2 g/t to 3.5 g/t vs 4.0 g/t to 4.2 g/t in exploration drilling. This is attributable to the fact that lithological boundary is not considered during the collection of samples in production drilling which is done on 1 m basis from the start of each drillhole, the net result – geological dilution.

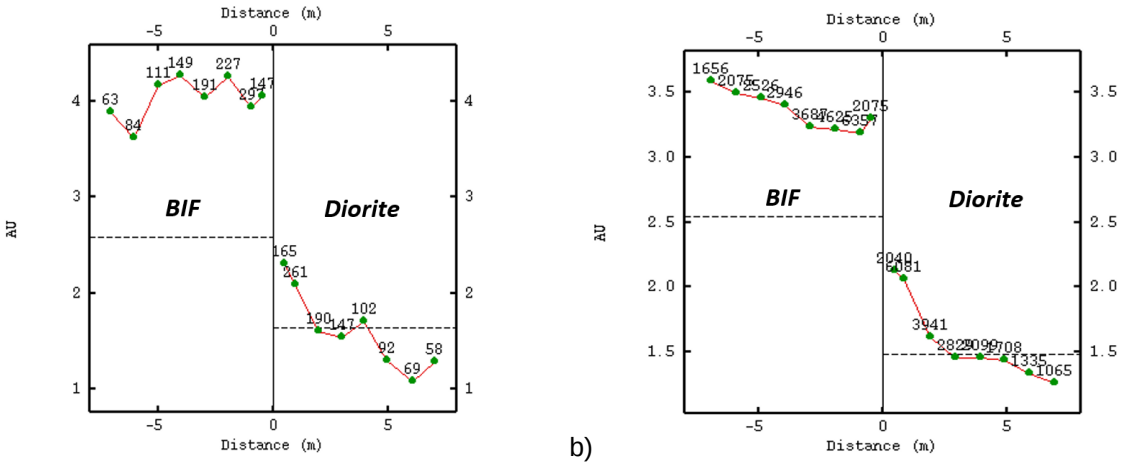
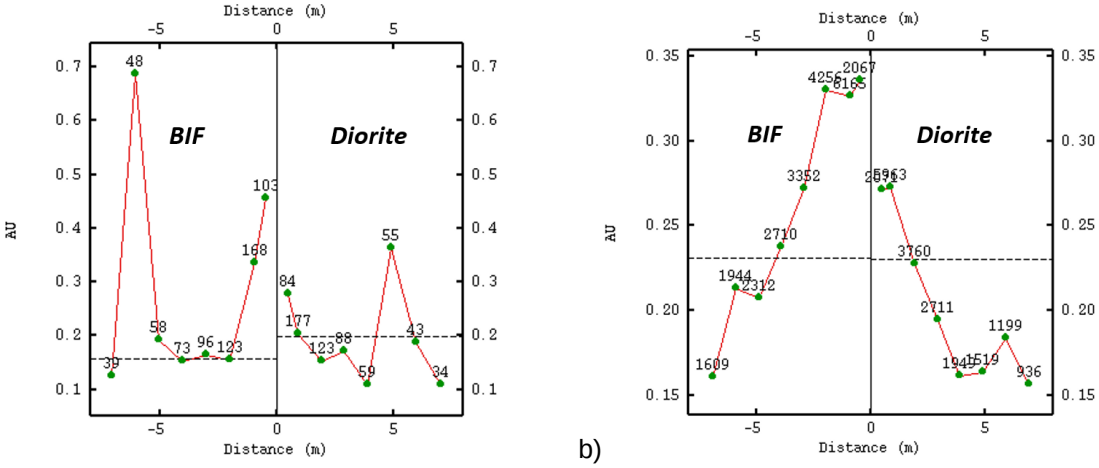


Figure 4-6. Boundary analysis within the shear zone for 1 m gold grade composites: a) in exploration, and b) in production holes

Outside the mineralised shear zone, the highest grades occur along the lithological contact (Figure 4-7a - exploration drilling and Figure 4-7b – production drilling). The reduction in average grade away from the contact is slightly more gradual in BIF and sharper in diorite. Within this zone of lower grade tenor, it demonstrates the fact that the sulfidation reactions, and, consequentially the mineralisation, were preferentially taking place along diorite-ironstone contacts, the zones of compositional difference (Sanislav et al., 2017). A feature common to structurally controlled deposits is observed - the grade patterns emphasise the structural feature that controlled the flow of mineralised fluid (Cowan, 2014).



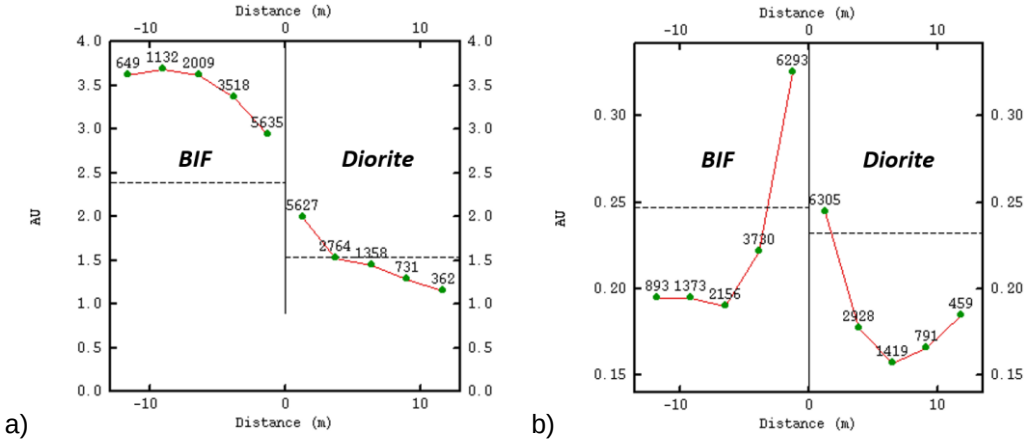
a) b)

Figure 4-7. Boundary analysis outside the shear zone for 1 m gold grade composites: a) in exploration, and b) in production holes

Boundary analysis for 2.5 m bench composites. Since the multiple-point statistics simulation is performed at the grid node resolution of 2.5 m x 2.5 m x 2.5 m, boundary analysis was repeated for 2.5 m composite support as well. The drillholes, exploration and production combined, were composited to 2.5 m intervals ignoring lithologies. The lithology prevailing within each interval was assigned to the interval post-compositing. Although not an ideal approach due to approximation, the effect of it was considered immaterial for the current research.

As to be expected, due to the change in support size, there is a more gradual boundary between the BIF and diorite within the mineralised zone (Figure 4-8a) in comparison to 1 m composites (Figure 4-6a).

Similar to 1 m boundary analysis (Figure 4-6b), elevated grades are exhibited along the contact of the two lithologies outside the mineralised shear (Figure 4-8b).

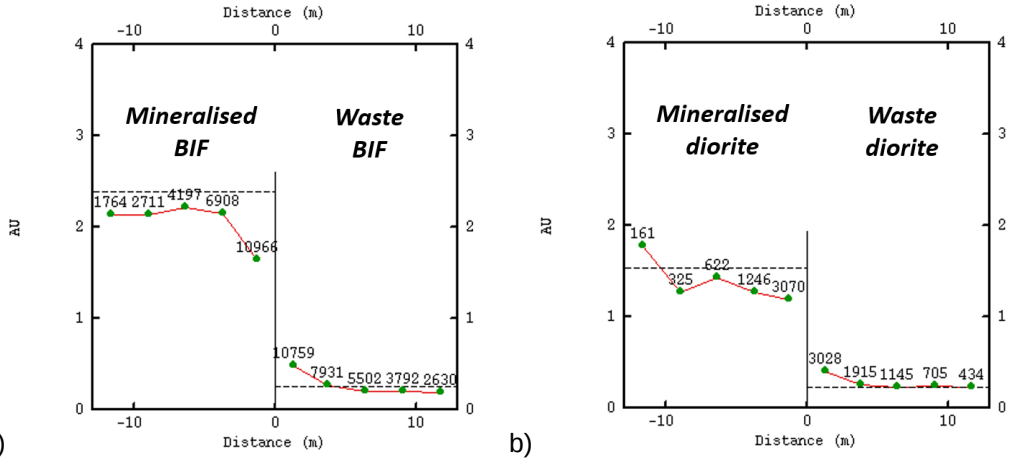


a) b)

Figure 4-8. Boundary analysis across lithological contact for 2.5 m gold composites: a) within the mineralised shear zone, and b) outside the mineralised shear zone

Boundary analysis for the 0.5 g/t cut-off grade used as a threshold between ore and waste is shown in Figure 4-9. Both BIF and diorite show quite a pronounced sharp contact along the mineralisation boundary: between mineralised BIF and waste BIF (Figure 4-9a), and between mineralised diorite and waste diorite (Figure 4-9b).

Later in this research, boundary analysis is repeated to validate gold simulation results.



a) b)

Figure 4-9. Boundary analysis across mineralisation contact for 2.5 m gold composites: a) within BIF, and b) within diorite

Understanding controls on mineralisation with spatial analysis

Conventionally, spatial behaviour of a variable is investigated by drawing experimental variograms to graph variability between pairs of points, followed by adoption of a suitable parametric model to define this behaviour in a stationary geostatistical domain. In the current research, due to high complexity of the geological patterns beyond two-point statistics, with the sigmoidal shape of ore shoots, absence of distinct geological proxies for delineation of mineralised shapes and non-stationarity, variography was not used to arrive at an understanding of spatial variability over the extents of the deposit. It was however resorted to later, at a post-simulation validation stage. Visual analysis and geological judgement were employed to understand important spatial relationships within the orebody. Leapfrog© geological software (Cowan et al., 2003) was used as a tool.

To understand first-order structures of the mineralisation, a method utilising X-ray down-plunge projection was used at the beginning of the spatial analysis. The method is based on two techniques. The first one, down-plunge projection, relies on dependencies between the patterns observable in grade and fluid flow paths which are associated with permeable zones resulting from structural deformation (Cowan, 2014). Down-plunge projection allows understanding the prevailing orientation of mineralisation by aligning the display of the grade data to look along the plunge of ore shoots. This direction often represents the intersection of main structural controls such as, in case of Geita Hill deposit, fold axial plane and bedding orientation. The down-plunge projection method has been used by geologists for the last century (Cowan, 2014).

It becomes particularly powerful when combined with another technique, Maximum Intensity Projection (MIP). Developed for applications in nuclear medicine (Wallis, Miller, Lerner & Kleerup, 1989), MIP allows projecting to the forefront of a point cloud the highest grade encountered along the rays running parallel from the viewing point to the plane of projection. The structures associated with localisation of extreme values, be it abnormal

nodules in tomography cancer screening or elevated mineralisation values in an orebody, are easily visualised.

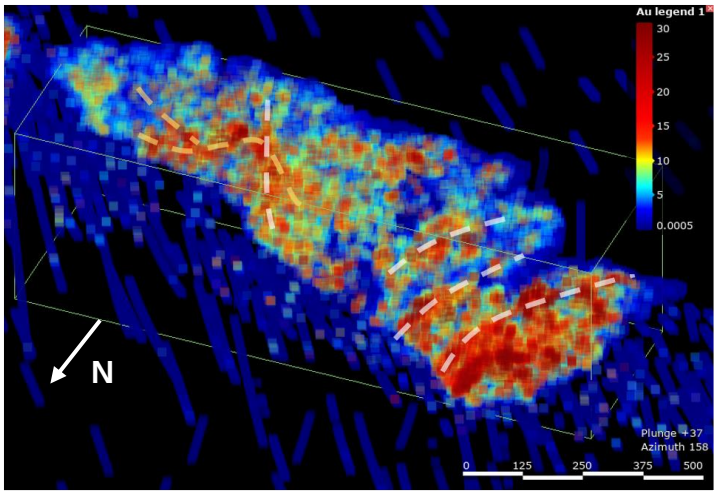
Example views of down-plunge projection method for Geita Hill deposit are presented in Figure 4-10. The lineations along important structural controls are shown in dashed lines of a different colour: the white lines represent the orientation of the lithological contact, the orange lines – the intersection of the 'braided' shears. Figure 4-10a shows prevailing controls on the high-grade shoots in different parts of the orebody: in the western part of the orebody, lithological boundary has the main influence on the orientation of ore shoots, in the east – it is the intersection of individual shear splays.

In Figure 4-10b, the view is aligned to look along the plunge of ore shoots predominant in the western part of the orebody, to north-east at 47^{o} from vertical. In Figure 4-10c, the MIP technique allows to see the 'braided' nature of the GHSZ, the splays are emphasised with dashed green lines.

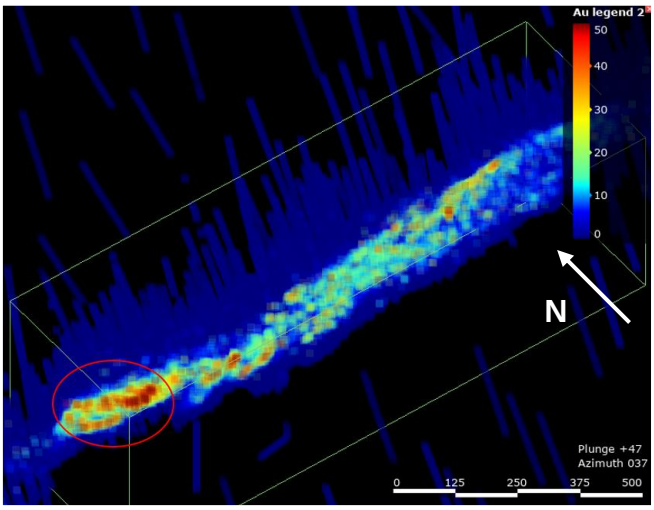
The process of identifying and interpreting individual shear splays was taken further by creating an isotropic interpolant of logarithmically transformed gold values using variogram-based RBF in Leapfrog© and relating it to different structural controls. This method facilitates structural interpretation in cases of densely and uniformly informed volumes as it allows for visualisation of the intersection of important structural controls easily while suppressing the effect of individual extreme values outliers.

Figure 4-11 shows grade shells at two cut-off thresholds enveloping gold mineralisation. Ore shoots are pronouncedly sigmoidal along the plane of the GHSZ.

a)



b)



c)

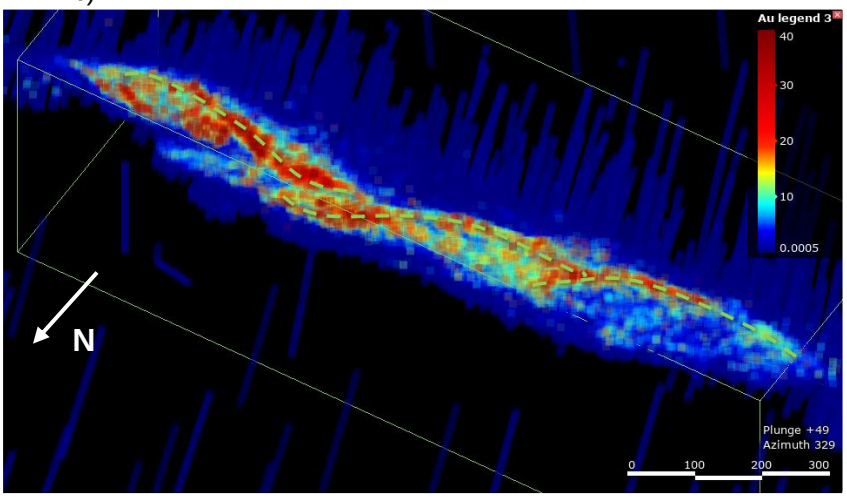


Figure 4-10. Views of Geita Hill deposit with Maximum Intensity Projection grade renderings:
a) view orthogonal to the GHSZ, b) view aligned to plunge direction of high-grade ore shoots in the west part of the deposit, c) view looking north-west highlighting the 'braided' nature of shearing. White dashed lines represent traces of bedding and lithological contact, orange dashed lines – intersection of 'braided' shears, green – 'braided' shear splays

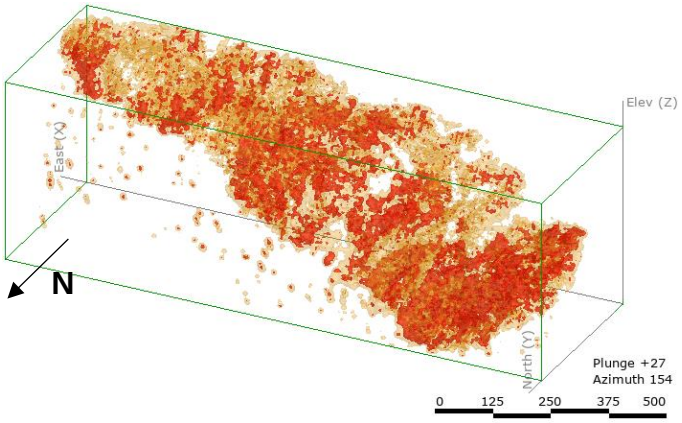
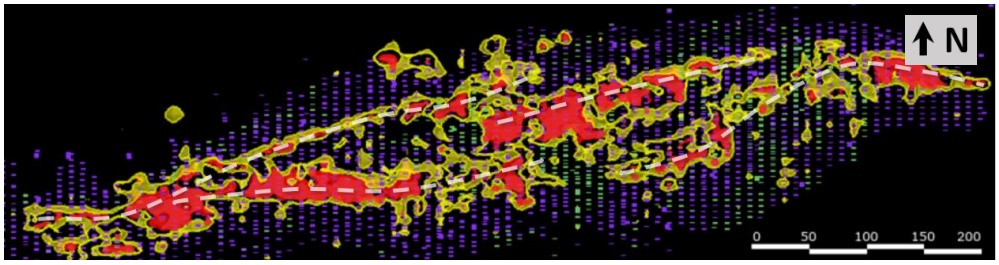


Figure 4-11. Isoclinal view of Geita Hill deposit showing two gold grade wireframes based on an isotropic interpolant of logarithmically transformed gold values

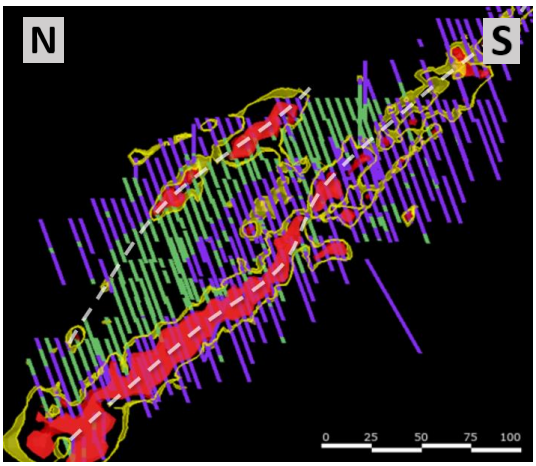
The envelopes at different cut-off values were scrutinised in 3D, as well as sliced on plans and sections to digitise significant shear splay surfaces. Figure 4-12 show examples of the process. The shear splays are mostly parallel to the lithological contact as highlighted in white dashed lines on Figures 4-12a and 4-12b, except rare exceptions when the attitude of the mineralisation has a steeper dip than the lithological contact. An example is shown in Figure 4-12c in orange dashed line. As a result of the iterative digitising exercise, five major shear splay surfaces were created (Figure 4-13).

To understand the interaction of the structural controls further, another approach was adopted - colouring the shear splay surfaces by the previously created implicit isotropic gold interpolant based on logarithmically transformed values and by distance to the lithological contact. It showed a strong dependence between the mineralisation and the frequency of the lithological contact per unit volume. Figure 4-14a displays the shear splays coloured by the isotropic gold interpolant, and Figure 4-14b the same surfaces coloured by lithology: diorite (green) and BIF (purple). The western part of the deposit has significant frequency of intercalation of the two main lithologies. Although not linearly dependent on the frequency of the lithological contact, the gold tenor is more robust in this part, with ore shoots elongated along the lithological contacts.

a)



b)



c)

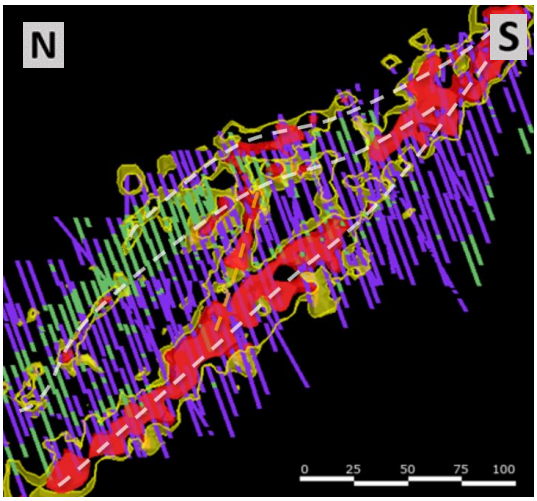


Figure 4-12. Plan (a) and section views (b) of the deposit showing the process of interpreting individual shear splays. The drillhole traces are coloured by logged lithology (BIF – purple, diorite – green). The yellow and red wireframes are grade shells of an isotropic RBF gold interpolant. The mineralised “braided” shears are shown as white dashed lines. Sometimes, the mineralisation is associated with brittle structures which have steeper attitude than the lithological contact (shown as an orange dashed line in (c)

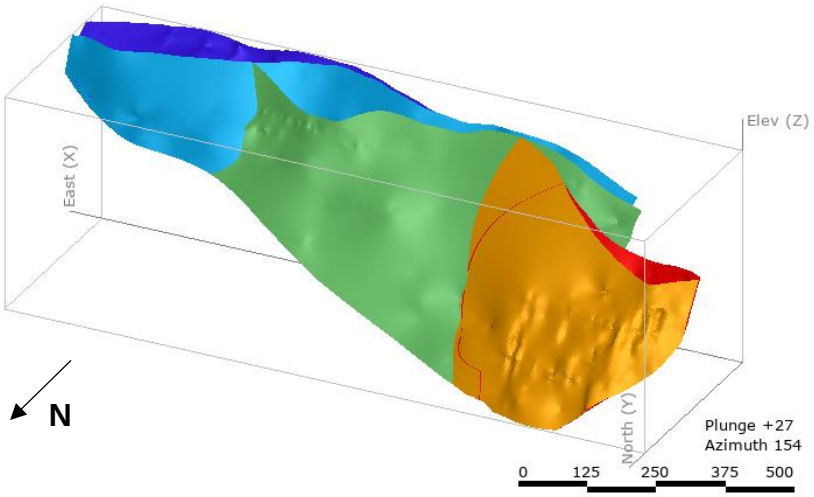
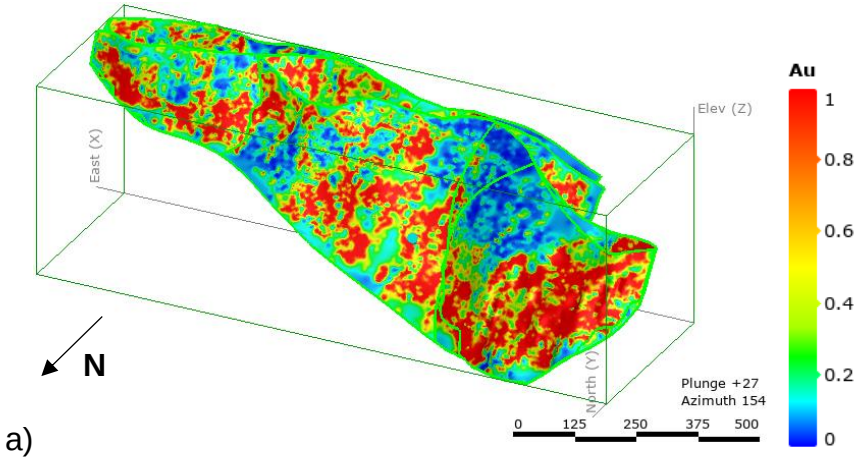
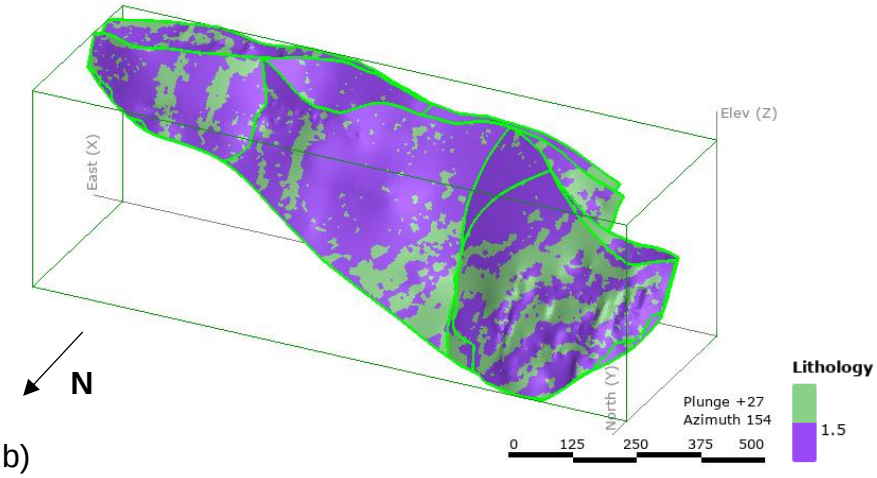


Figure 4-13. Isoclinal view of Geita Hill deposit showing five modelled "braided" shear splays



a)



b)

Figure 4-14. Isoclinal views showing "braided" shears coloured by: a) an isotropic gold interpolant based on logarithmically transformed gold values, b) lithology (BIF – purple, diorite – green)

Figure 4-15 displays the individual shear splays coloured by Au interpolant. The western part of the deposit exhibits strong tenor of mineralisation striking along the lithological contact. The eastern shallower modelled part of the deposit exhibits prevalence of different factors on mineralisation. The gold tenor is lower here and it is more a function of the intersection of individual shear splays, with the lithological contact playing a secondary role in the morphology of mineralisation.

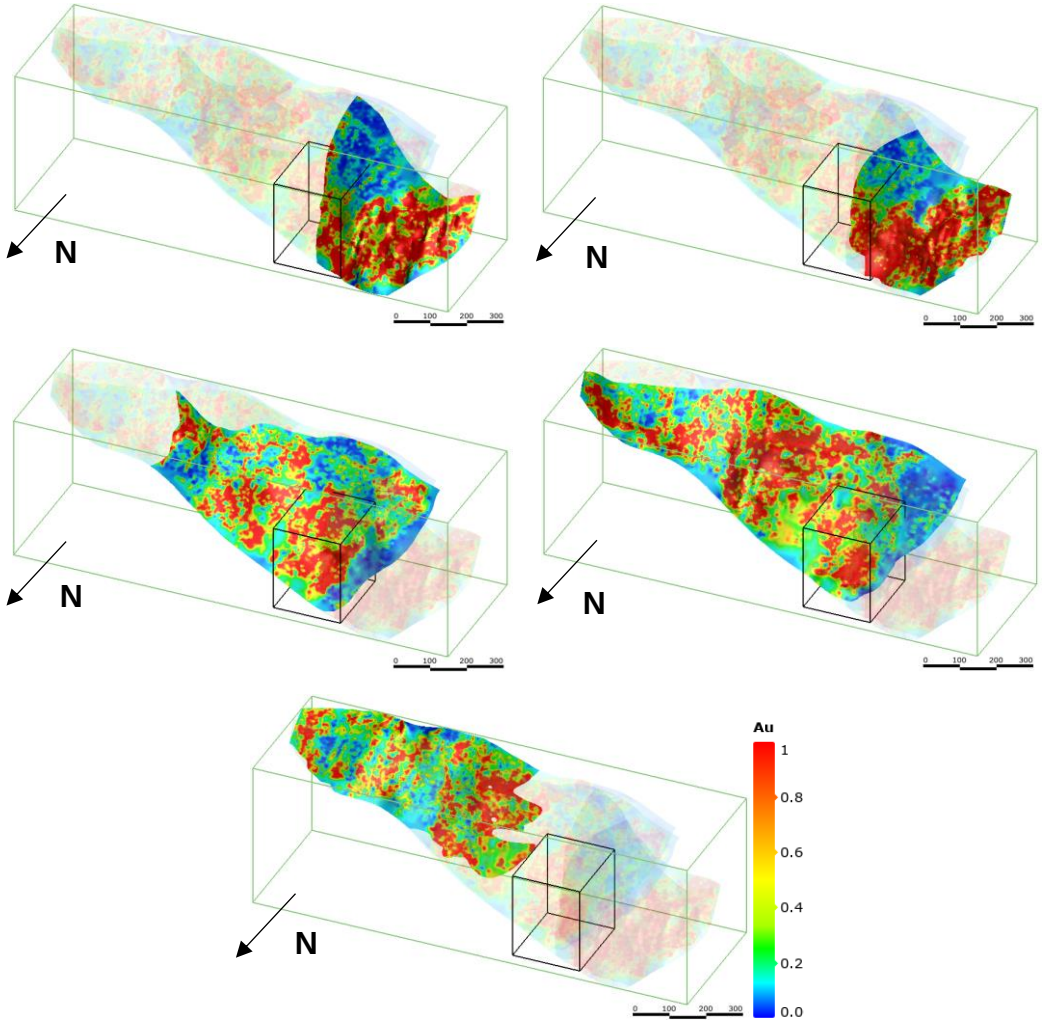


Figure 4-15. Isoclinal views showing each of the modelled shears coloured by an isotropic gold interpolant based on logarithmically transformed gold values. The box with black outlines represents the volume chosen for simulation.

The analysis described above aided understanding of the relation between the mineralisation morphology and features important for describing non-

stationarity, albeit no statistical tools have been utilised. It would be beneficial to advance this part of the research in future. For now, however, while the process of generating the 'true' model of the deposit continued for the whole extent of the mined-out portion, it was deemed appropriate to confine the extent of the simulation volume to the western portion of the deposit where the structural controls are more pronounced. The description of the simulation details will be presented further (Section 4.6).

For this area of the deposit, it appeared sensible to use the attitude of the lithological contact as a secondary variable to guide the simulation of gold in the non-stationary setting, due to elongation of the ore shoots in the direction of lithological contacts.

A potential field variable was created in Leapfrog© software to be used as a secondary exhaustively known variable to guide simulation of gold. Figure 4-16a shows an isoclinal view of the grid coloured by the variable. Figure 4-16b depicts "braided" shears with a slice through the potential field model. Non-stationarity of gold grade in Figure 4-14a is described by zonation of the potential field of the lithological contact variable in Figure 4-16b.

The principle of using potential field approach to geological modelling was pioneered by Lajaunie, Courrioux and Manuel (1997). The method considers the orientation measurements of a 3D field (azimuth, dip and plunge) and point observations taken on a geological interface, the latter being just one of a multitude of possible realisations of the field. The orientation data is assumed to sample the main anisotropy of the geological field – the observation points taken on the same interface have the same but unknown scalar value, the orientation measurements that do not belong to a specific interface have a scalar gradient of known values. The method proceeds by interpolating a scalar field the gradient of which is orthogonal to the structural orientations.

Once the potential field is generated it can be used to visualise any surfaces passing through the known points on interfaces while honouring the field.

These visualised interfaces represent iso-values of the potential field. The method has found broad application in geophysics and implicit modelling.

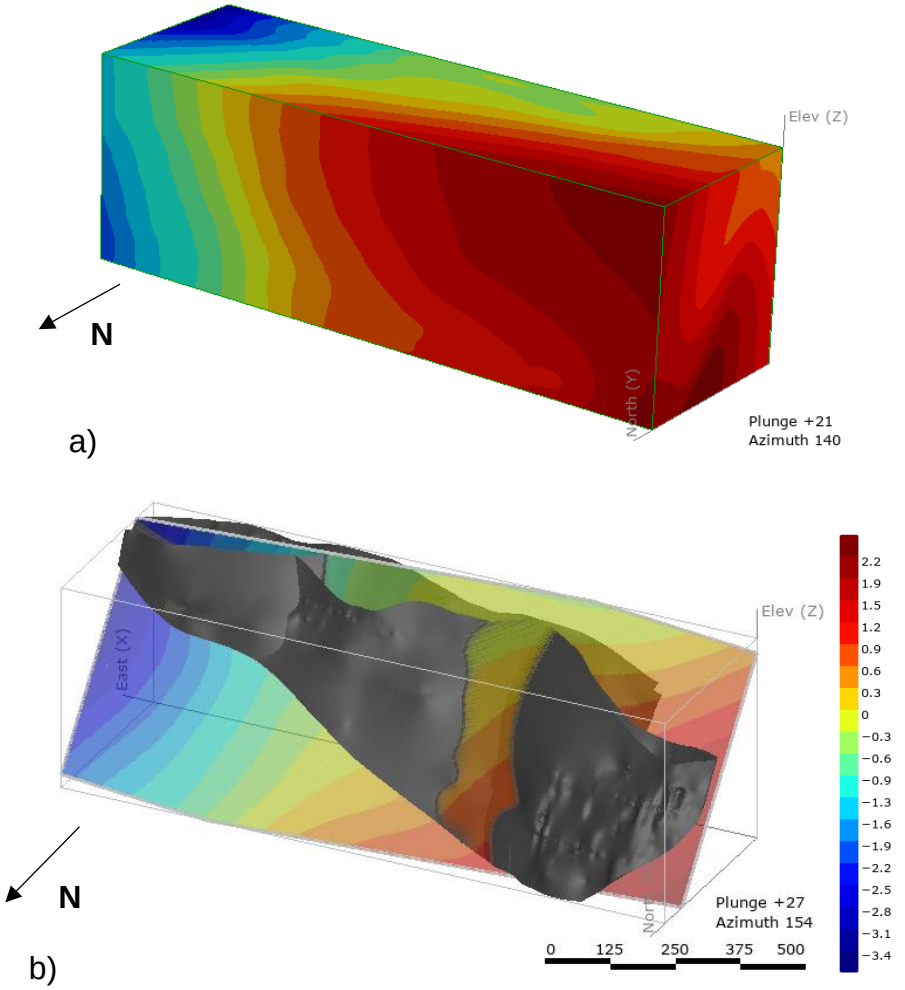


Figure 4-16. Isoclinal views showing: a) the potential field of lithological contact over the study area, and b) “braided” shears with a slice through the potential field model. Using the potential field variable to guide gold simulation is considered meaningful for Geita Hill deposit

When analysed against the modelled marker surfaces representing lithological interface (full details on the modelling process are in Section 4.4.3), it becomes apparent that the shear splays are localised strictly between the confining hinges of the S-fold, the zone of reduced stress field (Figure 4-17), as has been mentioned by different geologists (Cowan, 2009; Sanislav et al., 2017). It has been the first explicit 3D model relating this synergy for the deposit. The folded architecture of the deposit has been

described by different geologists however it has not been explicitly modelled beyond cartoons and sections.

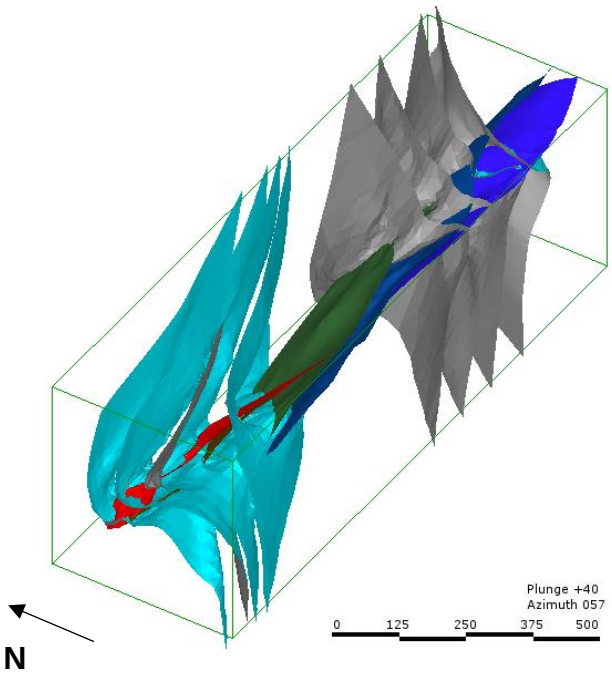


Figure 4-17. Isoclinal view of the 3D model of the folded lithological interface showing shear splays localisation between two hinge surfaces of the S-fold

Analysing the pattern of the lithological interface projected on the shear zone splays (Figure 4-14b), together with the slice through the potential field of the lithological contact (Figure 4-16b), it becomes apparent that the S-fold is plunging to the local north-west, and as such it is expected that the grade tenor might increase in that direction. This can be considered for further drilling targeting.

The spatial analysis described above assisted in understanding and identifying that the main auxiliary variables important to incorporate in the 'true' model and to further use in guiding MPS gold simulation are the attitude of the lithological contact, the attitude of shears, distance from the axial planes of the S-fold and, to some extent, the angle between the lithological contact surfaces and brittle shears.

Descriptions of explicit 3D models of mineral deposits are scarce in academic literature (Cowan, 2012). Mostly, schematics, maps and 2D

sections are presented, which may be simplified and insufficient to convey 3D morphology of deposits, not to mention, to incorporate it reliably into a Mineral Resource model. This is frequently a function of limited access university researchers have to exhaustive exploration and mining datasets. At the same time, published Mineral Resource reports seldom describe the morphology of the orebody genesis and structural framework (Cowan, 2012). The methods applied above are sound and available as part of commercial software (Cowan et al., 2003; Cowan, 2009): interpolation of planar structural measurements to create potential fields, modelling with respect to the structural geometric axes rather than plan and cross-sectional views and projections of important numeric interpolants on different structurally significant planes, as adopted above. Modelling the morphology of the orebody with reference to a structural geological framework lends itself to discovering unexpected patterns of mineralisation and allows to generate well-supported mineralisation domains for Mineral Resource modelling.

4.4.3 **'True' model construction**

The 'true' model with a total number of 14,200,320 nodes was created over the mined-out pit extent. Its geometry is presented in Table 4-1.

The distance of 2.5 m for node spacing in three directions was chosen for convenience of fitting into multiples of 5m to comprise different size mining units, and to arrive at a reasonable balance between the raw data point support of 1 m and the size of open pit SMUs while taking into consideration the considerable extent of the working area.

Table 4-1. 'True' model geometry

Parameter	*X*	*Y*	*Z*
Origin coordinates	52,052.5	9,061.25	1,091.25
Extents, m	1,290	400	430
Node spacing	2.5	2.5	2.5
Number of nodes	516	160	172
Total number of grid nodes	14,200,320		

The ‘true’ model was defined only in a well-informed portion of the grid, within a distance of 6 m from the available production drilling information (Figure 4-18). The drilling covered the mineralised part of the volume. The rest of the grid nodes were assigned background values for different variables.

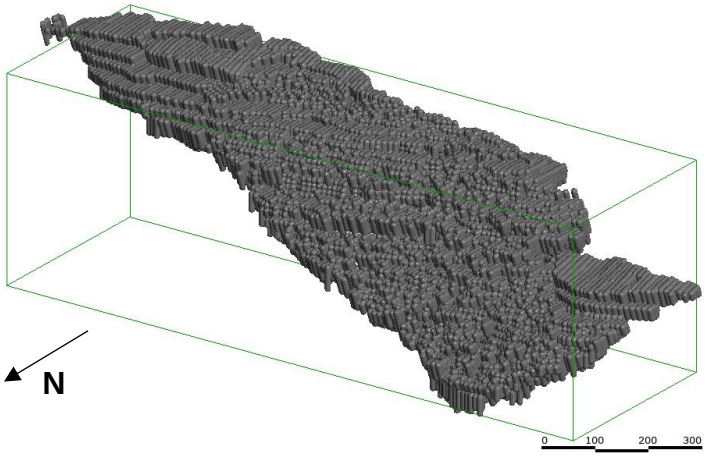


Figure 4-18. Isoclinal view through Geita Hill deposit showing the informed volume used to define the limits of the ‘true’ model

The following controls of significance for the morphology of ore shoots (Paranhos, 2008; Cowan, 2009; Brayshaw et al., 2010; Sanislav et al., 2017) were considered during the construction of the ‘true’ model:

- Folded structural architecture pre-dating emplacement of mineralised fluid flow;
- Brittle-ductile shear zones – conduits of gold-bearing fluids;
- Proximity to contact between the two main host rock lithologies, BIF and diorite intrusions.

The ‘true’ model contained several variables, grouped as primary – those impacting on the final Mineral Resource estimation (lithologies with associated densities, and gold grades), and secondary – those used to guide simulation of primary variables (limits of the fold hinge zone, attitude of lithologies and shears, distance from shear splays, etc). The process of identification of important variables was iterative.

The variables and the process of deriving them are described below. The variables' code names in the 'true' model are mentioned in brackets next to the full name for easier reference to the dataset if needed, e.g. *Lithology (variable 'LITH').*

Primary variables

Lithology (variable 'LITH')

The two main host rock types, BIF and diorite, comprise more than 99.9% of the deposit. Uncomposited exploration and production drilling was used for modelling the interface between the two lithologies. The logging of RC chips was of good quality and allowed for reliable definition of the lithological model.

Step 1: modelling significant marker surfaces of the overall attitude of the contact

Leapfrog© software was used to create a geological wireframe model of the interface by employing an implicit variogram-based RBF method on a categorical lithology variable. Albeit in an environment with an abundance of dense data, the implicit wireframe was not suitably representative of the spatial distribution of the lithologies, the connectivity of the units was not maintained, specifically in areas with thin intercalation, as shown in Figure 4-19. The same result would be expected from any variogram-based model. The idea of using variogram-based methods for the 'true' lithology model was abandoned and MPS resorted to. An approach of utilising an elementary training image in combination with rotational field was employed henceforth.

The RBF wireframe of the lithological contact was not discarded, however. It was used as a visual guide to produce a set of marker surfaces each one representing an interface of the same but unknown scalar value, to be used later for interpolation of the scalar field.

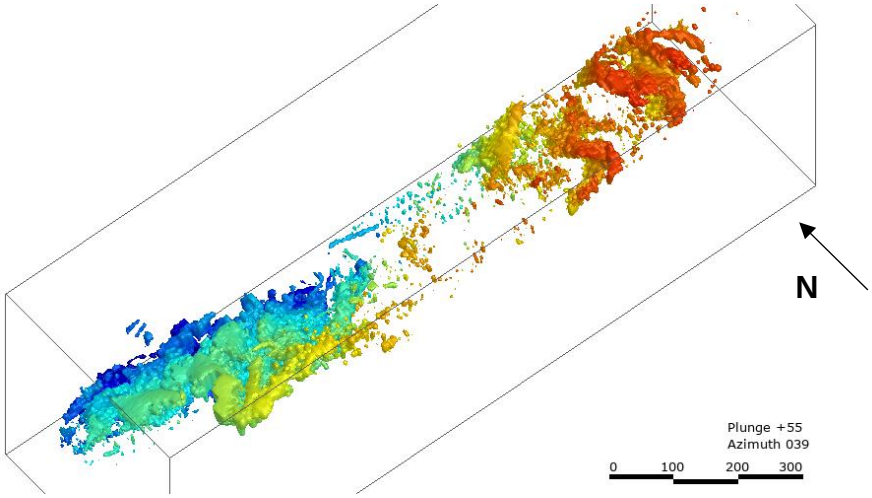


Figure 4-19. Isoclinal view of Geita Hill deposit showing the implicit diorite wireframe modelled from exploration and production drilling, coloured by elevation. A shortcoming of variogram based methods is apparent - connectivity of intrusive bodies is not maintained in the presence of fine intercalation even in densely informed areas

As mentioned before, this synthetic method resulted in a first 3D model of the kind to be produced for the folded architecture of the deposit. Isoclinal views of the marker surfaces are presented in Figure 4-20. Visual analysis of the images shows that shearing and elevated gold values are localised in the gold hinge zone.

Ideally, the scalar field is generated from structural readings of the contacts between the two lithologies which would require a consistent diamond drillhole data logging or orientation imaging, followed by declustering to generate the potential field. This approach can be used in future should the workflow be adopted for other orebodies.

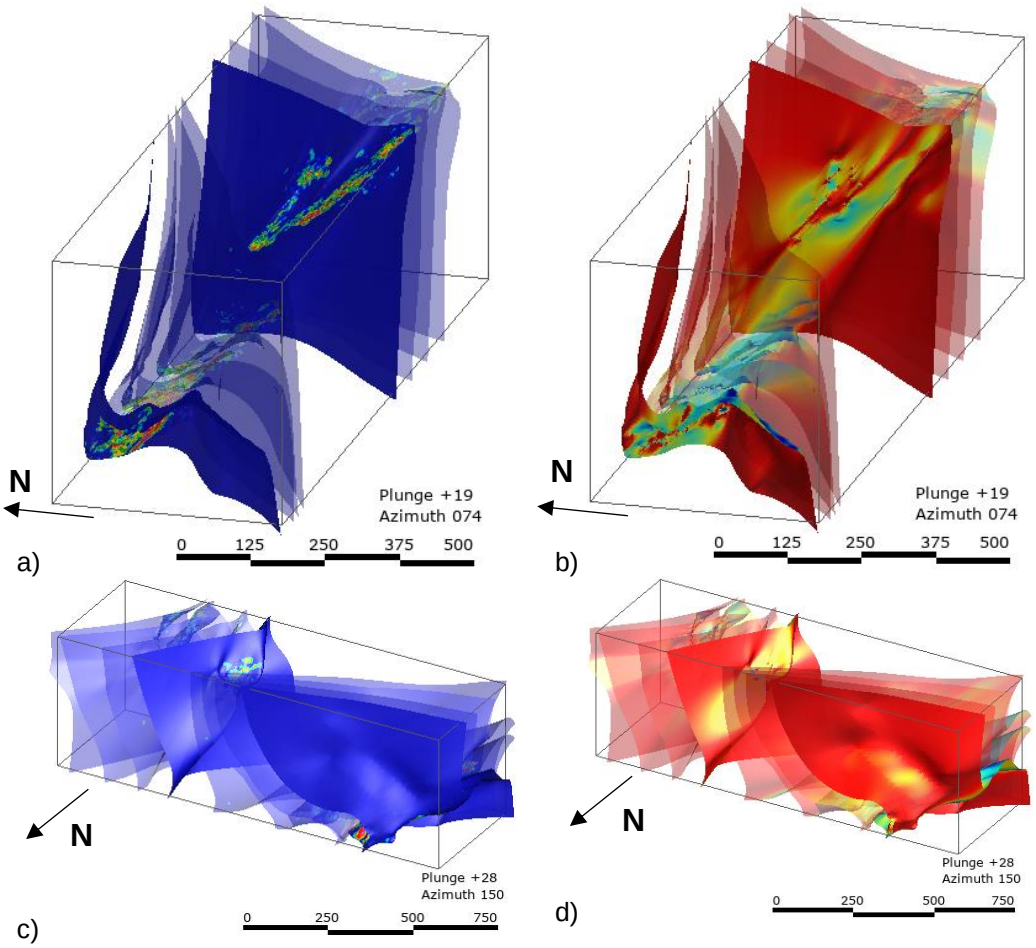


Figure 4-20. Isoclinal views showing the marker surfaces of the interface between BIF and diorite derived by manual digitisation.
a), c) surfaces coloured by an isotropic Au interpolant based on log-transformed values, b), d) surfaces coloured by the mesh face dip angle.
Visual analysis of the images confirms that shearing and elevated gold values are localised in the gold hinge zone

Step 2: creating rotational fields

The marker surfaces generated in the previous step were used to create rotational fields for the surfaces' true dip and dip direction to guide MPS simulation. The dip and dip direction were calculated using Datamine© software for each mesh triangle and stored in a point file with associated X, Y and Z coordinates representing the center of gravity of each triangle. The true dip and dip direction for the attitude of the contact were interpolated into the 'true' model using Inverse Distance Weighted (IDW) interpolation with the power of 1. Figures 4-21a and 4-21b present sectional views through the grid of true azimuth and true dip.

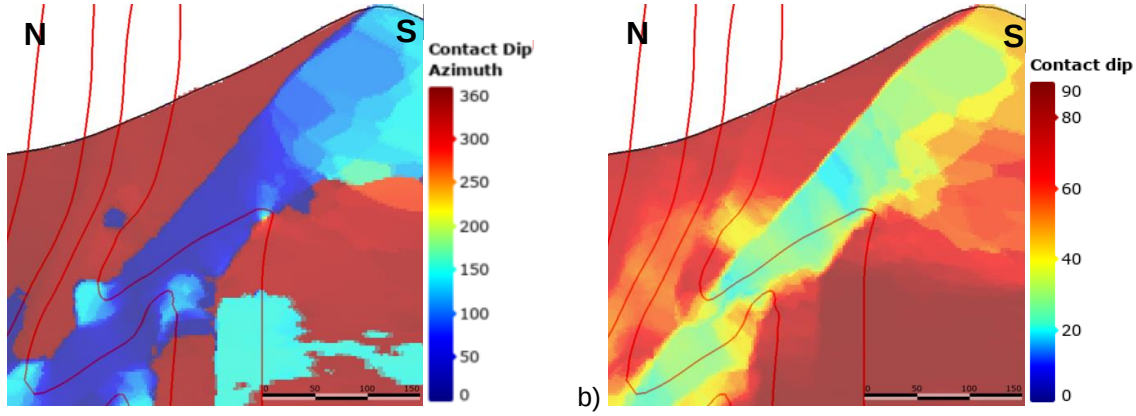


Figure 4-21. Sectional views through a grid with rotational variables for lithology contact simulation: azimuth (a) and dip (b). Red lines show attitude of lithological contact

Step 3: simulating a 'true' model for lithology

The 'true' model of lithology was created by MPS simulation using the DS. The elementary training image for the simulation consisted of a sequence of horizontally banded units of varying thickness, representative of those observed in different parts of the deposit: the minimum vertical size of the lithological units was 1 node, the maximum - 18 nodes. The geometry of the elementary training image is 50 x 50 x 380 nodes. An isoclinal view of the training image is shown in Figure 4-22a.

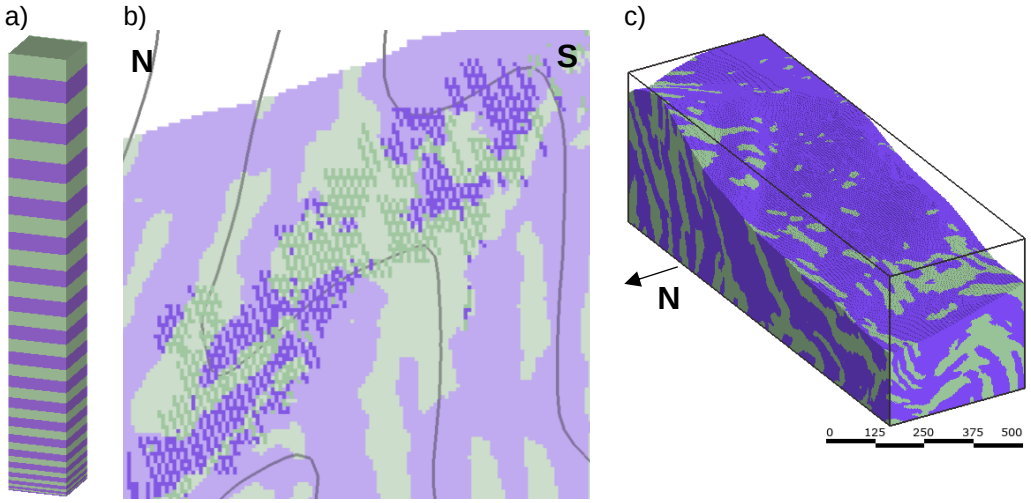


Figure 4-22. Objects in simulation of lithology for the 'true' model coloured by lithology: BIF - purple, diorite - green. a) elementary training image; b) vertical section through a conditional realisation (shown as semi-transparent) with the conditioning nodes shown as solid pixels. This simulation was used as the 'true' model of lithology shown in (c)

A conditional MPS simulation using the elementary training image is displayed in Figures 4-22b and 4-22c. The result of the simulation was adopted as the 'true' model for lithology. It displays a considerable improvement in continuity of finely intercalated units as compared to a wireframe generated using a variogram-based method (as shown in Figure 4-19).

A further enhancement to this step would consist of introducing conditioning by an additional continuous variable - true thickness of the logged units, and maybe even further – by employing dynamic calculations and eliminating a need for an explicit elementary training image representative of the true thickness.

Gold (variable 'AU')

A 'true' model of the gold grade was produced in Leapfrog© software using an RBF interpolant based on log-transformed gold grade values. To enable a successful log-transform, a pre-transformation shift was done by adding a constant of 0.0005 to each measurement, although non-negativity of the data was ensured prior to EDA. An RBF using a blending anisotropy trend along the five shear surfaces was used for the interpolation, with a variogram range of 20 m within the plane of continuity and anisotropy ratio of 5 : 1 (within the undulating plane of continuity : orthogonal to it). The structural anisotropic RBF trend approach was chosen to ensure better continuity along the shears, as they define the primary structural control on mineralisation.

Prior to adopting this approach, a few iterations of kriging had been attempted using raw gold values with different grade capping strategies, however, all of them yielded sub-optimal results, volumetrically overstating the high-grade values. This phenomenon is to be expected in the presence of highly skewed distributions of log-shape if no logarithmic transformation is applied to the data.

Figure 4-23 shows plan and section views through the 'true' model of gold.

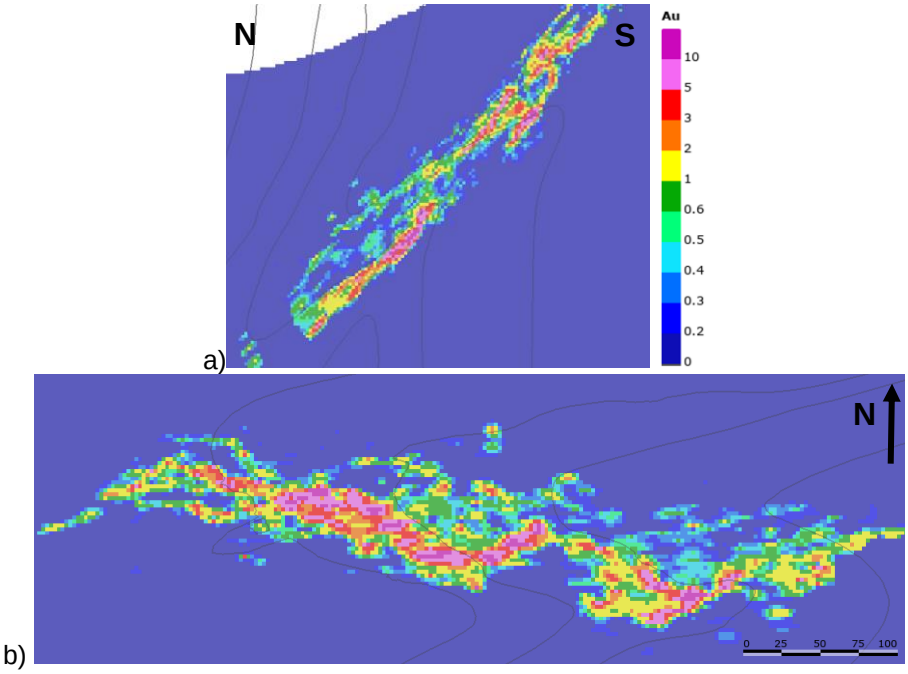


Figure 4-23. Section (a) and plan (b) views of the 'true' model of gold grade. The model was created using log-transformed gold values and a structural interpolant to maintain continuity along the five shear splay surfaces modelled in preceding steps. Grey lines show the attitude of the lithological contacts

Density (variable 'DENSITY')

Based on statistics of diamond core density measurements, global average values were assigned to the two lithologies: BIF – 3.0 g/cm^3, diorite – 2.7 g/cm^3. Estimation of density for the deposit is considered to be beyond the scope of the current research. In future, an MPS approach can be adopted for stochastic modelling.

Secondary variables

Simulation zone (variable 'SHVOI')

This variable has been defined broadly around the fold hinge zone (Figure 4-24), the latter associated with the volume of preferential stress field where the shearing and mineralisation have developed. Further in the text, besides being referred to as the "simulation zone" this zone is denoted as the "fold hinge zone" or the "shear zone". It is a loose usage of terms as volumetrically it has been modelled with insufficient accuracy and extends

slightly beyond the fold hinge zone into the footwall and hanging wall of the shear zone. This volume was used as a stationary domain confining the simulations. A mask object was applied to the rest of the simulation grid. The reasoning behind this was to reduce the unnecessary processing time of waste areas outside the fold hinge zone as the mineralisation at Geita Hill is shear-zone controlled and shears are confined to the fold hinge zone.

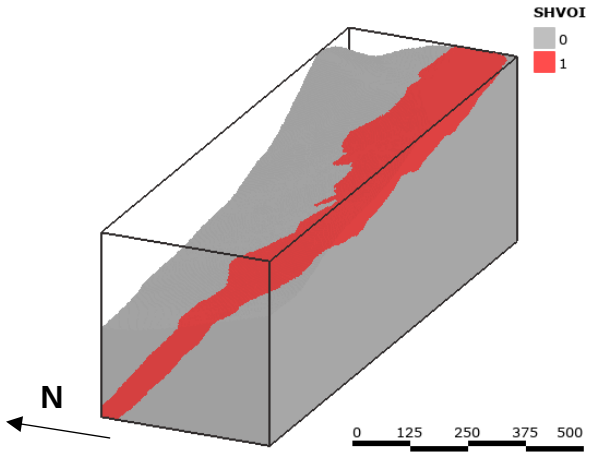


Figure 4-24. Isoclinal view of the 'true' model with the volume of the simulation zone (red)

The boundaries of the zone were modelled as two limiting surfaces from polylines digitised on sections (Figure 4-25). In future as an enhancement to the process can be developed - to eliminate manual digitisation and improve the precision in the definition of the stationarity, the modelling can be based on a derivative of a potential field representing the attitude of the bedding.

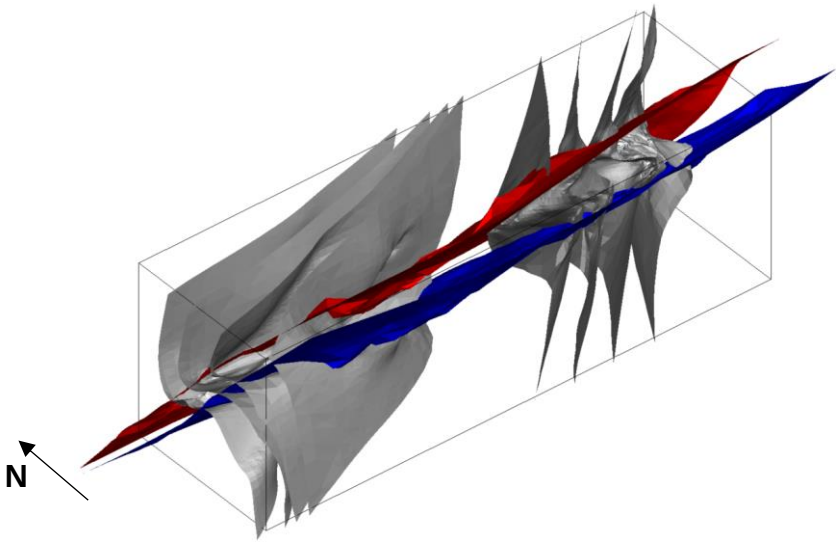


Figure 4-25. Isoclinal view showing folded contact surfaces (grey), the footwall (blue) and the hanging wall (red) surfaces of the fold hinge zone

Mineralised zone (variable 'SHSPLAYS')

The term mineralised zone has been used to describe a volume within the 'true' model with gold grade in excess of 0.3 g/t showing good overall connectivity. Extending along the GHSZ, this volume is distinct, connected and tabular with local irregularities introduced by minor "braided" shear splays.

This variable is a synthetic variable. It has been introduced to serve as an auxiliary variable to guide gold simulation. It would also allow an intermediate step of total mineralised volume simulation and validation prior to committing to simulation of gold values.

A few approaches to define the mineralised zone were tested in this research. In the first one, the zone was modelled as a tabular geological unit rather than a categorical gold threshold. This approach can be useful if brittle structures delineating the mineralised zone can be easily identified and correlated during diamond hole logging. An elementary binary training image similar to the one used in lithology simulation described above was employed (Figure 4-26a), together with the rotational fields for the shear zone true dip and azimuth (Figures 4-29a and 4-29b). The minimum thickness of the mineralised units in the TI was 1 node, the maximum – 15

nodes. The geometry of the elementary training image used in the simulation of the mineralised zone in the ‘true’ model is 50 x 50 x 258 nodes.

While depicting the main shear splays better geologically, with strong connectivity and tabularity, it also meant such a domain would include the whole range of gold values, including uneconomic grades. As such it would be less direct to perform primary validation on, and in several test simulation runs it proved to be such. Gold realisations utilising it as an auxiliary variable contained a whole range of gold values, including uneconomic grades below 0.5 g/t.

Figure 4-26 demonstrates the trial approach. The elementary TI for mineralised zone is shown in Figure 4-26a, the resulting simulation of the mineralised zone is in Figure 4-26b. This realisation has been used as an auxiliary variable to guide the simulation of gold shown in Figure 4-26e. The training image used as a source of patterns in this simulation is the ‘true’ model of gold (Figure 4-26d) and the mineralised zone (Figure 4-26c). This approach was abandoned as unsuccessful.

The next approach was to flag for the mineralised zone based on an anisotropic implicit gold interpolant without grade transformation. A cut-off value of 0.5 g/t was used for the indicator flagging. However, similar to the attempt above, several gold simulations performed using this approach also did not allow for direct validation of gold statistics. It was due to lack of direct connection between the ‘true’ model of Au, which was based on an implicit anisotropic RBF interpolant of logarithmically transformed grade values, and the mineralised zone indicator variable based on an implicit anisotropic RBF interpolant, however, of untransformed gold values. This approach had been discarded also.

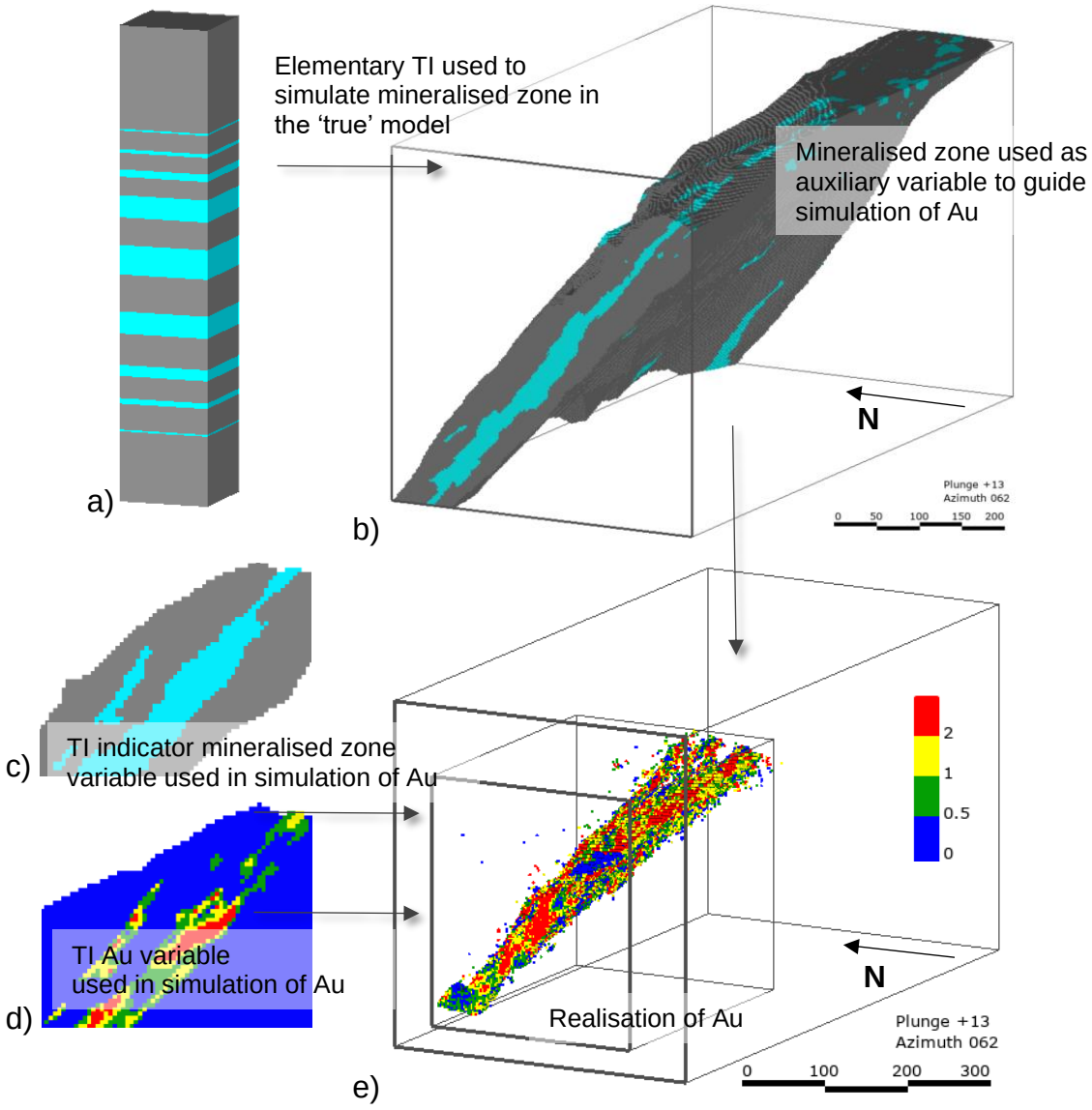


Figure 4-26. Objects in a trial simulation of the mineralised zone for the 'true' model: a) elementary training image of mineralised zone; b) isoclinal view of a conditional realisation of mineralised zone; c) and d) sections through the TI variables used in the simulation of Au; e) isoclinal view of a conditional realisation of Au. Gold realisations relying on this version of the mineralised zone contain a whole range of gold values, including uneconomic grades below 0.5 g/t which makes it less direct to perform an intermediate validation of the mineralised content

In the next attempt to define mineralised zone variable in the 'true' model, the indicator cut-off value of 0.5 g/t based on the logarithmically transformed gold krige of the 'true' model was used, however, this volume appeared quite disconnected.

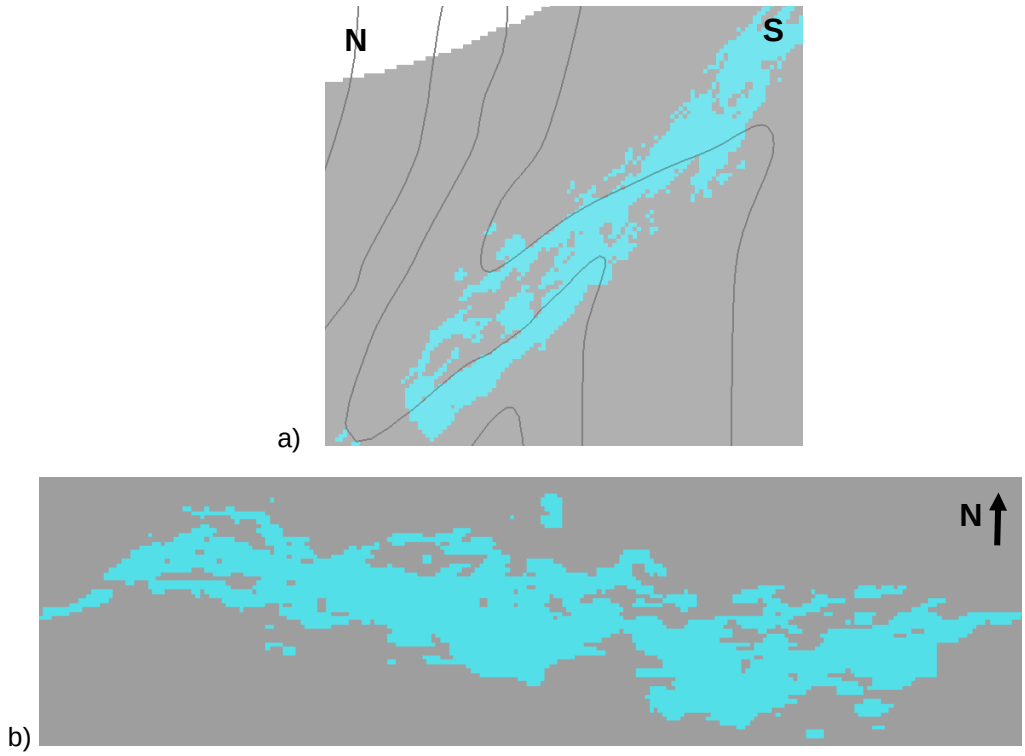


Figure 4-27. Section (a) and plan (b) views of the mineralised zone variable in the 'true' model. Grey lines on the section show the attitude of the lithological contacts and indicate a strong connectivity of the mineralised zone along the attitude of the contact

Finally, the value of 0.3 g/t of logarithmically transformed gold in the 'true' model was chosen to represent the mineralised volume better. In the final runs of simulation, the mineralised zone refers to the ensemble of nodes equal to or larger than 0.3 g/t. Figure 4-27 shows the final model of the mineralised zone, with strong connectivity displayed not only along the shear splays but also along the lithological contact.

Categorical gold variable (variable 'AUCAT')

In several initial runs of gold simulation, it has proven difficult to achieve good connectivity of high grades. Introduction of a categorical gold variable in a co-simulation mode with the continuous Au variable allowed to overcome the problem. Four cut-off values were chosen (0.3 g/t, 0.5 g/t, 1 g/t and 2 g/t): 0.3 g/t to coincide with the mineralised zone indicator variable, 2 g/t chosen to represent an underground mining cut-off value, 0.5 g/t and 1 g/t to visually allow sufficient representativity of the medium

grade categories in the mineralisation morphology. Figure 4-28 shows a plan and section views of the categorical gold variable.

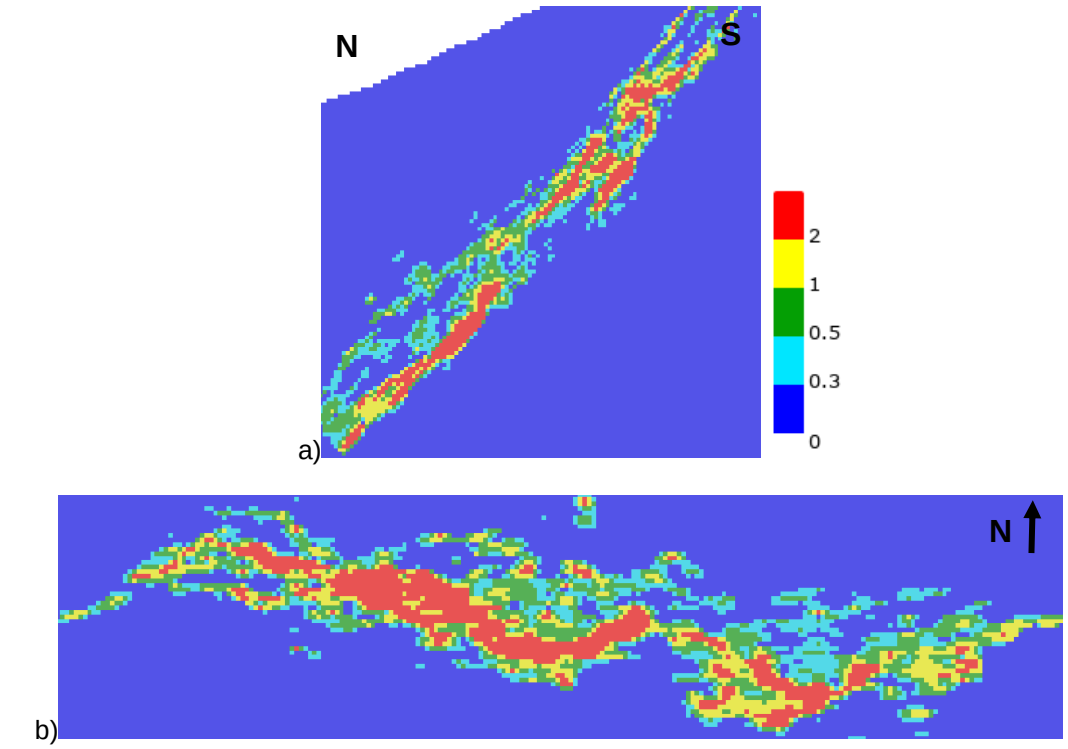


Figure 4-28. Section (a) and plan (b) views showing the categorical gold variable in the 'true' model

The decision for the number of categories and the choice of cut-off values is possibly another topic for research. Important considerations are relevance for mining and sufficient volumes of multiple-indicator node clusters of the mineralised morphology. It might be enough to limit the number of indicator values to two, namely, at waste/marginal-ore boundary and marginal-ore/full-grade-ore boundary. It might result in better pattern reproduction due to sufficient size of indicator geobodies and representative neighbourhood in each category. As shown in Figure 4-53 (Section 4.6.3), adopting three indicator thresholds improved the morphology of the simulated gold patterns.

Rotational field of the lithological interface (variables 'CODIP' and 'CODIPDIR')

The true dip and azimuth for the attitude of the lithological contact were estimated from the marker surfaces described above using the IDW

estimator suitable for angles. Figures 4-21a and 4-21b show example sections through the block model for the two variables. These variables were used as auxiliary to guide simulation of the two main lithological units based on an elementary training image when constructing the 'true' model. The same approach was used for the further conditional lithology simulation over the SG.

Rotational field of the shear splays (variables 'SHDIP' and 'SHDIPDIR')

Similar to the lithological contact, true dip and azimuth for the mineralised zone were estimated into the 'true' model. Figures 4-29a and 4-29b below show sectional views through the fields. These variables were used in the initial runs of the mineralised zone simulation in conjunction with an elementary training image, displayed in Figure 4-26a, to produce a simulation as shown in Figure 4-26b.

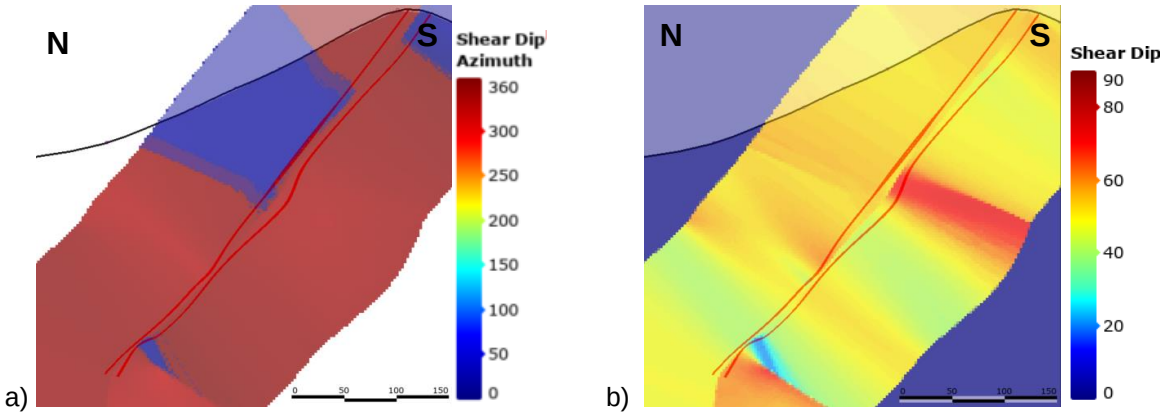


Figure 4-29. Section views through the 'true' model showing (a) true dip azimuth and (b) true dip of the attitude of mineralised shear splays.
Only cells within the simulation zone are kriged for the rotational field variables. Cells above topography are shown in semi-transparent colour.
Red lines represent shear splays surfaces

The potential field of the lithological contact (variable 'POTENCO')

As in most geological settings, the model area is characterised by high non-stationarity. Analysing the primary structural controls (shear splay surfaces) coloured by gold interpolant (Figure 4-30a) and comparing it to the shear splay surfaces coloured by lithology (Figure 4-30b) and further to surfaces coloured by the potential field interpolant of the lithological contact (Figure

4-30c) reveals substantial dependency. The volume over which simulation will take place in the following sections was chosen to cover a high-grade area with strong lithological control, albeit not stationary. It is shown as a black box in the Figures 4-27 a-c.

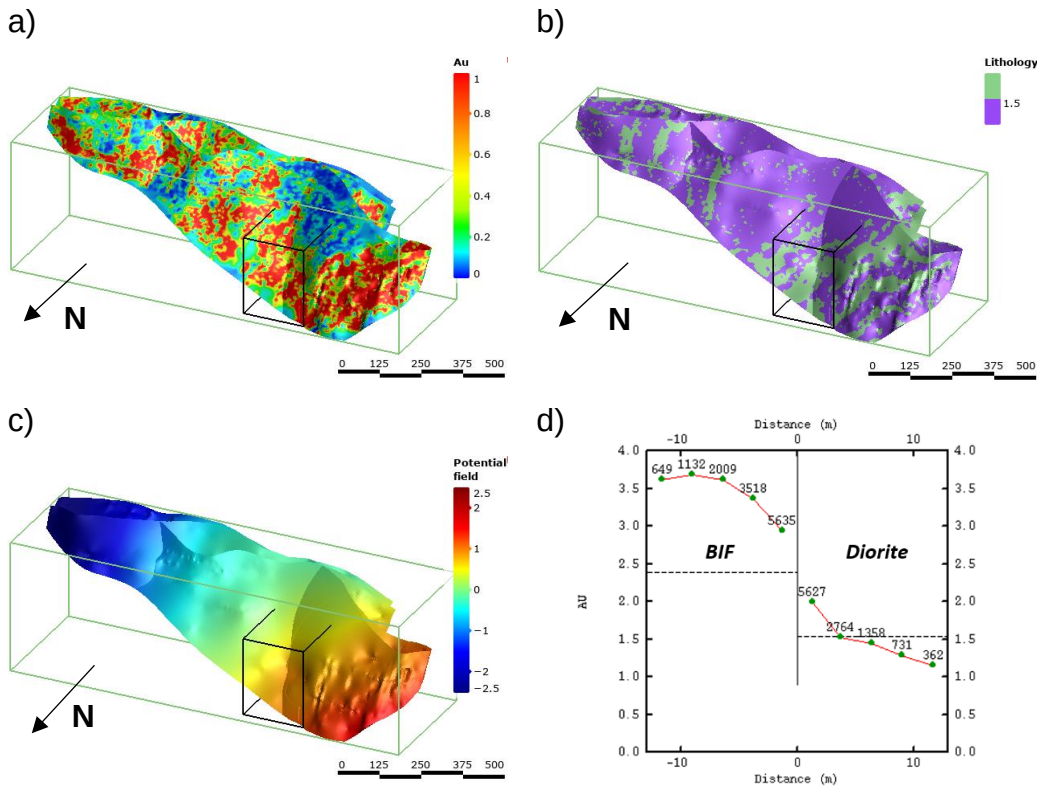


Figure 4-30. Schematic in support of using potential field of the lithological contact variable as the main variable describing non-stationarity.
a-c) Isoclinal views showing “braided” shears:
a) coloured by an isotropic gold interpolant based on logarithmically transformed gold values,
b) coloured by lithologies (BIF - purple, diorite - green);
c) coloured by the potential field of lithological contact.
Using the potential field variable to guide gold simulation is considered representative of the non-stationarity setting for the western portion of Geita Hill deposit

The boundary analysis between the lithologies (Figure 4-30d) indicates that the mean gold grade increases within BIF away from the lithological contact. As such, it appeared to make sense to use the attitude of the lithological contact as a secondary variable to guide the simulation of gold in the non-stationary setting. The process of choosing the variable relevant to non-

stationarity would certainly benefit from supporting rigorous numerical analysis. It is considered a good topic for further research.

The potential field variable for the lithological contact in the ‘true’ model is shown in Figure 4-31.

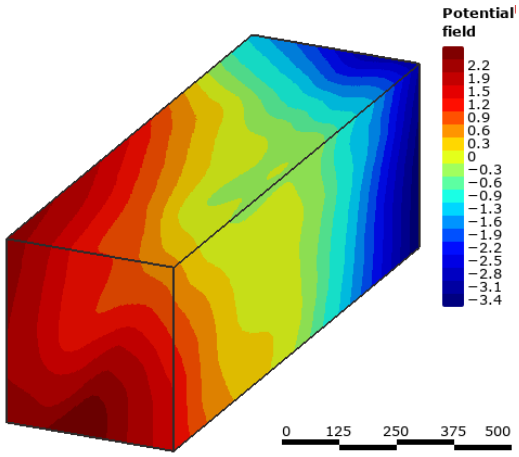


Figure 4-31. The potential field of lithological contact variable in the ‘true’ model

Distance to shear splays (variable ‘DIST2SH’)

A variable representing the distance to “braided” shear splays was created using the five shear splay surfaces. Usage of this variable did not result in significant success in the current research, but in future, it can be attempted to further help localise the gold simulation.

It would specifically make sense to use the variable for guiding simulations if major brittle discontinuities can be logged reliably in diamond core drillholes and explicitly identified and correlated. It can also be valuable for exploration potential targeting when MPS is used to simulate structures and lineations identified from exhaustively known geophysical data.

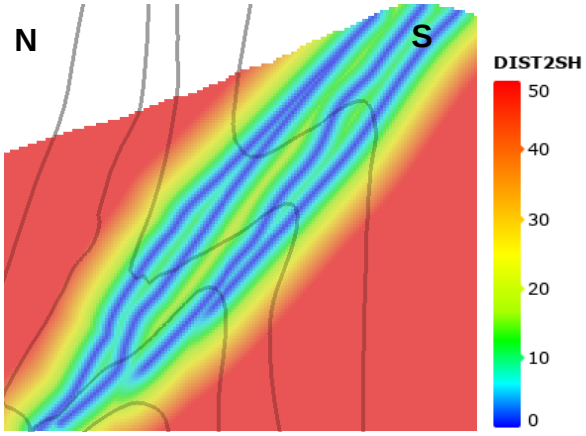


Figure 4-32. Section showing distance to shear splays variable in the 'true' model. Grey lines represent the overall attitude of the lithological contact

The interface between the training and validation volumes (variable 'OP1UG2')

To enable benchmarking the performance of simulations, the 'true' model has been split into two domains in similarity to ML framework and mimicking an interface between a 'mined-out' known and 'to-be-mined' unknown portions of the deposit. The top part of the model was used in further simulations as a training image. The simulations were performed over the bottom part. Subsequently, the results were compared to the 'true' model over the same volume. Figure 4-33 shows the interface coded in the 'true' model.

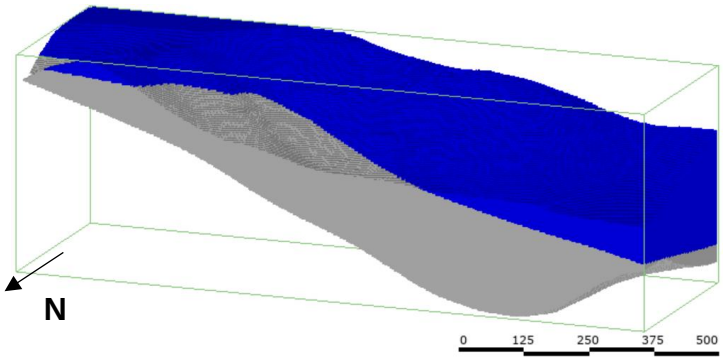


Figure 4-33. The interface between the training image 'known' (blue) and testing 'unknown' (grey) portions of the 'true' model

4.5 Derived conditioning data and its uncertainty

Derived conditioning data was generated from the 'true' model, by first creating 84 subsets (1/3 of all possible combinations) of pseudo drillholes,

with spacing of 16 x 16 nodes in X and Y directions which is equivalent to 40 mE x 40 mN. This spacing served was used in the study for either, a conservative approach to categorising the material into the “Inferred” Mineral Resource, or an optimistic approach for the “Indicated” Mineral Resource. The 84 subsets were coded into the ‘true’ model and, subsequently, the corresponding nodes extracted for statistics.

Prior to undertaking simulation, 84 instances of pseudo-drillhole sets were accessed to understand inherent uncertainty associated with random placement of a drillhole grid of the same spacing. The comparison was done within the fold hinge zone. Though the pseudo drillholes are regularly spaced and choosing any of them as a conditioning data set going forward is not expected to result in any bias it was still considered of interest to understand the spread of uncertainty among the data sets.

The uncertainty in the gold statistics associated with the placement of a 40 mE x 40 mN grid of conditioning data is described below.

4.5.1 **Mineralised zone proportions uncertainty**

The box-whisker plot in Figure 4-34 shows the spread of mineralised zone proportions across the 84 scenarios in comparison to the proportion in the ‘true’ model (red line). The proportions of the nodes flagged as mineralised zone in the data scenarios are very close together. The proportions within the 50% of the scenarios contained between the 25th and the 75th percentiles are between 0.763-0.77. The spread between the maximum and minimum values is very narrow, 0.757-0.777, which is 1% around the ‘true’ model mean value. This is indicative that no significant bias in the mineralised zone proportions can be expected when any one of the sampling grids is chosen as a conditional dataset for further simulation.

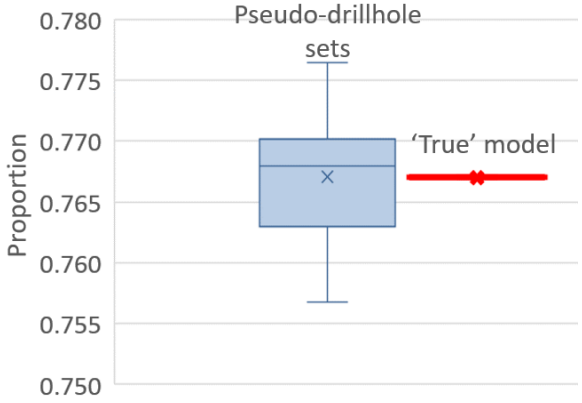
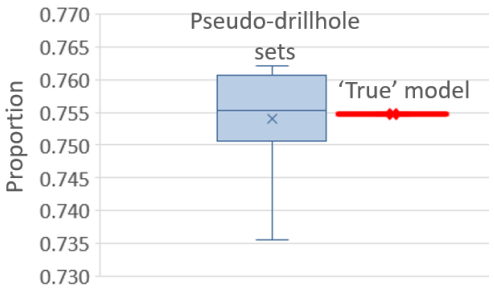


Figure 4-34. Box-plot of the mineralised zone proportions across pseudo drillhole scenarios (blue) vs the 'true' model proportion (red)

4.5.2 Lithology proportions uncertainty

The proportion of lithologies across all the scenarios has insignificant spread about the proportions in the 'true' model. The differences in mean values between the scenarios vs the means of the 'true' model are immaterial. For the fold hinge zone (Figure 4-35a), the two mean values are approximating 0.755, for in the mineralised zone - 0.797. For the fold hinge zone, the spread of simulated proportions of BIF is between +1% and -3%. For the mineralised zone, the spread of minimum and maximum values is symmetrical, ±3% around the 'true' model mean.

a)



b)

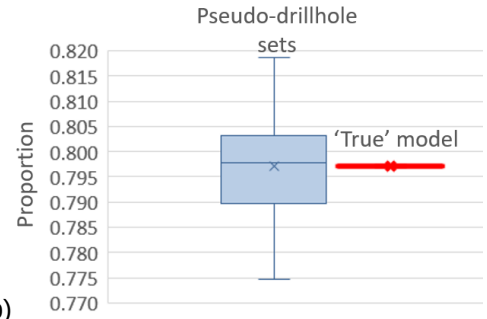


Figure 4-35. Box plots of BIF proportions across pseudo drillhole scenarios (blue) vs the 'true' model (red): a) in the fold hinge zone, b) in the mineralised zone

The conclusion stemming from the analysis of the categorical variable proportions across the pseudo drillhole sets vs the proportion in the 'true' model is that the pseudo drillhole grids reproduce the global mean very well

and show acceptable symmetry and a tight spread of the lithology proportions around the values in the ‘true’ model. Similar to the analysis of the mineralised zone proportions, it indicates that negligible bias is expected when any sampling grid is chosen from the ensemble as a conditional dataset for further simulation.

The information on the spread of proportions (+1% and -3%) can be used in future when imposing the global proportions conditioning for lithology simulations of the deposit, to define allowed spread of uncertainty.

4.5.3 **Gold grade uncertainty**

Comparison of gold grade statistics across all 84 grid scenarios exhibits a more pronounced spread than for the mineralised zone or lithology proportions. Cumulative histograms of gold within BIF and diorite are shown in Figures 4-36a and 4-36b respectively. The Au histograms of the scenarios in its entirety are having a symmetrical spread around the ‘true’ model cdf, at first glance, implying a good reflection of the underlying histogram of gold grade in the ‘true’ model.

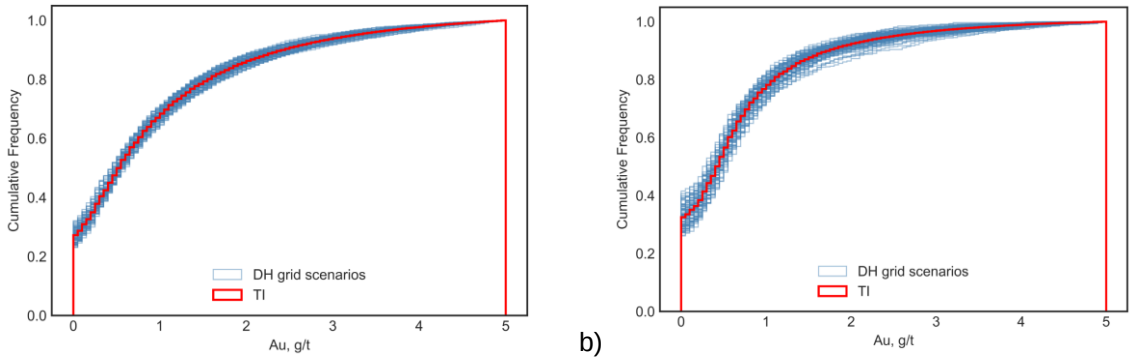


Figure 4-36. Cumulative density functions of Au across pseudo drillhole scenarios vs ‘true’ model: a) Au in BIF; b) Au in diorite.
The nodes within the mineralised zone in the ‘true’ model are selected

The analysis is taken further, to understand the uncertainty in the mean value among the scenarios. Figure 4-37 shows distributions of the mean values for the 84 data scenarios in the mineralised shear zone: Figure 4-37a for the BIF, 4-37b for the diorite, and 4-37c for the two combined lithologies. In all three histograms, the mean values for 90% confidence

intervals (shown as black dashed lines) across the grid scenarios spill outside of the desirable ±15% error intervals (shown as blue dashed lines), in the upper confidence limit. This indicates that the pure choice of positioning the 40 m x 40 m unbiased sampling grid in the given 'true' model brings into the simulation some uncertainty prior to committing to any estimation technique. Although, comparison of the errors at a specified confidence interval is done for point support rather than being related to a mining period, as required by reporting codes, it is considered to be of interest in understanding the inherent uncertainty of positioning an unbiased sampling grid in the deposit.

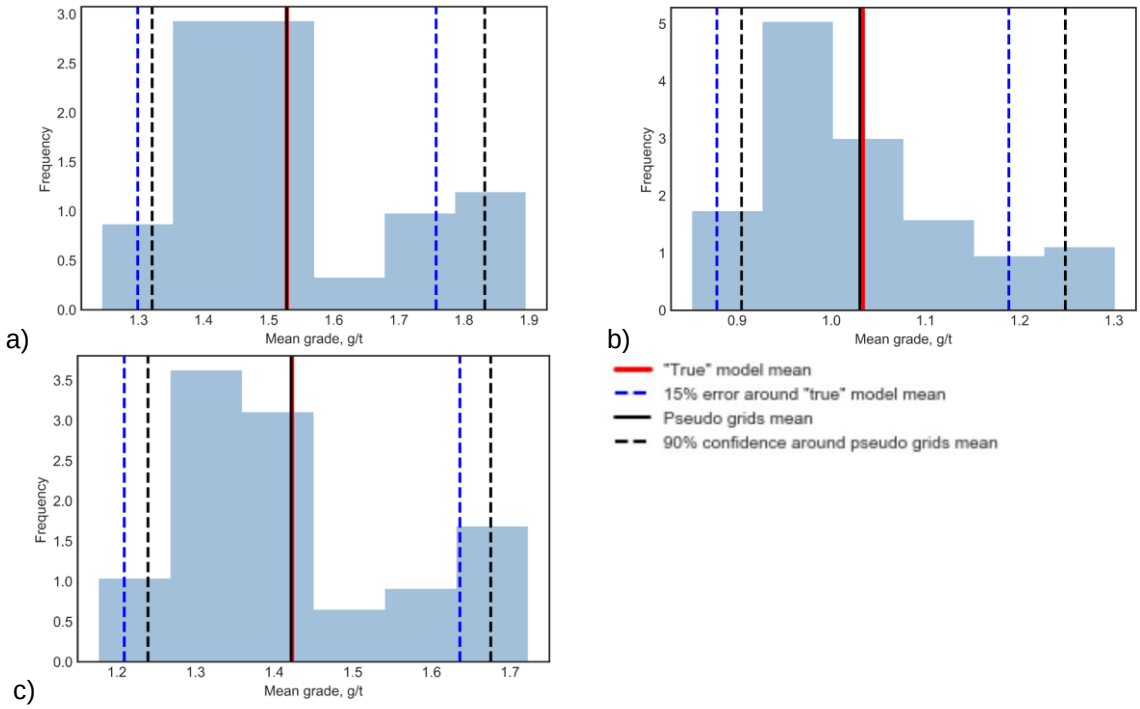


Figure 4-37. Histograms of the mean grade across pseudo drillhole scenarios in the mineralised shear zone: a) in BIF, b) in diorite, c) in two combined lithologies. There is no bias in gold mean values observed in the ensemble of the scenarios at the point support, however, individual scenarios spill outside of the desired 90% at 15% error at the upper confidence limit

A more pronounced spread of gold statistics across the drillhole scenarios in comparison to the categorical variables can be explained by the fact that the categorical variables have a more pronounced continuity while the gold grade distribution is more erratic and nuggetty. As such, there is significant

uncertainty associated with the location of the drill grid, albeit an unbiased sampling pattern is ensured.

4.6 Simulation with sparse data conditioning

To benchmark the performance of simulations, the ‘true’ model was divided into two domains as described above: the top portion was preserved as a training image, while the simulations were undertaken over the bottom half. Figure 4-33 demonstrates the division of the informed volume into the two domains. Only the volume within the fold hinge zone was subjected to simulation. Nodes along artificial drillhole traces on a 40 mE x 40 mN grid orthogonal to the attitude of the orebody were extracted from the ‘true’ model and used to condition the simulations. This set of nodes corresponded to one of the conditioning grid scenarios described in the preceding section. The conditioning grid also included all informed nodes in the ‘mined-out’ top portion of the ‘true’ model, to ensure a continuum between the known and the to-be-simulated unknown parts of the model. The simulation results were validated versus the training image, pseudo drillholes and the ‘true’ model available over the unmasked simulation volume. This approach is followed in each variable simulation.

A few iterations of the simulation were undertaken in different parts of the deposit. Finally, the volume for simulation was chosen in a highly non-stationary part of the deposit (Figure 4-38) with well-understood controls on mineralisation to investigate the performance of the DS in such settings.

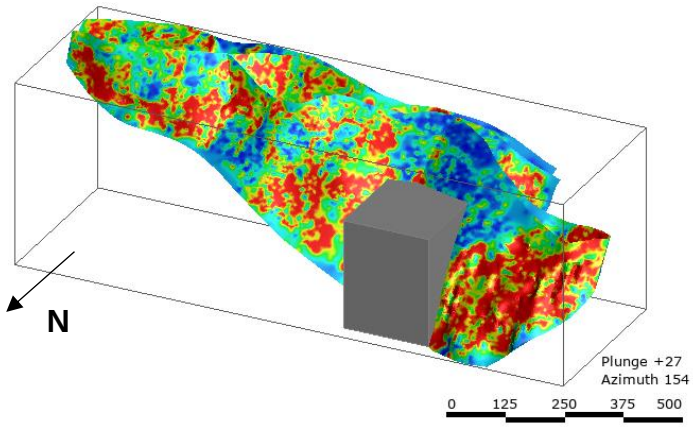


Figure 4-38. Location of the simulation grid (grey box) in relation to the 'true' model of Geita Hill deposit

The simulation grid comprised 776,628 nodes: 81 in easting, 94 in northing and 102 nodes in the vertical direction. The geometry of the simulation grid is described in Table 4-2.

Table 4-2. Geometry of the simulation grid

Parameter	*X*	*Y*	*Z*
Origin coordinates	52,382.50	9,196.25	1,096.25
Extents, m	202.5	235	255
Node spacing	2.5	2.5	2.5
Number of nodes	81	94	102
Total number of nodes	776,628		

The size of the simulation grid enabled modelling of the quantity of mineralised material sufficient to describe an annual volume of production by an underground mining method.

4.6.1 **Lithology simulation**

Simulation parameters

The main DS parameters for lithology simulation are shown in Table 4-3. No global or local proportions conditioning was considered necessary. The complete parameter file can be found in Appendix F. 100 realisations of lithology were produced in the final ensemble.

Table 4-3. Main parameters of the Direct Sampling algorithm used in the lithology simulation (simulation ID 83) and processing time

Parameter	*Value*
Input parameters	
Conditioning data	Categorical binary lithology variable in the top portion of the deposit and along the pseudo drillholes on 40 mE x 40 mN grid
Consistency of conditioning data	0.05
Maximal number of conditioning nodes	30
Weight factor for conditioning data	2
Distance threshold	0.01
Tolerance for flagging nodes	0
Maximal scan fraction of the ti	0.5
Processing characteristics	
Number of simulated nodes	263,951
Simulation time per realisation, sec	861.45

Simulation process description

Simulation of lithology followed the same route as in the generation of the 'true' model (See Section 4.4.3 'True' model construction). The same elementary training image was used together with the rotational fields for true dip and azimuth of the contact. The conditioning data was provided in the form of a simulation grid with conditioning nodes along the pseudo drillhole traces on a 40 m x 40 m grid. The upper presumably 'mined-out' portion of the grid was included in the conditioning data for simulation in a reconstructive mode.

Validation

Three realisations (simulation ID 83) are shown on sections in Figure 4-39. For reference, the marker lithological contacts are indicated on the sections as grey lines.

The realisations show good honouring of conditioning data and rotational fields, however, might be understating the uncertainty associated with the lithological interface. To introduce more variability in the simulated morphologies, simulation with dip and azimuth fields tolerances can be attempted in future. In such a case, the tolerance can be input as an

exhaustively informing variable with a tolerance of 0 along the drillhole traces and increasing tolerance values away from drillholes.

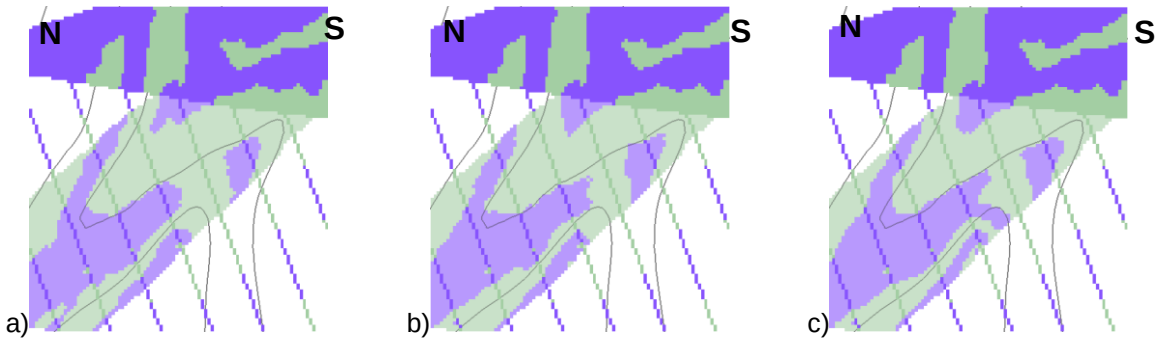
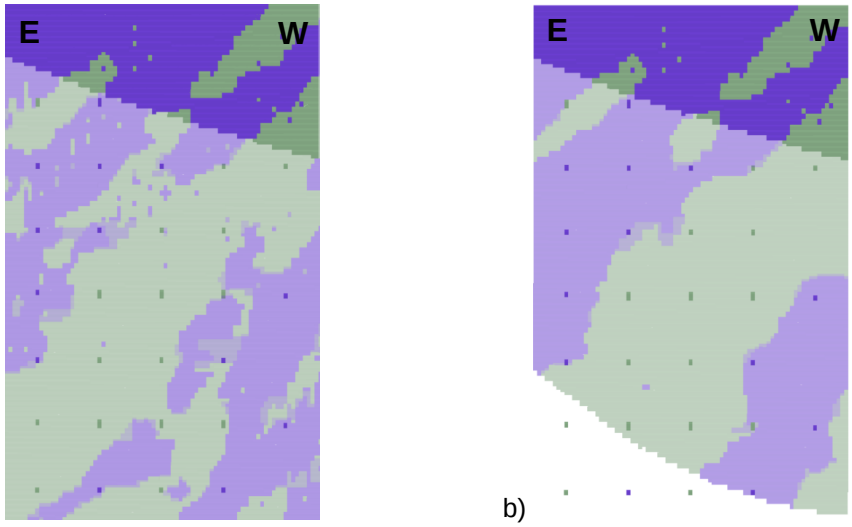


Figure 4-39. Sections showing three final realisations of lithology in semi-transparent colour: BIF (purple) and diorite (green). The solid colour indicates conditioning data. Grey lines represent the overall attitude of the contact

Analysing the patterns of the simulated lithology (Figure 4-40b) vs the 'true' model (Figure 4-40a) along a slice parallel to the shearing revealed that the patterns are too bulky in the simulation, while finer intercalation is observed in the 'true' model. At the same time, the conditioning data is well honoured. This, due to direct correlation between the mineralisation and frequency of the lithological contact per unit volume can have a material effect on the simulated gold grades. Statistics of subsequent gold grades simulation, not shown here, confirmed the finding.

As mentioned above, in future, it can be beneficial to attempt lithology simulation using a continuous variable - true thickness of the units which can be calculated from apparent thickness using structural measurements in drillholes. This would allow guiding the algorithm in selecting the patterns of appropriate thickness from a similar part of the elementary training image.



a) b)

Figure 4-40. Isoclinal view parallel to the direction of the shear zone, showing a slice through lithology in: (a) the 'true' model, and (b) one realisation. The displayed slices are 3m thick. The solid colour indicates pixels of hard data conditioning in the top part of the grid and along the 40 mE x 40 mN pseudo-drillhole grid. The simulation shows bulkier and 'clear-cut' geometries, while the 'true' model demonstrates some noise possibly due to the inferior quality production drilling as well as finer intercalation of units not reflected in the conditioning data

Bar plots with reproduction of categorical proportions are shown in Figure 4-41. The simulated proportions display narrow spread across the 100 realisations. The proportions in the 'true' model in the top portion of the deposit (referred to in this case as 'TI') are also shown. They vary considerably from the rest of the data in the simulation grid and demonstrate a highly non-stationary setting. It is worth noting that the training image spanning over the whole top portion of the deposit has been used for the proportions validation, while a smaller up-plunge portion should have been considered.

Figure 4-41. Bar-plots showing reproduction of the lithological categories' proportions in the simulation (BIF - purple, diorite - green). The whiskers (shown as narrow black lines) in the simulated proportions bars display a very narrow spread of uncertainty for the 90% confidence limits

The table of comparative statistics with lithology proportions can be found in Appendix A.

4.6.2 **Mineralised zone simulation**

Simulation parameters

The most important parameters of the DS used in the final simulation of the mineralised zone are shown in Table 4-4. The complete parameter file can be found in Appendix G. 20 realisations were produced as the final ensemble.

Table 4-4. Main parameters of the Direct Sampling algorithm used in the mineralised zone simulation (simulation ID 99) and processing time

Parameter	*Value*	
Input parameters		
Conditioning data	*Potential field of lithological contact* - exhaustively informed over the whole simulation grid	
	Lithology - exhaustively informed over the whole simulation grid. A unique realisation of lithology per realisation of the mineralised zone	
	Categorical binary variable for the mineralised zone – exhaustively informed in the top portion of the deposit (to ensure reconstructive simulation) and along the 3-nodes diameter cylinders along the pseudo drillholes on 40 mE x 40 mN grid	
Consistency of conditioning data	0.02	
Maximal number of conditioning nodes	Potential field Lithology Mineralised zone	40 40 80
Weight factor for conditioning data	Potential field Lithology Mineralised zone	1 1 1
Simulation path	PATH_RANDOM_HD_DISTANCE_PDF strength of the distance	1 (most guided by the distance)
Distance threshold	Potential field Lithology Mineralised zone	0.05 0.01 0.01
Block data	Provided for the mineralised zone as a linear weighted average of the 40mE x 235mN x 40mRL blocks from the 'true' model	
Maximal scan fraction of the TI	0.25	
Tolerance threshold	0.01	
Processing characteristics		
Number of simulated nodes		251,845
Simulation time per realisation, sec		18,627.20

Simulation process description

Several approaches to modelling the mineralised zone were tested.

In the beginning, a method similar to the simulation of lithology described above was adopted, namely, using an elementary training image with rotational fields. The conditioning data was flagged as mineralised zone where Au>0.5 g/t and as waste otherwise. An example showing the training image and a section through the resulting realisation is shown in Figures 4-42a and 4-42b. The reproduction of proportions was perfect, within one

percent deviation from the conditioning target proportions. The conditioning data, however, was not honoured well in places of fine banding.

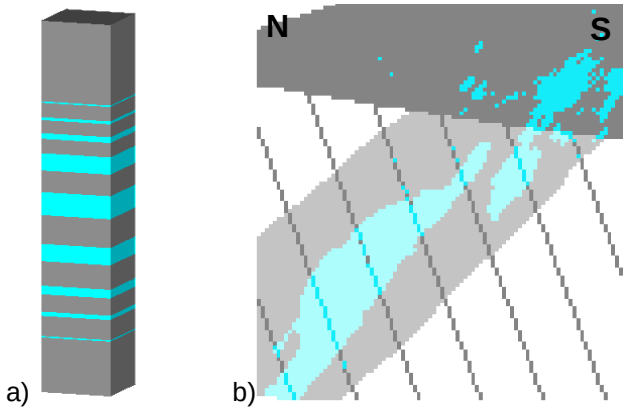
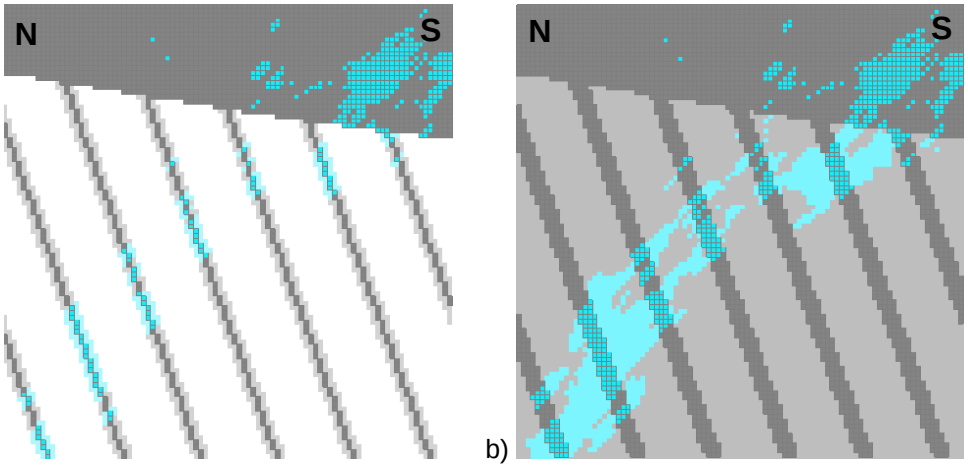


Figure 4-42. Sections through objects in trial simulation of the mineralised shear zone employing an elementary training image as source of patterns: a) the elementary training image; b) one conditional realisation using the elementary training image.
The conditioning nodes are shown as solid pixels, the simulated nodes – semi-transparent. Follow-up gold simulations overstated the contained metal due to bulkiness of mineralised zone morphology

Follow-up gold simulations based on this realisation of the mineralised zone were all overstating the contained metal (the statistics are not shown here). Upon scrutiny, it became clear that it was the bulkiness of the mineralised morphology originating from the tabular shapes of the mineralised zone in the elementary training image. Figure 4-27 shows sections through the ‘true’ model with much more discontinuous and patchy patterns than those in the simulation in Figure 4-42b above. For further simulation of the mineralised zone, it was deemed appropriate to utilise the ‘true’ model as a source of realistic mineralised zone morphology.

To improve local honouring of the conditioning data while providing continuity, nodes within 4 m radius (an equivalent of two grid nodes) around the pseudo-drillhole traces were coded for the same indicators as the nodes along the pseudo drillholes. Local anisotropy of the attitude of shear splays was ensured during the flagging. The resulting conditioning data is shown in Figure 4-43a. Figure 4-43b presents a section through one of the realisations of the mineralised zone over-imposed on the conditioning data.



a) b)

Figure 4-43. Sections through objects in simulation of the mineralised shear zone using the 'true' model as source of patterns: a) conditioning data, and b) one conditional realisation. The conditioning nodes are shown as solid pixels, the simulated nodes – semi-transparent

In the final round, 100 realisations of the mineralised zone were generated. Each simulation run contained one realisation. The conditioning data for each simulation consisted of three variables: 1) potential field of lithological contact exhaustively informed over the whole simulation grid, 2) previously simulated lithology, one unique realisation per each simulation of the mineralised zone, exhaustively informed over the whole simulation grid, 3) the mineralised zone, with informing nodes defined within 4 m radius from the pseudo drillholes as well as all the nodes in the upper 'mined-out' portion of the simulation grid (Figure 4-43a).

The potential field of the attitude of the lithological contact (Figure 4-44) has been provided as an auxiliary variable due to dependence of the intensity of mineralisation on a local scale from the folding architecture. Tolerance of 0.05 of the potential field was used during the simulation. This translates to the width of about 10-15 m along GHSZ. More testwork can be undertaken in future on the sensitivity of this value.

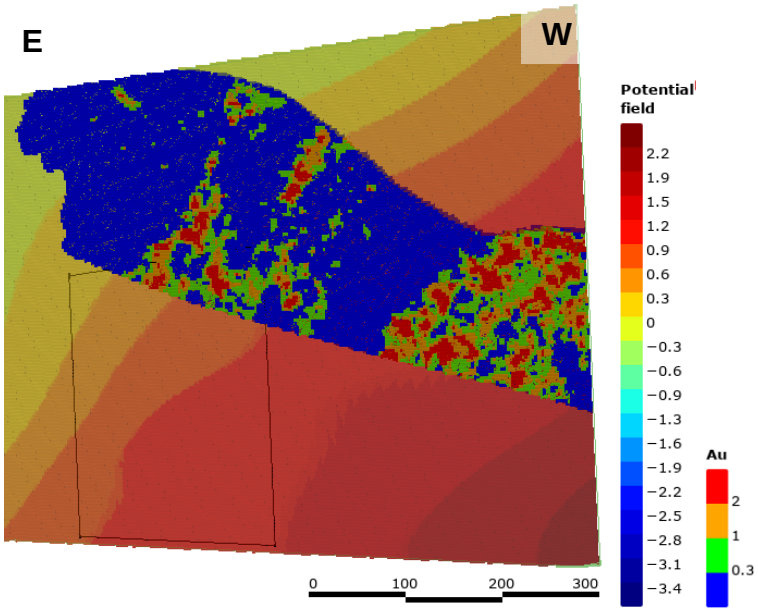


Figure 4-44. Section through the 'true' model along the shear zone showing the exhaustively informed variable of the potential field of lithological contact and gold in the 'known' upper portion of the 'true' model. The grey outline represents the simulation volume. The elongation of the ore shoots along the lithological contact potential field is apparent

Lithology was provided as another exhaustive variable to ensure the multi-variate dependencies of mineralisation morphologies with this structural control are honoured. The previously simulated lithology was used, one realisation of lithology per one realisation of the mineralised zone.

To improve better local reproduction of the mineralised content, block data conditioning was attempted. Two block geometries were tested: 40 mE x 40 mN x 40mRL and 40 mE x 235 mN x 40 mRL.

Figure 4-45 shows the extents of the simulation grid, the pseudo drillholes and one of the conditioning blocks of 40 mE x 235 mN x 40 mRL.

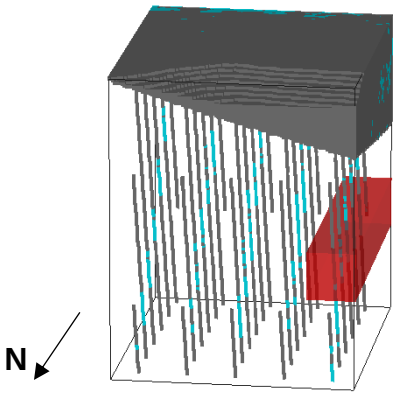


Figure 4-45. Isoclinal view showing the extents of the simulation grid (grey box), conditioning pseudo drillholes and one of the conditioning blocks (red) of 40 mE x 235 mN x 40 mRL size used in mineralised zone simulation with block conditioning

For both geometries, the conditioning block values were calculated as linear averages of the mineralised material proportion from the ‘true’ model nodes. While the post-simulation validation of global mean statistics of the proportions was good for both shapes, the cubic conditioning block geometry produced a suboptimal spatial connectivity pattern and artefact geometries of the mineralised zone, as well as a wider spread of block-on-block scatter. Because the shear zone is a tabular structure it was deemed more appropriate to utilise cuboidal blocks intersecting the total thickness of the shear zone. The elongated conditioning block geometry was chosen over the cubic shape to avoid introducing an overfitting problem when the algorithm tries to honour smaller conditioning cubic blocks.

Ideally, the block geometry should be rotated so that the elongated axes of the cuboidal blocks are orthogonal to the overall attitude of the shear zone. Blocks of cylindrical shapes around the conditioning drillholes for better spatial pattern reproduction can be attempted. It is beyond the scope of the current research.

Validation

Visual validation is used as a primary validation step in this research. A trained eye performs complex image matching process faster and often better than any algorithms developed so far.

Sectional views through three realisations of the final simulation of the mineralised zone are shown in Figure 4-46. It is apparent that the simulation zone (Figure 4-24) as used in the current research defines too broad a domain around the fold hinge zone, and the mineralised zone has 'spilt' outside of this stationary domain (Figure 4-46b). Another conditioning variable is required in future, to accurately delineate the zone between two axial surfaces of the S-fold. It can be created by utilising derivatives of the potential field representing the attitude of the bedding.

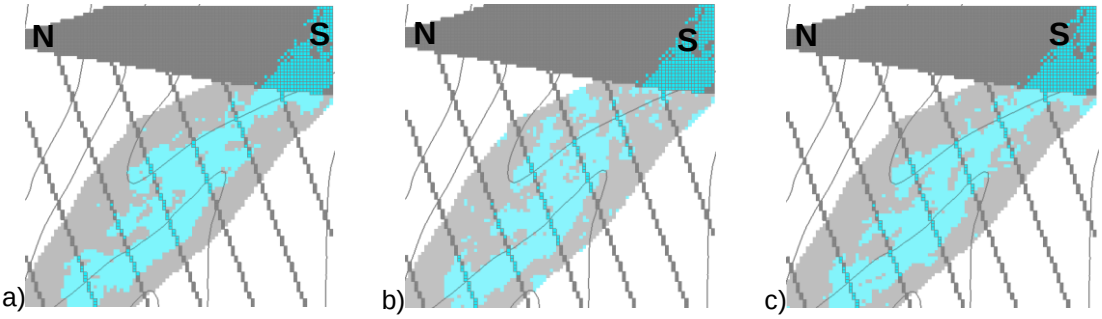


Figure 4-46. Sections through three realisations of final simulation of the mineralised zone using the 'true' model as source of patterns and block data conditioning. Cyan – mineralised shear zone, grey – waste. The solid colour indicates pseudo drillholes. The simulated nodes are shown in semi-transparent colour. Grey lines show the overall attitude of the contact. The simulated mineralised zone in the top portion of (b) spills outside the fold hinge zone due to insufficiently accurate definition of the latter

A section along the shear zone is shown in Figure 4-47b. It displays reasonably good reproduction of patterns observed in the training image (Figure 4-47a), albeit not as strong zonation along the orientation of lithological contact in the bottom right corner.

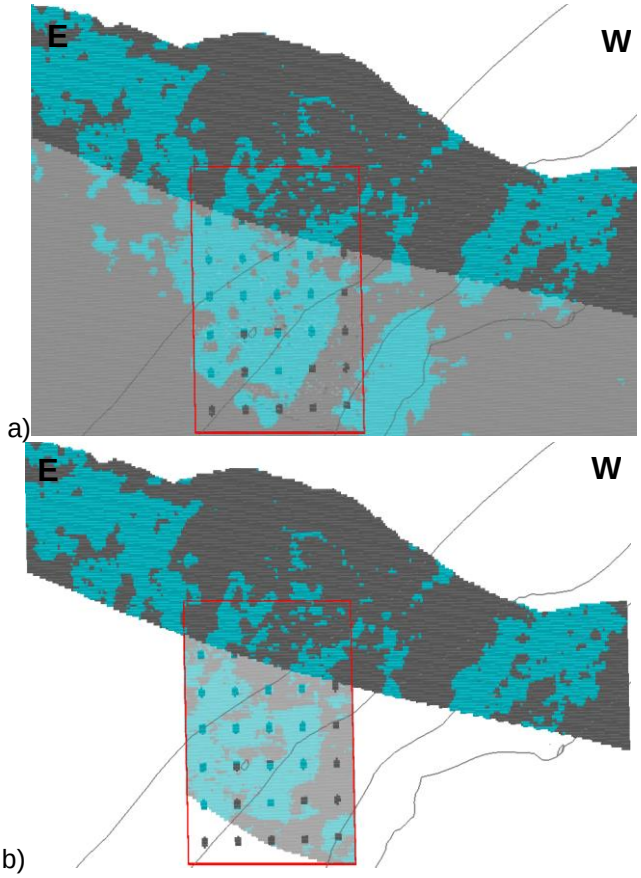


Figure 4-47. Sections along the GHSZ, shown within the red rectangle: a) the 'true' model of the mineralised zone, b) one of the final realisations of the mineralised zone. In both images, solid nodes in the top portion represent the training image, in the bottom – conditioning nodes, semi-transparent nodes in (a) are 'true' model, in (b) – the simulation. Grey lines show the overall attitude of the lithology contact

Bar charts for indicator proportions are presented in Figure 4-48. The reproduction of proportions within the unmasked portion of the simulation grid is good, contradictory to the fact that the training image in the upper portion of the deposit has a lower proportion of the mineralised material: 26.6% in the training image vs 34.4% in the 'true' model over the simulated volume. The mean values of the mineralised material in the conditioning data, the 'true' model and the realisations are very close: 35.4%, 34.4% and 33.3% correspondingly, indicating 1.1% negative bias in the simulation vs the 'true' model. The whiskers on the bar-plot represent the 90% confidence interval. There is a very low spread of uncertainty among the 20 realisations that demonstrated a good performance of the algorithm. Appendix B contains tabulated statistics for the simulated indicator proportions.

Figure 4-48. Bar-plots showing mineralised material proportions in the training image, conditioning data, 'true' model and simulation.
The whiskers (shown as narrow black lines) display a very narrow spread of uncertainty within a 90% confidence interval in simulation. The simulated proportions show about 2% difference from the conditioning data proportions and 1% from the 'true' model, even though the TI serving as source of patterns has a much lower proportion of the mineralised material, 25.9%

Block on block scatterplots for the mineralised proportions in conditioning blocks for four selected realisations are given in Figure 4-49. The marginal distributions of the proportions in the conditioning block dataset derived from the 'true' model upscaling and in realisations are displayed along the X and Y axes correspondingly. The scatter is narrow, mostly within ±10% of the 45^{o} gradient line, indicative of good regression.

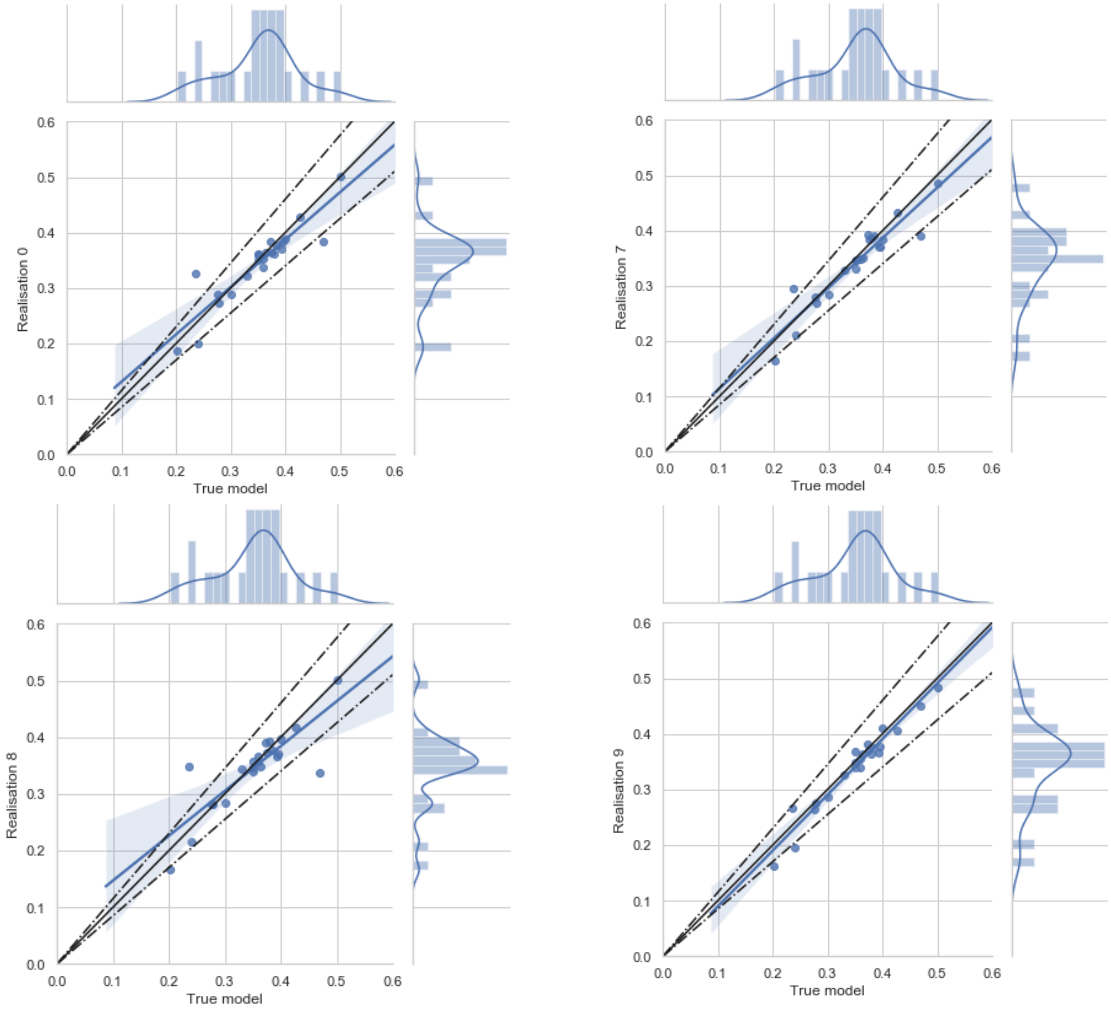


Figure 4-49. Scatterplots of block proportions of the mineralised zone in four simulations vs the 'true' model.
Note, the axes on the graphs are swapped as compared to a conventional representation where the simulated realisation is shown on the X-axis and the 'true' model on the Y-axis

4.6.3 **Gold grade simulation**

Simulation parameters

The most important parameters of the DS used in the final simulation of gold grades are shown in Table 4-5. The complete parameter file can be found in Appendix D. 20 final realisations of gold were produced.

Table 4-5. Main parameters of the Direct Sampling algorithm used in the gold grade simulation (simulation ID 109) and processing time

Parameter	*Value*	
Input parameters		
Conditioning data	*Potential field of lithological contact* - exhaustively informed over the whole simulation grid	
	Lithology - exhaustively informed over the whole simulation grid. A unique realisation of lithology per simulation of gold	
	Categorical binary variable for the mineralised zone – exhaustively informed over the whole simulation grid. A unique realisation of mineralised zone per simulation of gold	
	Categorical variable of gold with categories determined by 4 cut-off values: 0.3 g/t, 0.5 g/t, 1 g/t and 2 g/t. The variable is exhaustively informed in the top portion of the deposit (to ensure reconstructive simulation) and along the pseudo drillholes on 40 mE x 40 mN grid	
	Gold representing the grades - The variable is exhaustively informed in the top portion of the deposit (to ensure reconstructive simulation) and along the pseudo drillholes on 40 mE x 40 mN grid	
Consistency of conditioning data	0.02	
Maximal number of conditioning nodes	Potential field Lithology Mineralised zone Au categorical Au continuous	40 100 100 100 100
Weight factor for conditioning data	Potential field Lithology Mineralised zone Au categorical Au continuous	1 1 1 2 2
Simulation path	PATH_RANDOM_HD_DISTANCE_PDF strength of the distance	1 (most guided by the distance)
Distance threshold	Potential field Lithology Mineralised zone Au categorical Au continuous	0.1 0.01 0.01 0.01 0.01
Block data	None	
Maximal scan fraction of the ti	0.25	
Tolerance threshold	0.01	
Processing characteristics		
Number of simulated nodes		263,951
Processing time per realisation, sec		40,440.20

Simulation process description

An acceptable simulation of gold grade values was not achieved immediately. Below is a summary of the problems encountered and lessons learnt.

In the beginning, difficulties arose in reproducing continuity of the gold mineralisation. Figure 4-50 demonstrates the results of one of the attempts when the orientation of the lithological contact produced a considerable improvement in high-grade continuity. The unilateral path used in this simulation is attributed to be the main reason for improvement. Another contributing factor was the fact that the TI and SG in this simulation run were defined over the same volume, which resulted in the extraction of identical neighbourhoods.

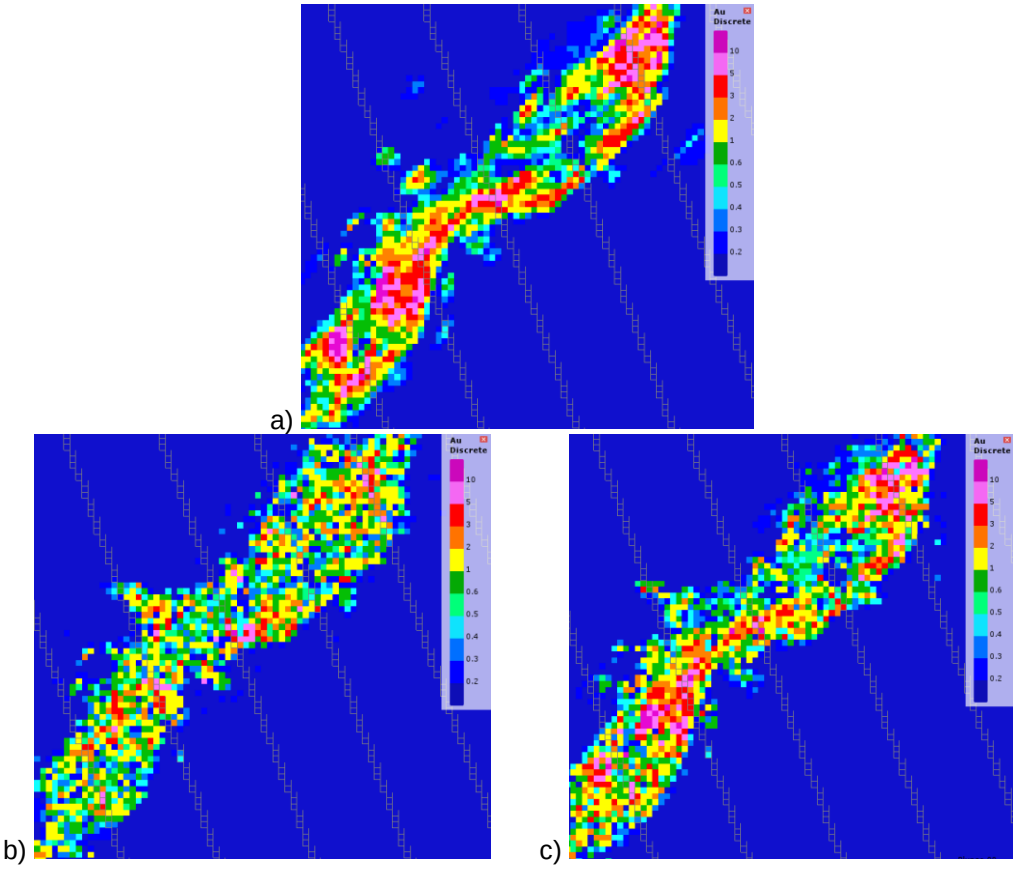


Figure 4-50. Sections through the training image (a), and two simulations of gold: without (b) and with (c) the rotational field of lithological contact attitude as exhaustive variable conditioning. Introduction of the conditioning dip and azimuth ensured significant improvement in gold continuity (c) in comparison to the simulation without it (b)

Further attempts did not produce significant improvement in the framework. These attempts were:

- Using the distance to shear as an auxiliary exhaustively informed variable. Although it did not show a visible improvement, it should not be discarded, specifically if it can be incorporated in the data collection workflow, when individual major brittle discontinuities can be explicitly logged and correlated in drillholes. In this case, another simulation step may need to be introduced post-lithology simulation – simulation of the “braided” shear splays;
- Fixing the radii of the neighbourhood window to 10 x 10 x 10 nodes or even smaller, to 3 x 3 x 3 nodes;
- Hierarchical simulation of gold grade per lithology, with gold grade in BIF simulated first and used further as a conditioning dataset when simulating gold in diorite;
- Setting up the masked area to include only cells simulated previously as mineralised zone. Failure to achieve good pattern reproduction, in this case, was attributed to the fact that minute variations of neighbourhoods ‘cookie-cut’ by the intricate mineralised zone shape could not be easily found in the training image;
- Introducing conditioning variables defining difference in angle between the attitude of shears and lithological contact. The reasoning for using this approach was based on the premise that small angle values (e.g. shear splays running subparallel to the lithological contact) are associated with higher Au grades, although some high-grade mineralisation does occur within brittle structures of dip steeper than that of the lithological contact attitude. The two angles, the difference in dip and difference in dip direction, were combined into one using trigonometric calculations (Cowan, 2018). While this test did not show any improvement, it is believed it has potential and can be tested further, tolerance of the angle variable might need to be increased during the simulation.

Nothing seemed to improve the result until a multiple-indicator variable of gold grade was introduced in a co-simulation model, in the sequence preceding the gold grade variable (e.g. two variables simulated in the same run: the multiple-indicator of gold grade, and gold grade).

Changing the cut-off for the mineralised zone from 0.5 g/t to 0.3 g/t (as explained in Section 4.4.3) also had a marked improvement on continuity and allowed to have an intermediate validation stage in the workflow - confirming proportions of the total mineralised material prior to committing to gold simulation.

The next step contributing to an improvement in simulated gold grade patterns was conditioning gold grade simulations to the mineralised zone, the latter derived using the 'true' model as the source of training image patterns rather than an elementary training image. Figure 4-51 demonstrates the problem: using an elementary training image resulted in the tabular and bulky mineralised zone for conditioning gold grade simulations.

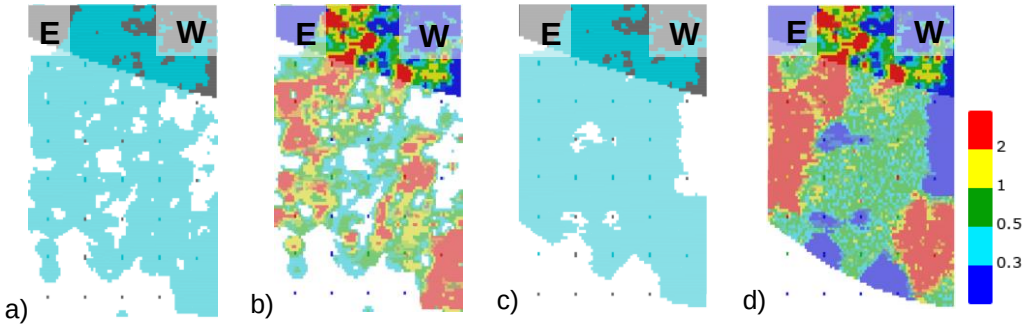
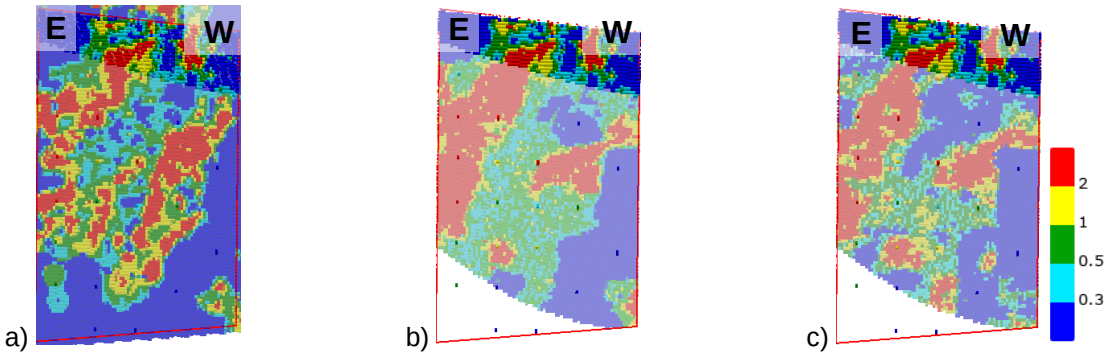


Figure 4-51. Sections parallel to the shear zone showing overstated continuity of simulated mineralisation when using an elementary training image for shear zone simulation.
a) 'true' model mineralised zone, b) 'true' model gold grade, c) mineralised zone simulation achieved using an elementary training image, d) gold simulation using the mineralised zone in (c) as exhaustively informed conditioning grid. Solid cell colour – conditioning data, semi-transparent cells – 'true' model cells in (a) and (c) and simulated nodes in (c) and (d)

Figures 4-51a and 4-51b display the 'true' model of mineralised zone and gold grade values respectively. Figure 4-51c shows a section through a realisation with tabular and bulky mineralised zone generated using an

elementary training image. Figure 4-51d displays a section through the realisation of gold grade utilising the tabular mineralised zone. Morphology of the high grade is oversized, while low- and medium-grade zones (below 1 g/t) do not have sufficient connectivity and look ‘pixelated’.

Figure 4-52 shows the progress made subsequently by using the ‘true’ model as source of patterns when simulating mineralised zone. The ‘true’ model morphology of gold grade patterns is displayed in Figure 4-52a, the simulated morphologies are: the realisation in Figure 4-52b – with tabular mineralised zone simulated from an elementary training image, vs the realisation in Figure 4-52c – improved albeit still not ideal patterns due to using the ‘true’ model as a training image for mineralised zone simulation.



a) b) c)

Figure 4-52. Sections parallel to the shear zone showing improved morphology of simulated gold when using the ‘true’ model as source of patterns for shear zone simulation.
a) ‘true’ model over the simulation grid, b) simulation achieved with tabular mineralised zone, the latter constructed using an elementary training image, c) superseding simulation achieved using the ‘true’ model training image as source of patterns for the mineralised zone.
Solid cell colour – conditioning data, semi-transparent cells – simulated nodes

Other factors contributing to the improved quality of gold simulation are:

- Simulating controlling variables in the sequence of geological precedence - lithology and the mineralised shear before gold grade;
- Using the potential field of the lithological contact as an auxiliary variable – to help localise the neighbourhood in the training image in the stationarity-similar up-plunge area of the deposit. The motivation for it is apparent from analysis of the effect lithological contact

attitude has on the stationarity of gold grade (Figure 4-30). For this purpose, the size of the training image should be larger than the simulation grid and include the area with the potential field similar to that encountered over the simulation grid. To economise on the processing time, one can use an extract of the training image utilising only the portion with the range of the potential field values encountered in the SG. Tolerance of 0.1 was chosen to ensure search over a sufficiently wide zone of the training image while providing the width equivalent to about one variogram range across the strike of the orebody;

- While it was not necessary to use a large amount of conditioning nodes, reducing this number too much resulted in deteriorated quality. This is attributed to the fact that the gold grade is closely related to the contact morphology and a larger number of nodes helps to extract an appropriate neighbourhood. In several simulations, it was discovered that using the same number of nodes for lithology, mineralised zone, and gold would improve the quality of validated statistics significantly. The 100 conditioning nodes specified in the input to the final simulation runs is still considered insufficient for a 3D environment and can be increase in future;
- Morphology of the simulated mineralised zone should have the same size of patterns as gold grade to ensure similar statistics. It is better to be consistent: if the 'true' model is to be used as a training image for the gold simulation it should be used for the mineralised zone simulation as well.

As a final test to improve the morphology of the simulated gold mineralisation, another approach to defining thresholds was attempted. Instead of having four indicators (0.3 g/t, 0.5 g/t, 1.0 g/t and 2.0 g/t), a multiple-indicator variable with three thresholds was used (0.3 g/t, 0.8 g/t and 2.0 g/t). It was presumed that a lesser number of cut-offs would result in larger geobodies and more robust neighbourhood search. The lower and upper thresholds were fixed (0.3 g/t as the indicator for the definition of the

mineralised zone, 2.0 g/t as a value chosen for the definition of the underground cut-off value), leaving flexibility to decide on the middle threshold. Ensuring sufficiently robust clusters of low and medium grade indicators was a guide in the decision.

Note: This version (shown in Figures 4-53 b, c, d and 4-54 b, c, d) was produced after the final uncertainty estimation and stope design were completed and was not used in any further statistical analysis.

A statistically more rigorous approach to the decision on the number of cut-offs and the specific cut-off values can be investigated in future.

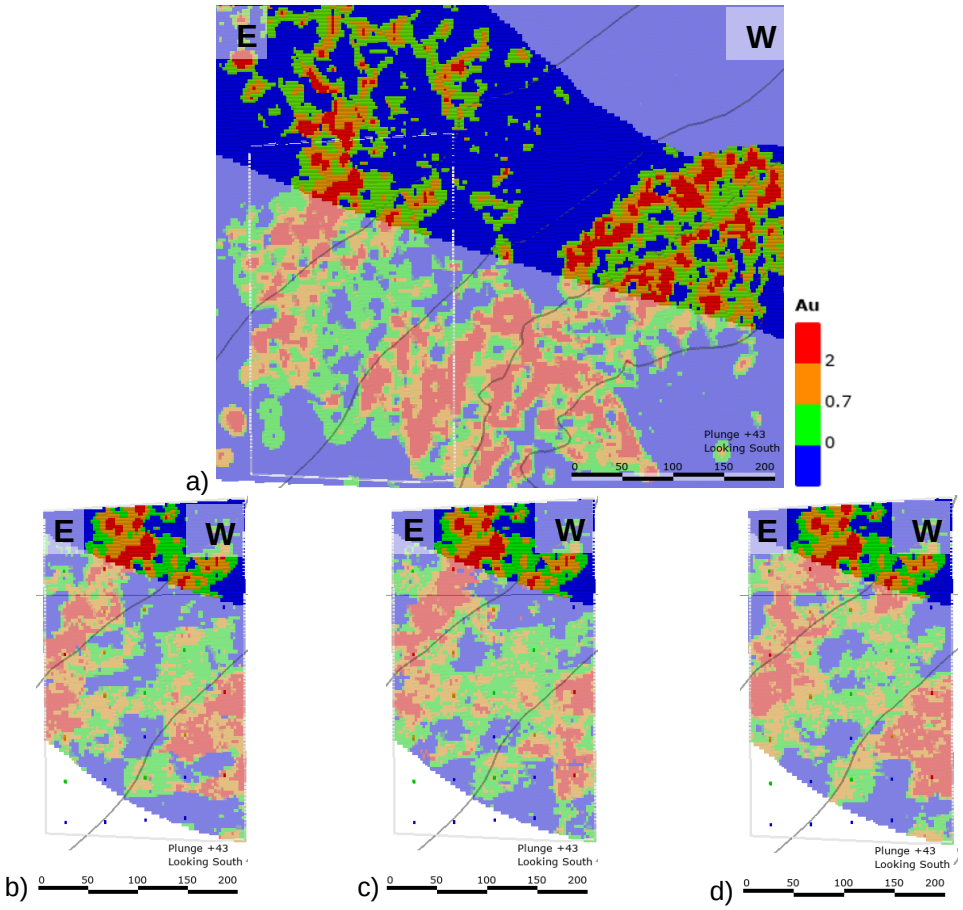


Figure 4-53. Sections parallel to the shear zone showing improved morphology of gold patterns when using the three-indicator gold grade variable in co-simulation mode with gold grade: a) the 'true' model over the simulation grid, (b-d) three realisations achieved with indicator Au variable with three thresholds (0.3 g/t, 0.8 g/t and 2 g/t). Solid cell colour in (a) – training image, in (b-d) - conditioning data. Semi-transparent cells in (a) – 'true' model, in (b-d) - simulated nodes

Figures 4-53 and 4-54 display long- and cross-sections through the ‘true’ model of gold grade and three realisations. Overall, the high-grade shoots honour intersection of the main structural controls, bedding and shear zones, and display good reproduction of high-grade continuity. While the patterns in the realisations approximate the ‘true’ model, there is still considerable noise in the simulated patterns. Increasing the maximum number of conditioning nodes allowed can be tested in future.

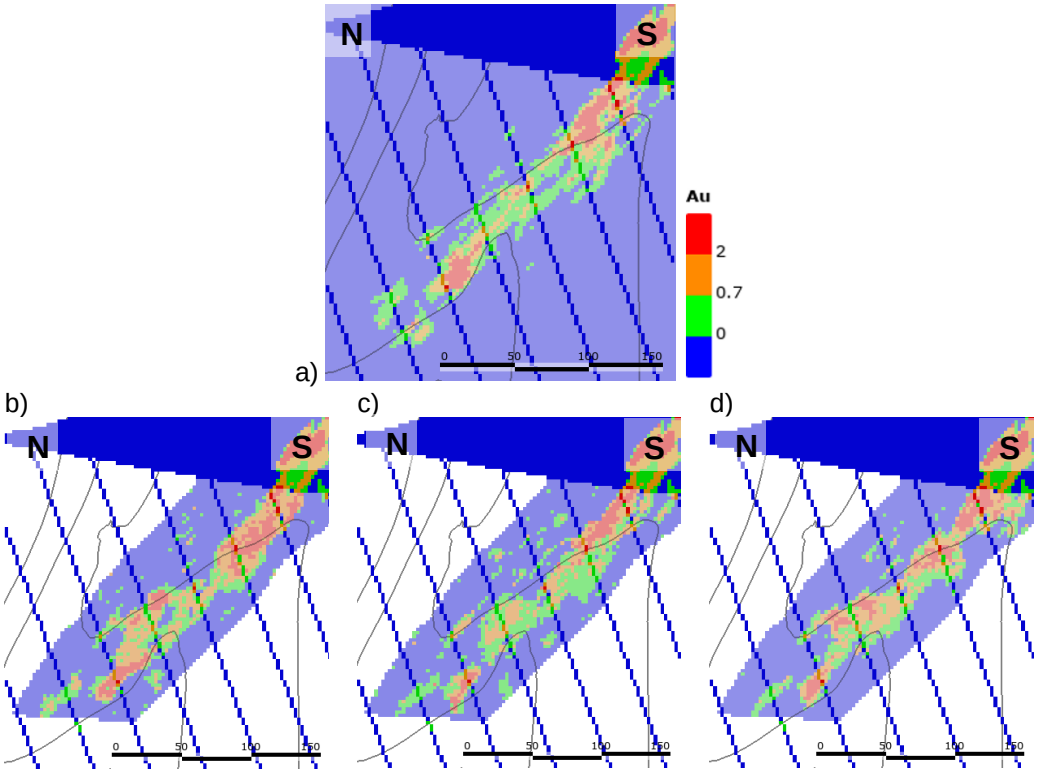


Figure 4-54. Cross-sections showing improved morphology of the simulated gold patterns when using the three-indicator gold variable in a co-simulation mode with gold: a) the ‘true’ model over the simulation grid, (b-d) three realisations achieved with indicator Au variable with three thresholds (0.3 g/t, 0.8 g/t and 2 g/t). Solid cell colour in (a) – training image, in (b-d) - conditioning data. Semi-transparent cells in (a) – ‘true’ model, in (b-d) - simulated nodes. Grey lines show the overall attitude of lithological contact

Validation

The validation process of the gold grade simulations comprised of visual analysis in 3D, global statistics validation, scatterplots, boundary analysis,

variogram reproduction and grade-tonnage curves. The latter is covered in the following section 4.6.4 Global Uncertainty Assessment.

Twenty realisations of gold grade were produced for the final ensemble. Sections through three of these grade realisations are shown in Figure 4-55. Note: these do not represent the best achieved simulation as mentioned previously.

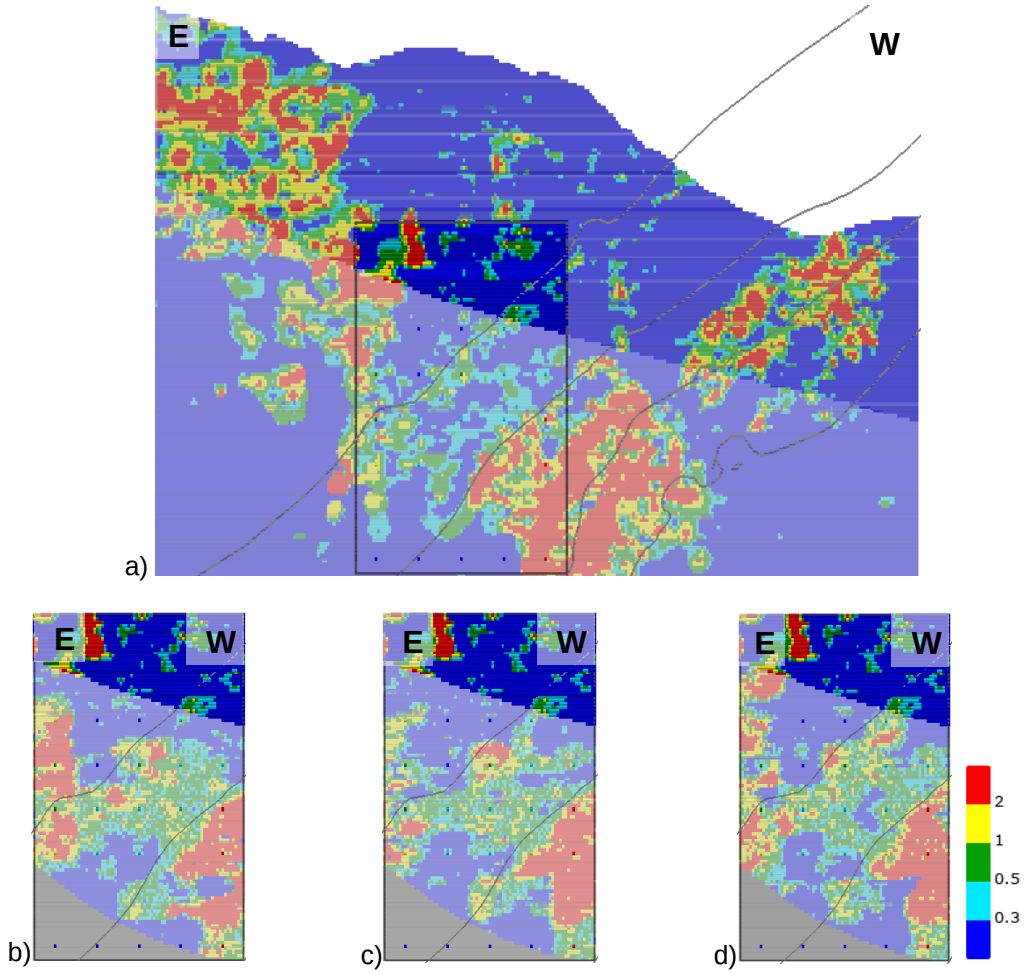


Figure 4-55. Sections parallel to the shear zone showing final realisations of gold grade: a) 'true' model, b-d) three realisations. The top portion of the 'true' model served as a TI for the gold grade simulation. Solid pixels in (b-d) display conditioning nodes, grey lines represent the overall attitude of the lithological contact. The realisations show coarser morphology of patterns than that of the 'true' model

It is apparent that the simulated gold grade patterns display larger morphology than those in the training image and the 'true' model over the simulation volume. Also, the lower range of values from 0.3 g/t to 1 g/t is too

scattered and ‘pixelated’. This geological contact morphology is an important factor to ensure correct prediction of the recoverable Mineral Resource, specifically at the grade range close to the mining cut-off as it will affect the selectivity between ore and waste.

Statistics validation. Statistical validation included a comparison of the main statistics in the simulated gold grade realisation vs the training image, conditioning data and ‘true’ model. The tabulation for the corresponding statistics is given in: Appendix C – for BIF within the mineralised zone, Appendix D – for diorite within the mineralised zone, Appendix E – for combined BIF and diorite within the mineralised zone.

Histograms of simulated mean values for the 20 realisations of gold are shown in Figure 4-56: a) mineralised BIF, b) mineralised diorite, and c) combined lithologies within the mineralised shear zone. The vertical solid lines on the graph represent mean values of the reference data, namely: training image mean (red), ‘true’ model mean over the simulation volume (green), conditioning data mean (blue), and the mean value in the simulated gold grade realisations (black). The blue dashed lines represent ±15% error around the conditioning data mean value, the black dashed lines – the 90% confidence interval around the simulated mean. This and subsequent comparisons in the Section 4.6.3 are drawn at point support for the volume representative of the annual production by an underground mining method. Mining selectivity takes place at the stope support and while being of no consequence for the Mineral Resource classification, it is deemed important and informative for validation purposes.

For BIF (Figure 4-56a), the simulated mean of all the realisations is slightly lower than the conditioning data mean (1.97 g/t vs 2.06 g/t). It is very close to the ‘true’ model mean. The 90% of simulated realisations mean values plot within the ±15% error from the conditioning data mean.

For diorite (Figure 4-56b), the simulated mean of the realisations is lower than the conditioning data mean (0.80 g/t vs 1.04 g/t). The 90% of simulated

realisations mean values plot outside the ±15% error from the conditioning data mean.

For the combined mineralised material (Figure 4-56c), the mean of realisations is 1.40 g/t while the mean of conditioning data is 1.51 g/t. The spread of the simulated mean values is quite narrow, and negative bias is apparent across the ensemble of realisations. The mean values of 19 realisations out of 20 (95%) plot below the conditioning data mean. Nonetheless, 90% of the simulated realisations mean values are within the ±15% error from the conditioning data mean.

The negative bias demonstrates a well-recorded phenomenon of MPS – gravitation of the simulation statistics towards the training image statistics.

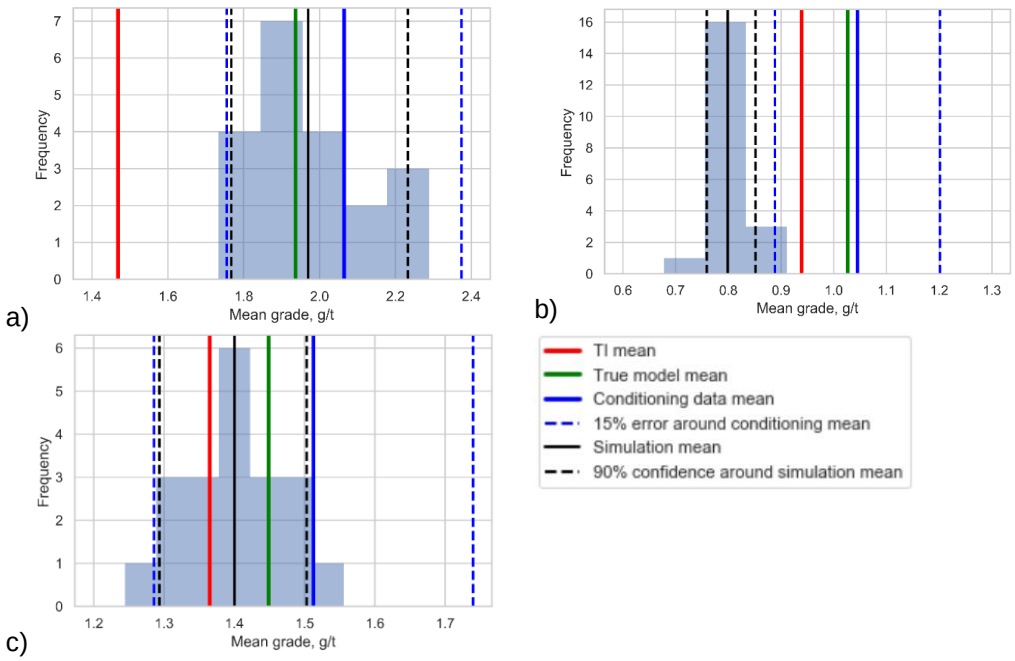


Figure 4-56. Histograms of the simulated mean values of gold grade within the mineralised zone with 90% confidence interval and ±15% error threshold at point support: a) in BIF, b) in diorite, c) in combined lithologies. The black lines show the mean of all realisations (solid line) with a 90% confidence interval (dashed lines). The blue lines show the conditioning data mean (solid line) with 15% error around it (dashed lines). The histogram bins spill marginally outside the desired interval at the lower confidence margin of the conditioning data. This is explained by the fact that the training image shows lower mean gold grade value than the conditioning data

The comparison of the simulated mean values to those in the training image, however, is not direct, because the contact potential field was used with a restricted neighbourhood. In effect, only the western portion of the training image was used during retrieval of suitable patterns during simulation. The statistics should be adjusted to extract only the respective part of the training image within the range of potential fields values as present over the SG.

Reproduction of variance is considered an important factor for traditional variogram-based estimation methods and subsequent change-of-support models for the correct prediction of the recoverable Mineral Resource. Figure 4-57 shows spread of grade variance among the datasets. The variance statistics for the ensemble of realisations (mean variance, minimum and maximum variance, 5th and 95th percentiles) were established by calculating the variance of simulated gold values for each realisation, and then deriving the appropriate statistical measure of the variance for the ensemble. The variance is under-stated in the simulations, while the maximum values are higher than in the conditioning data, approaching those in the training image. The simulations variance is plotting outside the desired interval at the lower confidence margin of the conditioning data. The statistics of the simulated gold grade realisations vs the input data can be found in Appendices C, D and E.

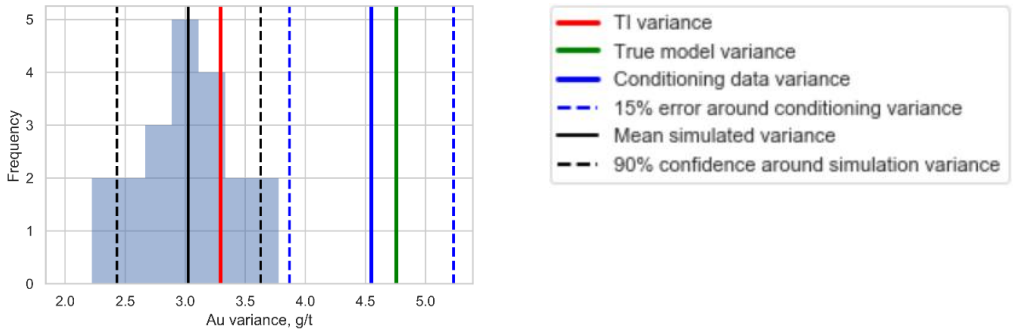


Figure 4-57. Histogram showing spread of Au variance in simulated realisations. The black lines represent the variance mean of all realisations (solid line) with a 90% confidence interval (dashed lines). The blue lines show the conditioning data variance (solid line) with 15% error around it (dashed lines)

Under-representativity of variance vs the real phenomenon is a known issue of MPS which tries to achieve a trade-off between the reproduction of patterns and reproduction of variability. Some developments are taking place in this area, for instance, an optimisation-based approach for MPS simulation has been suggested (Abdollahifard, Mariethoz & Mohammadi, 2019), which is based on an idea of variance maximisation based on optimisation of an energy function. Besides satisfying the conditioning constraints of an MPS simulation and minimising similarity distance, another condition imposed in simulations is to increase the local variance to limit the reproduction of similar patterns at the same location across realisations.

Block scatter. No block-on-block conditioning was provided when simulating gold grade values, nonetheless, it was considered of significance to check scatterplots for the 40 mE x 235 mN x 40 mRL blocks which were used for conditioning the simulation of the mineralised shear zone. Block-on-block scatterplots of four realisations vs the conditioning data are presented in Figure 4-58. The marginal distributions of the proportions in the conditioning block dataset and in realisations are given along the X and Y axes. It should be noted, the axes on the graph are unintentionally swapped as compared to a conventional representation where the simulated realisation is shown on the X-axis and the 'true' model on the Y-axis.

The scatters for the four realisations plot in ellipse along a 45^{o} line, albeit with some scatter outside ±10% of the line, showing quite broad but acceptable regression.

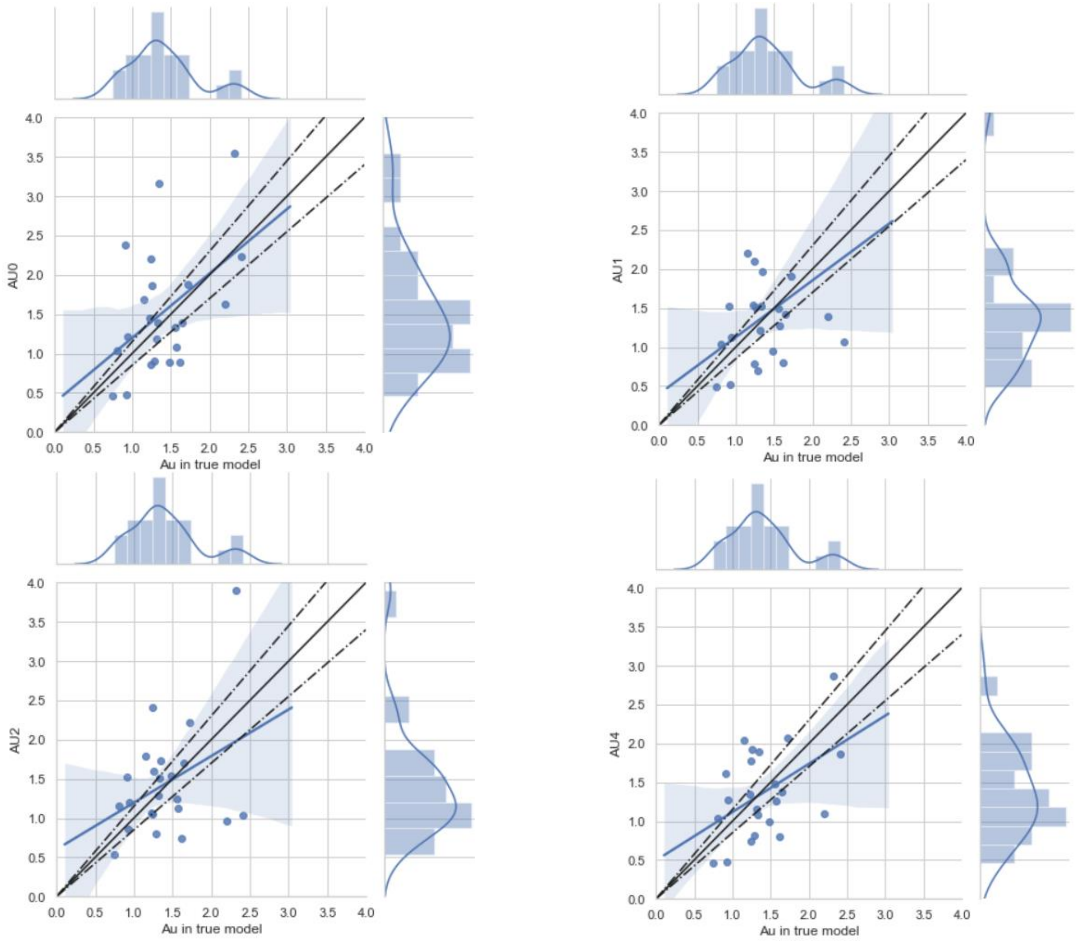


Figure 4-58. Scatterplots of average gold values for 40 mE x 235 mN x 40 mRL blocks: simulated gold in four realisations vs the gold in the 'true' model.
Note, the axes on the graphs are swapped as compared to a conventional representation where the simulated realisation is shown on the X-axis and the 'true' model on the Y-axis

For the author's interest, another scatter plot was generated for the 40 mE x 235 mN x 40 mRL blocks, comparing the scatter of the conditioning pseudo drillholes averages within the blocks vs the 'true' model averages for the same blocks. It is shown in Figure 4-59. The conditioning data display an elliptical scatter vs the 'true' model means with a considerable amount of points plotting outside ±10% of the target 45^{o} line. It indicates that the scatter validation as such is subject to considerable uncertainty even when comparing the average of the 'true' model vs the averages of the sparse data sampled from the same model. In light of this finding, the scatterplots in Figure 4-58 can be considered satisfactory.

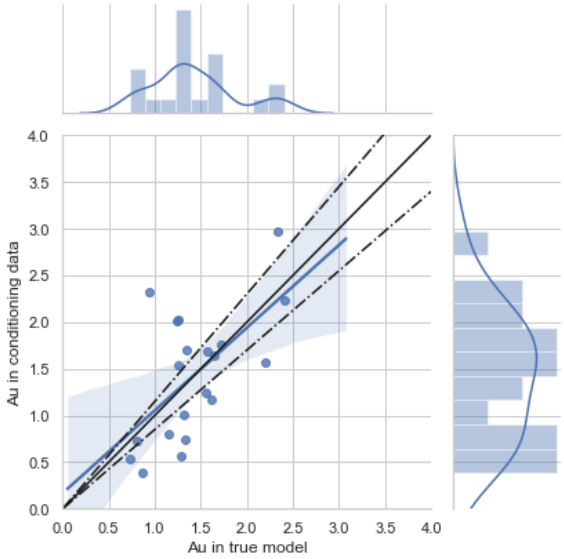


Figure 4-59. Scatterplot of average gold values for 40 mE x 235 mN x 40 mRL blocks: gold in the conditioning block data vs gold in the 'true' model

Boundary analysis. The boundary analysis algorithm in Isatis© was utilised for variogram values calculation, the results exported and plotted using Python scripts in Jupiter notebooks. The algorithm in Isatis© requires the data to be supplied in the form of a desurveyed drillhole data file. For this, a pseudo-drillhole dataset consisting of a subset of nodes from each realisation was created using 20 mE x 20 mN spacing. The downhole 2.66 m intervals (the result of 2.5 m vertical composites along drillholes dipping at 60^{o}) were evaluated for the proximity of the grid nodes: for gold grade simulations, 'true' model grids, training image and the conditioning dataset nodes with gold values extracted appropriately.

The graphs in Figures 4-60 and 4-61 show comparative behaviour of gold grades in the mineralisation across four different boundaries: between mineralised BIF and diorite (Figure 4-60a), between waste BIF and diorite (Figure 4-60b), between mineralised BIF and waste BIF (Figure 4-61a), between mineralised diorite and waste diorite (Figure 4-61b).

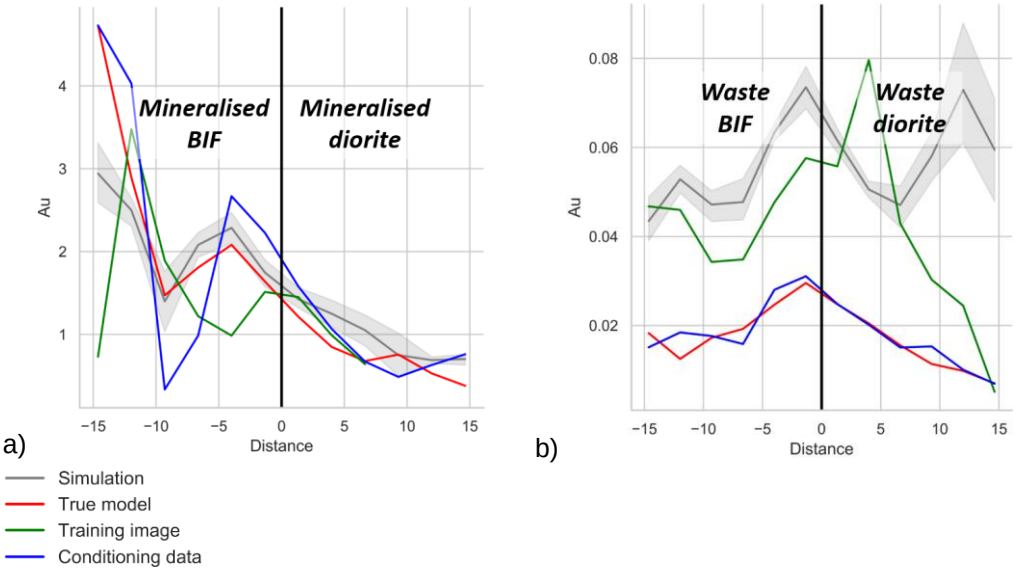


Figure 4-60. Boundary analysis validation for gold simulations across the lithological contact: a) within the mineralised zone, b) outside the mineralised zone

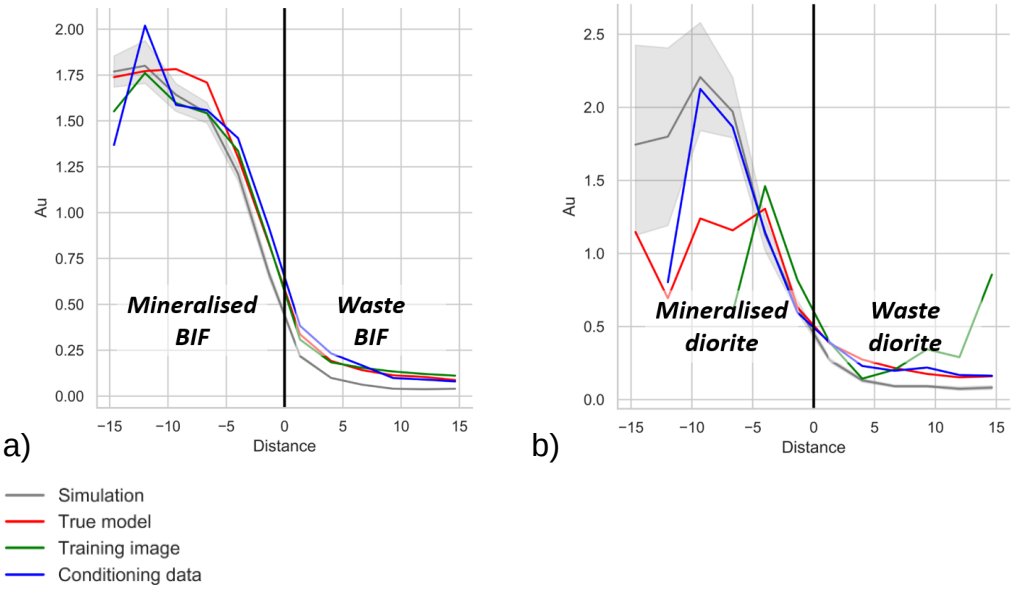


Figure 4-61. Boundary analysis validation for gold simulations across the mineralisation contact: a) in BIF, b) in diorite

All the interfaces exhibit good reproduction of the boundary behaviour in the simulation vs the input data and the 'true' model, except the interface between waste BIF and waste diorite (Figure 4-60b) where the simulation curves approach the TI curve, however, the mean grade difference is immaterial: 0.06 g/t vs 0.02 g/t.

Comparing the pseudo conditioning data curves above (Figures 4-60 and 4-61) to the curves established on the composited source drillholes during EDA (Figures 4-8 and 4-9) shows similar behaviour for all the interfaces. The comparison is not direct, however, and the mean grade values on either side of a specific boundary differ since different spatial extents were covered by the two datasets.

Variograms. Variogram validation was performed in the primary direction of anisotropy, GSLib angles dip/dip direction/plunge: 125/0/10 (Deutsch and Journel, 1998).

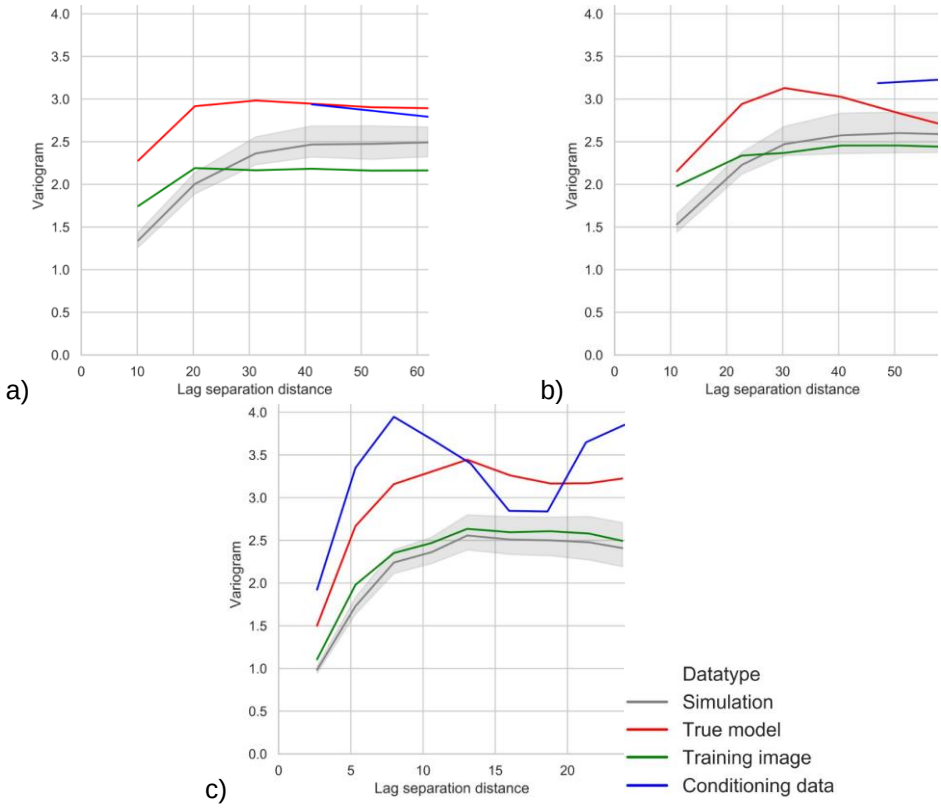


Figure 4-62. Variogram validation for gold simulations.
a) two lithologies combined in the mineralised zone,
b) BIF in the mineralised zone, c) diorite in the mineralised zone.

The simulated variograms plot close to the training image variograms (Figure 4-62), which is to be expected - the training image serves as the source of two-point as well as multiple-point events. The experimental variograms using the conditioning data show a lack of structure stemming from sparse data spacing in the plane of mineralisation. This is a common

problem experienced by conventional two-point statistical algorithms due to difficulties in establishing reliable variogram models in cases when no close spaced drillholes are available.

The author fully acknowledges that there is space for improvement in the variogram validation, for instance, to extend it to all three directions of anisotropy and it can be considered as scope for future work.

Point support grade-tonnage curves. Albeit mechanised mining cannot take place at 2.5 m grid node support, the grade-tonnage curves at point support in Figure 4-63 are shown as a validation tool – to compare the performance of the simulation vs the conditioning data. The group of curves with negative slope of tangents are tonnage above cut-off and those with positive slope describe the average grade above cut-off.

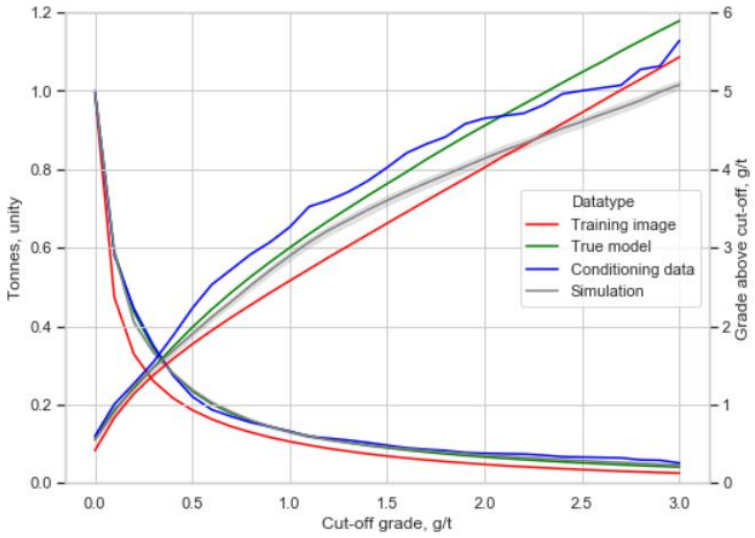


Figure 4-63. Grade-tonnage curves validation at point support (2.5 m x 2.5 m x 2.5 m). The group of curves with negative slope of tangents are tonnage above cut-off and those with positive slope describe the average grade above cut-off

For the reference data (TI, conditioning data and the ‘true’ model), there is a significant variation in tonnes and grade curves above cut-off, present between the ‘known’ (the TI) and the ‘unknown’ (the conditioning data and the ‘true’ model). It indicates that there is a non-stationarity present between the compared ‘known’ and ‘unknown’; the ‘unknown’ mineralisation is more robust.

Average grade above cut-off: the conditioning data curve plots above the ‘true’ model curve in the cut-off grade range 0.3 g/t to 2 g/t and the opposite is true above 2 g/t. The training image average grade above cut-off curve runs parallel to the ‘true’ model curve exhibiting a lower tenor along the whole grade range. The simulation curves mimic the ‘true’ model very well up to 0.5 g/t, deviating further above the cut-off grade 1.2 g/t and slumping below all the reference curves at cut-offs above 2.2 g/t. A well-known phenomenon is present – when the simulated statistics reproduce those of the training image. The lack of resemblance to the reference curves above the cut-off grade of 2.2 g/t can be explained by a lack of connectivity of the extreme gold grade values in the ensemble, calling for further improvement in the area, possibly by using connectivity constraints.

Tonnage curves show that the TI tonnes above cut-off have lower values along the whole range of cut-offs above 0.1 g/t, which is reflective of the fact that the SG was defined over a robust portion of the deposit down the plunge of mineralisation. The training image describes a larger and more variable domain than the volume over the simulation grid. Ideally, the comparison should be drawn using the portion of the training image referenced during the search for data events. The strongest controlling factor during search to ensure a stationarity-similar portion of the deposit is utilised was the potential field of the lithological contact.

The simulation tonnage curves plot very close to both, the ‘true’ model and conditioning data curves, showing good reproduction.

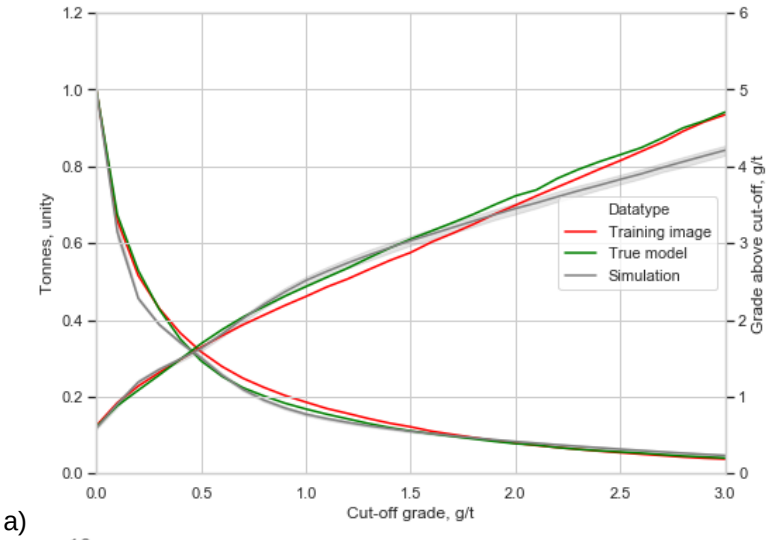
Overall, the point-support validation can be considered good, specifically with regards to the simulated tonnage reproduction, however, at the important high-grade cut-offs, above 2 g/t, the grade is underestimated in comparison to the ‘true’ model and conditioning data by approximately 10%.

4.6.4 **Global uncertainty assessment**

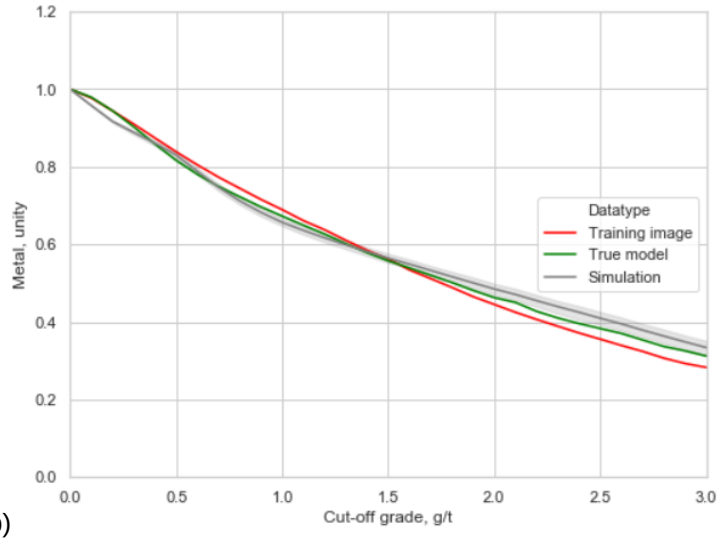
Assessment at the open pit SMU support and open pit cut-off grade

Gold grade realisations at point support were upscaled to open pit SMU support size (10 mE x 10 mN x 5 mRL) to understand the spread of uncertainty by open pit mining method. This exercise does not form a part of the Mineral Resource classification approach at the grid spacing of 40 m x 40 m as the simulated volume of economic mineralisation exceeds the annual production considerably. Tonnage, grade and metal curves at this support are presented in Figure 4-64: grade and tonnage are shown in image a, the metal curves in image b. Due to the absence of conditioning data at SMU support, the direct comparison is made to the ‘true’ model available over the simulation grid.

The average grade above cut-off curves of all three models plot closely below 2 g/t range, the simulation curves run in proximity to the ‘true’ model. Above the cut-off grade 1.5 g/t, the simulations grade curves deviate down from the ‘true’ model and the training image curves. At the cut-off 2 g/t the difference comprises about -5%. This is an undesirable behaviour since one of the research objectives is to validate the performance of the MPS approach for an underground scenario at a cut-off grade of 2 g/t.



a)



b)

Figure 4-64. Grade-tonnage (a) and metal (b) curves validation at open pit SMU support 10 mE x 10 mN x 5 mRL. In the image a, the group of curves with negative slope of tangents are tonnage above cut-off and those with positive slope describe the average grade above cut-off

For the tonnage, the simulated curves plot with unnoticeable spread of uncertainty from the median of the simulations, however, slightly below both, the training image and the 'true' model in the cut-off grade range 0.1 g/t to 0.4 g/t. The understatement of tonnes at low cut-off grade range is attributed to the poorer quality of pattern continuity reproduction in this range, and, hence, the lower SMU block averages due to geological dilution. Overall, the spread of uncertainty on the tonnage is very narrow.

The metal curves of the simulations plot very close to the 'true' model (Figure 4-64b), slightly overstating the contained metal above the cut-off 1.5 g/t – contribution of the positive trend in tonnes.

The tabulation for the open pit cut-off at 0.5 g/t is given in Table 4-6.

Table 4-6. Tabulation of the mineralised material at an open pit SMU support and an open pit cut-off grade 0.5 g/t (differences are tabulated vs the 'true' model)

Datatype	*Realisation*	*Tonnes, unity*	*Average grade above cut-off, g/t*	*Metal, unity*	*Difference vs true model*		
					Tonnes	*Grade*	*Metal*
Training image		0.32	1.64	0.84	8.1%	-2.7%	2.7%
True model		**0.29**	**1.69**	**0.82**			
Simulation	0	0.30	1.75	0.84	3.3%	3.3%	2.9%
	1	0.30	1.65	0.83	2.3%	-2.5%	1.1%
	2	0.32	1.58	0.85	8.5%	-6.5%	4.5%
	3	0.31	1.67	0.85	4.7%	-1.1%	4.6%
	4	0.29	1.60	0.81	-1.4%	-5.5%	-0.5%
	5	0.29	1.48	0.80	-1.2%	-12.6%	-1.7%
	6	0.31	1.63	0.85	6.8%	-3.9%	4.4%
	7	0.30	1.71	0.83	2.1%	1.4%	1.9%
	8	0.30	1.71	0.83	1.1%	1.4%	1.7%
	9	0.29	1.66	0.82	0.1%	-1.9%	1.0%
	10	0.30	1.51	0.81	1.6%	-10.5%	-0.4%
	11	0.29	1.58	0.81	-2.0%	-6.4%	-0.8%
	12	0.30	1.59	0.82	2.0%	-5.8%	0.2%
	13	0.33	1.68	0.87	12.8%	-0.9%	6.4%
	14	0.30	1.61	0.82	1.2%	-4.6%	0.7%
	15	0.29	1.67	0.83	0.6%	-1.4%	1.2%
	16	0.30	1.61	0.83	3.5%	-5.1%	1.1%
	17	0.30	1.57	0.82	3.6%	-7.0%	0.9%
	18	0.30	1.52	0.82	2.2%	-10.0%	0.1%
	19	0.30	1.63	0.82	0.9%	-3.8%	1.0%
Spread of uncertainty in simulations							
Simulation	**Mean**	**0.30**	**1.62**	**0.83**	**2.6%**	**-4.2%**	**1.5%**
	5th percentile	0.29	1.51	0.81	-1.5%	-10.6%	-0.9%
	95th percentile	0.32	1.72	0.85	8.7%	1.5%	4.7%

At 0.5 g/t cut-off, the tonnes are overestimated on average with 2.6%, the grade is underestimated (-4.2%), resulting in 1.5% more contained metal. The spread of uncertainty is very narrow with 90% of realisations showing an acceptable spread of errors: -1.5% to 8.7% for tonnes, -10.6% to 1.5% for grade and -0.9% to 4.7% for contained metal.

For the simulation results to be acceptable for stating the "Indicated" Mineral Resource category at the tested drill grid of 40 mE x 40 mN, the deviation

of the 5th and 95th percentiles should be within ±15% from the 'true' model values. In this case, the 90% confidence interval for all three parameters, tonnes, grade and metal above the cut-off of 0.5 g/t, are within 15% error from the 'true' model. The volume of the mineralised material is considerably larger than the required annual production volume. As noted by Verly et al (2014) in reference to confidence interval uncertainty by kriging estimation variance, the overall uncertainty can be significantly reduced by increasing the production or combining production from different areas. Conclusion stemming from this analysis is that the proposed MPS approach can be adopted for modelling the Mineral Resource of "Indicated" category in case of an open pit mining scenario at the spacing of 40 m x 40 m, albeit a larger production volume is considered.

Assessment at the open pit SMU support and underground cut-off grade, free selection assumed

For underground material to be classified as a Mineral Resource, it has to satisfy the adopted error threshold conditions at an underground cut-off value with an appropriate mining method applied for reporting. The volume of the mineralised material needs to be equal to annual production volume in case of "Indicated" Resource. In the case of an "Inferred" Mineral Resource, the error may exceed the adopted threshold parameters of 15% error at the 90% confidence limit. This intermediate tabulation step has been introduced to understand the confidence in the reported figures before applying a specific underground mining method. The same open pit SMU size of 10 m x 10 m x 5 m is used. The mineralised simulation volume is about 1.5 times larger than the required annual production would be and it would serve as a good indication of the compliance of the figures with the reporting codes, prior to appropriate mining method application.

A tabulation at the underground cut-off 2.0 g/t is presented in Table 4-7 assuming free and independent selection is achievable at SMU support 10 m x 10 m x 5 m. Figure 4-64a depicts the grade-tonnage relationship at the cut-off of 2 g/t.

The average grade and contained metal above 2 g/t in 90% of realisations are reporting within acceptable limits of ±15% error around the mean of the base case. The simulated tonnes, however, are tending to an overly optimistic side – the 95th percentile is 17.1% above the target mean.

Table 4-7. Tabulation of the mineralised material at an open pit SMU support and an underground cut-off grade 2.0 g/t (differences are tabulated vs the 'true' model)

Datatype	*Realisation*	*Tonnes, unity*	*Average grade above cut-off, g/t*	*Metal, unity*	*Difference vs true model*		
					Tonnes	*Grade*	*Metal*
Training image		0.079	3.49	0.445	1.8%	-3.4%	-3.8%
True model		**0.078**	**3.62**	**0.463**			
Simulation	0	0.091	3.62	0.522	16.6%	0.3%	12.8%
	1	0.082	3.59	0.491	5.4%	-0.7%	6.2%
	2	0.081	3.53	0.486	4.4%	-2.2%	5.1%
	3	0.087	3.60	0.518	11.3%	-0.5%	11.9%
	4	0.079	3.31	0.462	2.2%	-8.6%	-0.2%
	5	0.069	3.20	0.414	-11.2%	-11.6%	-10.6%
	6	0.084	3.52	0.497	8.4%	-2.6%	7.4%
	7	0.089	3.57	0.516	14.7%	-1.2%	11.6%
	8	0.085	3.73	0.517	9.2%	3.1%	11.7%
	9	0.083	3.50	0.493	7.2%	-3.3%	6.6%
	10	0.073	3.29	0.430	-6.7%	-9.1%	-7.0%
	11	0.078	3.37	0.468	0.3%	-6.7%	1.2%
	12	0.080	3.43	0.473	3.3%	-5.1%	2.3%
	13	0.098	3.42	0.525	26.0%	-5.4%	13.5%
	14	0.081	3.36	0.467	4.1%	-7.0%	0.9%
	15	0.089	3.32	0.495	14.0%	-8.1%	6.9%
	16	0.083	3.32	0.470	7.4%	-8.1%	1.5%
	17	0.080	3.27	0.454	3.5%	-9.4%	-1.8%
	18	0.076	3.24	0.443	-1.9%	-10.5%	-4.4%
	19	0.084	3.34	0.482	8.6%	-7.7%	4.3%
Spread of uncertainty in simulations							
Simulation	Mean	**0.08**	**3.43**	**0.48**	**6.3%**	**-5.2%**	**4.0%**
	5th percentile	0.072	3.23	0.43	-6.9%	-10.5%	-7.2%
	95th percentile	0.091	3.63	0.52	17.1%	0.4%	12.8%

A conclusion is that prior to applying an appropriate underground mining method the material would still qualify for an "Inferred" or even an "Indicated" Mineral Resource category adopted in this research, as the metal content error is within the 15% threshold 90% of the times, albeit the tonnage is positively biased exceeding the upper threshold.

Assessment at the underground cut-off and mining method applied

To take into cognisance an appropriate mining method for reporting, the simulated realisations were subjected to processing in the Mineable Shape Optimiser to produce underground stopes. 20 simulated models were used as input to MSO, each one based on a unique realisation of lithology and gold values. No underground development design was taken into consideration. No further merging of stopes or elimination of single stopes, no consideration for support pillars which would be an appropriate approach in practical situations, has been taken cognisance of. The MSO parameters are summarised in Table 4-8.

Table 4-8. Main parameters of the Mineable Shape Optimiser used in the definition of underground stopes

Parameter	*Value*
Economics	
Optimisation method	Maximise Stope Grade/Value above cut-off
Cut-off grade, g/t	2
Shape Framework	
Section interval increments, m	5
Number of sections	41
Level interval increments, m	25
Number of levels	11
Controls on a stope shape	
Stope width minimum, m	5
Stope width maximum, m	50
Minimum pillar between parallel stopes	5
Near stope dilution	0
Far stope dilution	0
Stopes hanging wall dip angles minimum	40
Stopes hanging wall dip angles maximum	70
Stopes footwall dip angles minimum	40
Stopes footwall dip angles maximum	70
Maximum dip angle change, degrees	20
Stope strike angle minimum, degrees	-45
Stope strike angle maximum, degrees	45
Maximum strike angle change, degrees	20
Maximum stope thickness ratio, top to bottom	2.25
Maximum stope thickness ratio, left to right	2.25
Options	
Stope smoothing match type	gap
Stope smoothing match direction	U direction
Match tolerance	5
Stope merging	U direction
Stope merging interval	2
Stope merging minimum-maximum range	0-4

Figure 4-65 shows examples of stope designs: 4-65a for the ‘true’ model and Figures 4-65b, 4-65c and 4-65d show stope designs for three realisations of gold grade. In all displayed versions of the stope design, the central diagonal of the models has not been covered by stopes, while two opposite corners (the top south-east and bottom north-west) have high confidence of being within the mineable model. This points towards a good performance of the DS in translating the non-stationarity settings into the mine design. Local uncertainty and Mineral Resource classification would be an appropriate extension of the workflow for the engineering purposes.

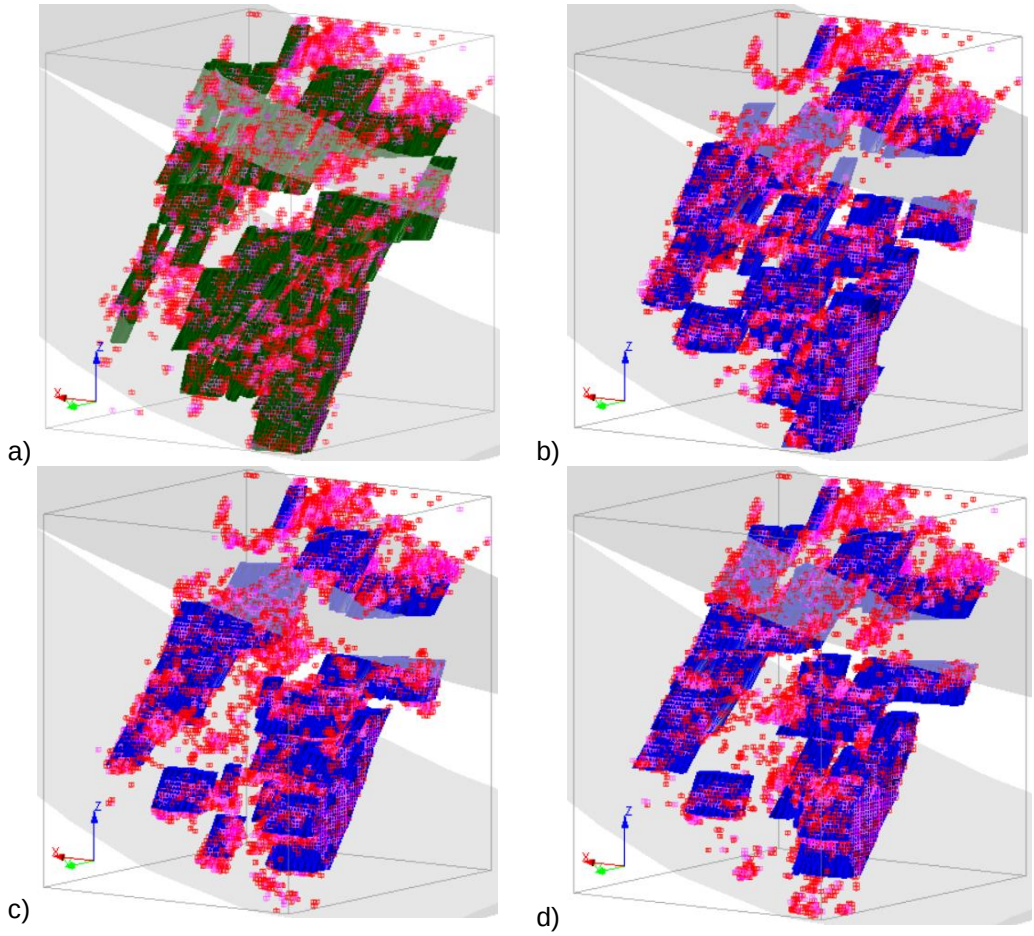


Figure 4-65. Isoclinal views showing stopes created with Mineable Shape Optimisation algorithm using realisations of gold based on the DS simulation. Cells above 1 g/t are shown in magenta, above 2 g/t – in red. a) stopes designed on the ‘true’ model, b-d) stopes designed for three realisations of gold (ID 0, 1 and 3). The two grey semi-transparent surfaces represent the top and bottom of the simulated volume. The ‘known’ top portion of the ‘true’ model (above the top grey surface) is frozen in all simulations, it has been used as conditioning data in the DS reconstructive mode

No grade, tonnage or metal curves are presented in this section as there is no free selection and the material mined from the stopes is to be fed to the plant in its entirety. The assessment of global uncertainty in the SG volume vs the 'true' model is presented by tabulation only (Table 4-9). A volume equivalent to annual production by the underground mining method is used.

The reporting is done for all input models, the 'true' model and the simulated ones, within the corresponding stope designs. The upper portion of the volumes (above the top grey surface in Figure 4-65) which was used in exhaustive conditioning of the simulated models was excluded from the comparison, to avoid risk understatement.

Table 4-9. Tabulation of the mineralised material tonnes, grade and metal in the underground stope designs comparing simulated results vs the 'true' model. The volume used for comparison represents an annual production volume. Cells showing difference against the 'true' model in excess of 15% are shaded in grey

Datatype	*Realisation*	*Tonnes, unity*	*Grade*	*Metal, unity*	*Difference vs 'true' model*		
					Tonnes	*Grade*	*Metal*
"True" model		**1.00**	**5.13**	**1.00**			
Simulation	real 0	1.31	4.61	1.18	30.8%	-10.0%	17.7%
	real 1	1.13	4.69	1.03	13.1%	-8.6%	3.3%
	real 2	1.12	4.54	0.99	11.9%	-11.5%	-1.0%
	real 3	1.24	4.54	1.10	24.1%	-11.4%	10.0%
	real 4	1.04	4.43	0.90	3.8%	-13.6%	-10.3%
	real 5	0.90	4.25	0.74	-10.3%	-17.0%	-25.6%
	real 6	1.20	4.47	1.05	20.4%	-12.8%	5.0%
	real 7	1.29	4.61	1.16	28.9%	-10.2%	15.8%
	real 8	1.23	4.72	1.13	23.1%	-7.9%	13.4%
	real 9	1.11	4.53	0.98	10.9%	-11.6%	-1.9%
	real 10	0.97	4.27	0.81	-3.0%	-16.7%	-19.2%
	real 11	1.04	4.44	0.90	4.0%	-13.4%	-10.0%
	real 12	1.11	4.48	0.97	10.9%	-12.5%	-3.0%
	real 13	1.36	4.37	1.16	36.3%	-14.8%	16.1%
	real 14	1.07	4.35	0.90	6.8%	-15.2%	-9.5%
	real 15	1.22	4.31	1.03	22.3%	-15.9%	2.9%
	real 16	1.11	4.41	0.95	11.0%	-14.0%	-4.6%
	real 17	1.09	4.27	0.91	8.8%	-16.6%	-9.3%
	real 18	1.02	4.29	0.86	2.3%	-16.3%	-14.4%
	real 19	1.18	4.34	1.00	18.4%	-15.3%	0.3%
Spread of uncertainty in simulations							
Simulation	**Mean**	**1.14**	**4.45**	**0.99**	**13.7%**	**-13.3%**	**-1.2%**
	5th percentile	0.97	4.27	0.81	-3.4%	-16.7%	-19.5%
	95th percentile	1.31	4.69	1.16	31.1%	-8.6%	16.2%

The mean of simulations shows overestimation of tonnes (13.7%) and underestimation on grade (-13.3%) in comparison to the 'true' model, resulting in closely reproduced contained metal (-1.2%). The spread of uncertainty, however, is quite considerable and falls outside of the desired limits of ±15% error for the "Indicated" Mineral Resource. At the 5th percentile end of realisations, the tonnes are below the 'true' model mean values by 3.4%, the grade by 16.7%, and the metal by 19.5%. At the 95th percentile, the tonnes are higher than the 'true' model mean by 31.1%, the grade is lower by 8.6% and the resulting metal exceeds the 'true' model figure by 16.2%. These results are outside the desired limits for the "Indicated" Mineral Resource, however, can be considered acceptable for the "Inferred" Mineral Resource category.

Recalling Section 4.5 "Derived conditioning data and its uncertainty", there is some risk associated with the placement of the drillholes. Namely, the 95th percentile spills outside the desired +15% error from the 'true' model mean (1.68 g/t vs 1.64 g/t in Figure 4-37c), which also contributed to the risk observed.

4.7 Simulation workflow chart

The workflow for the practical application of MPS established during this study is shown in Figures 4-66 – 4-70. It has been broken into segments by task. This workflow is an extension to the general one proposed in the beginning of Chapter 4, in Figure 4-1.

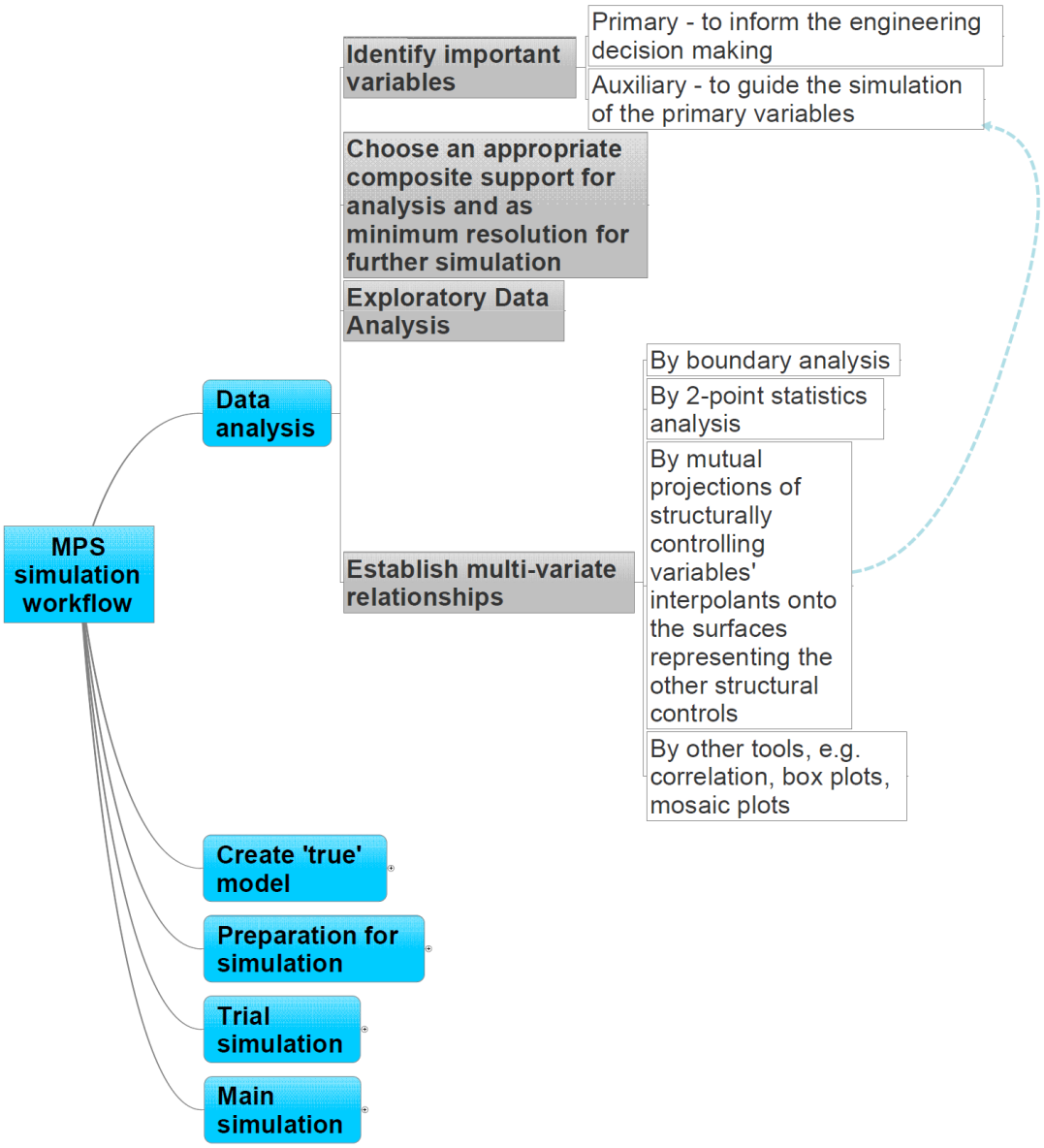


Figure 4-66. Workflow chart for performing MPS simulation with the DS algorithm in Geita Hill deposit environment – Data analysis

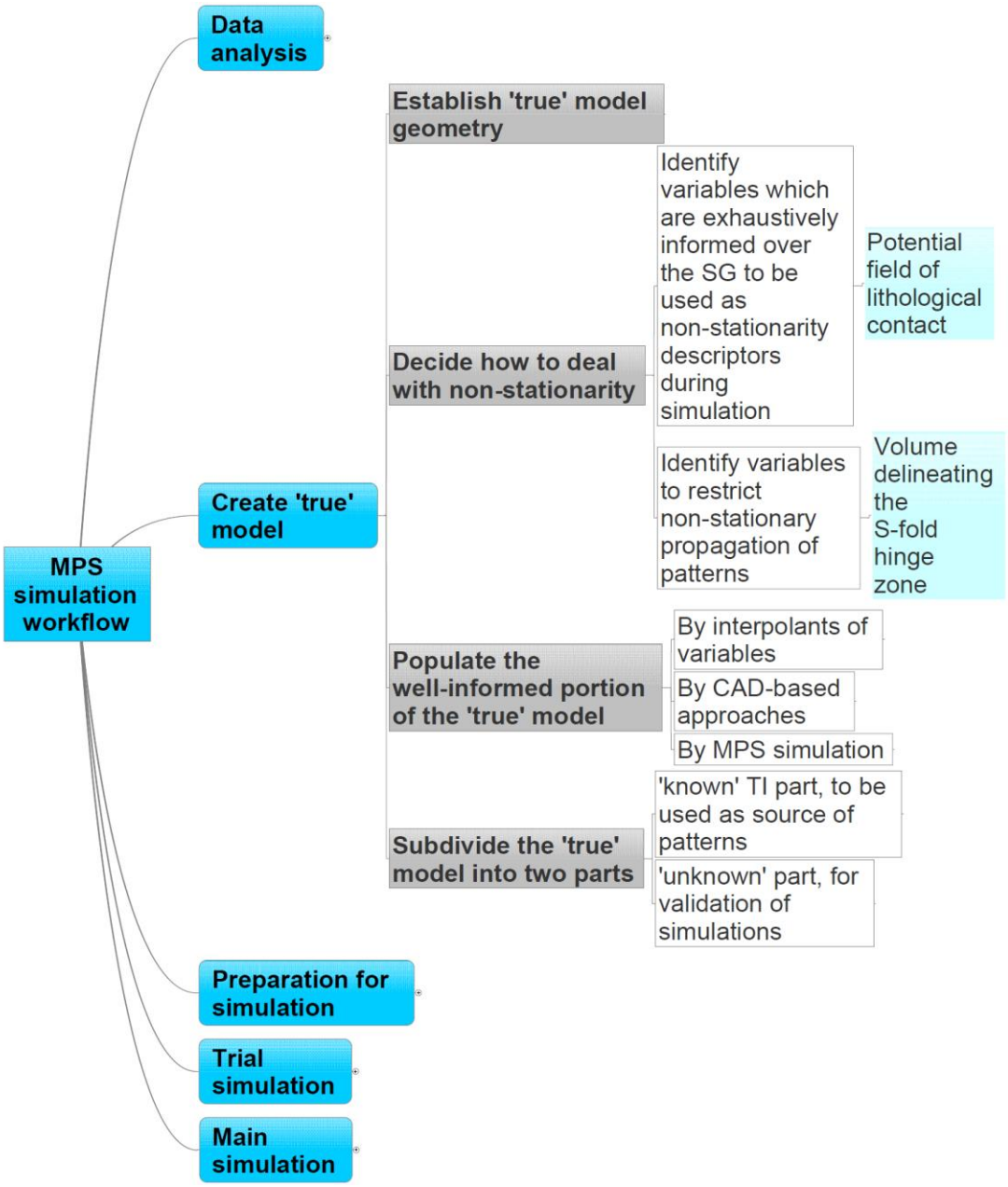


Figure 4-67. Workflow chart for performing MPS simulation with the DS algorithm in Geita Hill deposit environment – Creating ‘true’ model

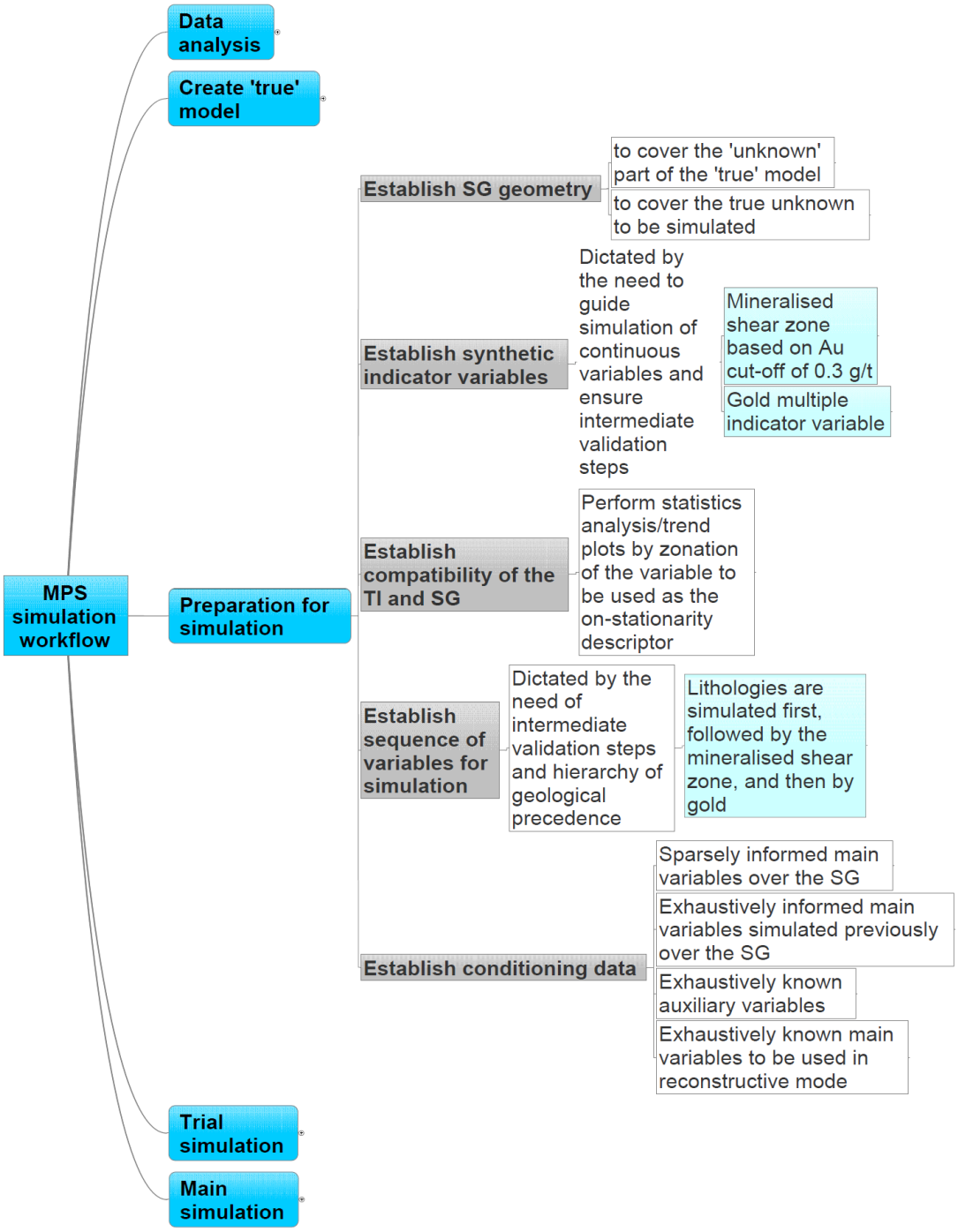


Figure 4-68. Workflow chart for performing MPS simulation with the DS algorithm in Geita Hill deposit environment – Preparation for simulation

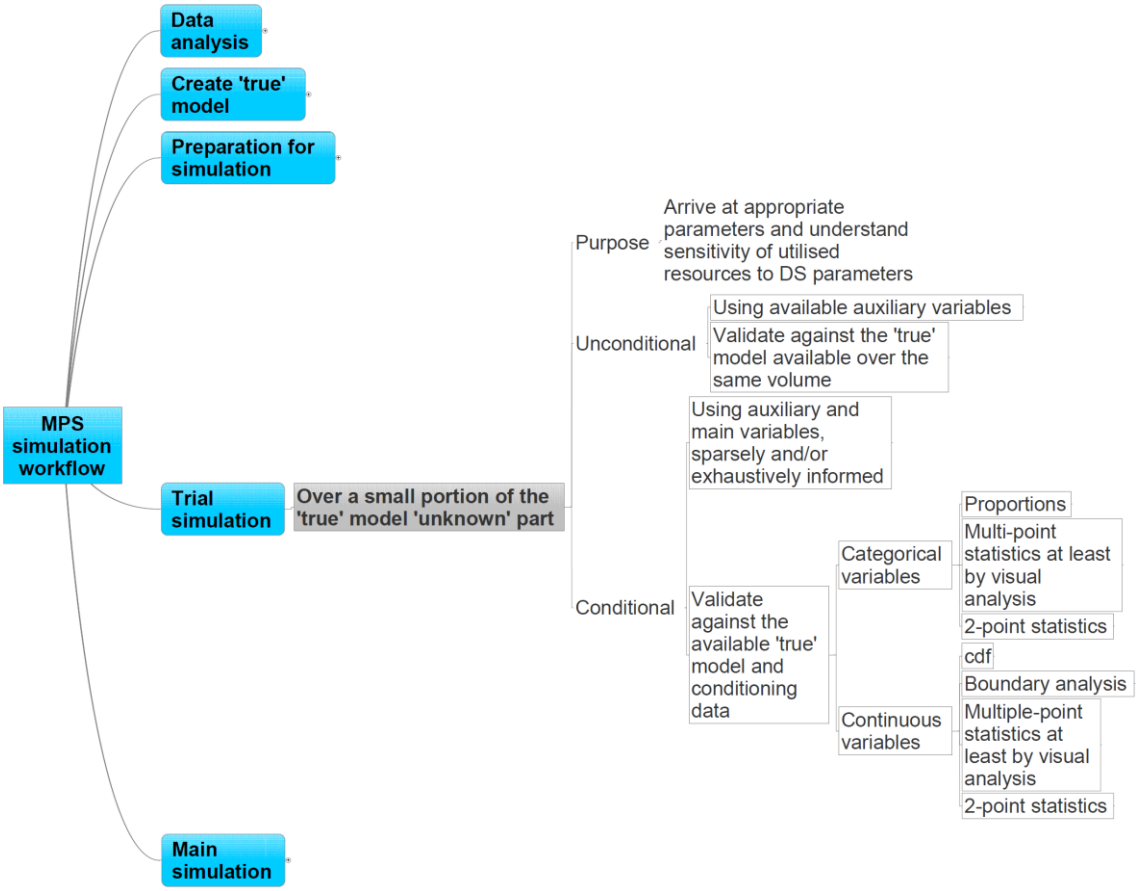


Figure 4-69. Workflow chart for performing MPS simulation with the DS algorithm in Geita Hill deposit environment – Trial simulation

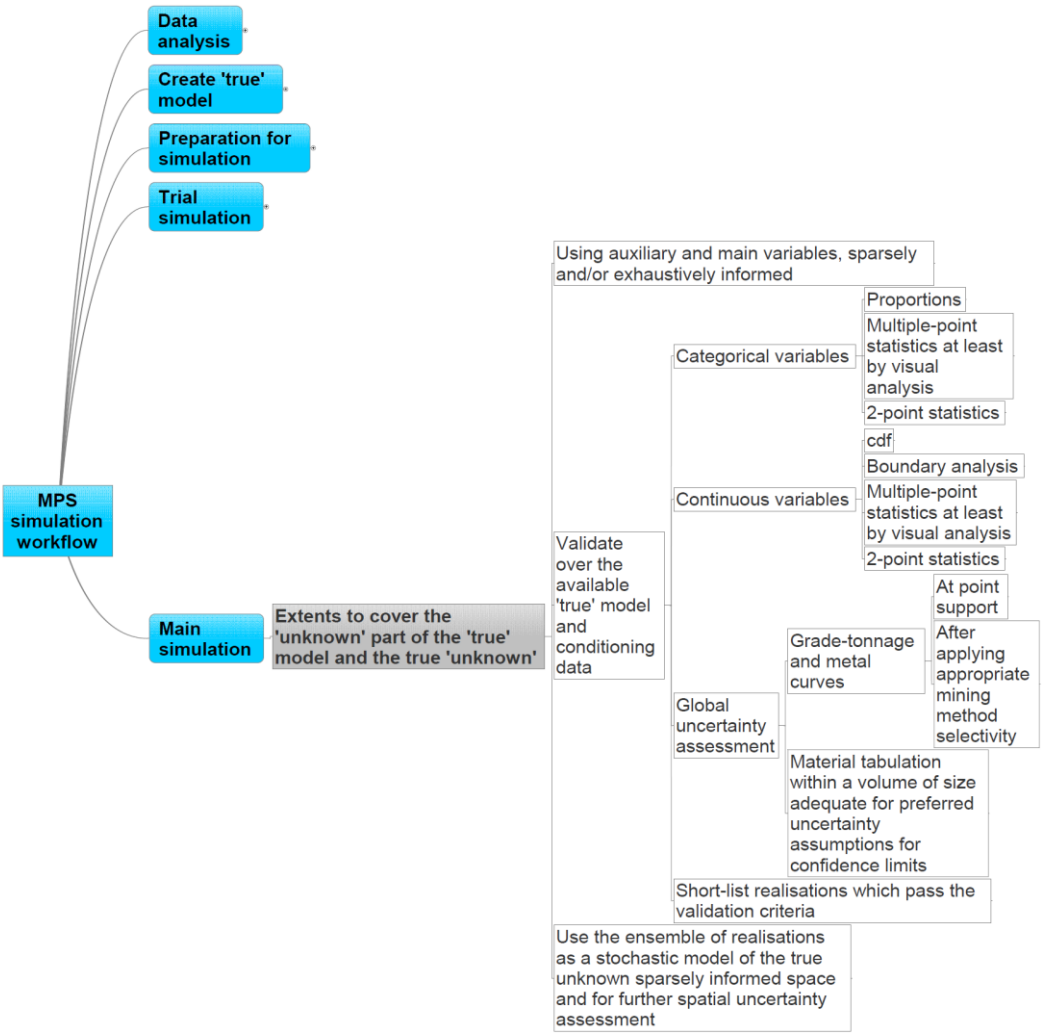


Figure 4-70. Workflow chart for performing MPS simulation with the DS algorithm in Geita Hill deposit environment – Main simulation

4.8 Chapter summary

A practical case of performing an MPS simulation in a complex geological setting has been covered in detail in the current chapter. It commenced with a description of the deposit geology and was followed by data analysis. Emphasis was placed on understanding the complex controls on mineralisation, where the macroscale structural architecture and competency contrast between the two main lithologies having a primary role in the propagation of shearing, combined with sulphidation of iron-rich BIF affected the deposition of gold mineralisation.

EDA covered boundary analysis for the point support: 1 m sample composites as well as 2.5 m bench composites. It was done to understand the character of mineralisation in the vicinity of the contact at the point support and use it further as one of the validation tools for the simulation performance.

Considerable effort was dedicated to the process of creating a representative ‘true’ model of the deposit, to be further used in an ML-like simulation framework: when the top part of the ‘true’ model is used as a training image guiding simulations while the bottom part is reserved for validation.

A 3D spatial analysis was undertaken to understand complex relationships between different structural controls. It was given precedence over the understanding of two-point statistical relationships in the deposit due to the complex nature of mineralisation and pronounced non-stationarity. In the process, an explicit 3D structural model of the deposit was created revealing the folding architecture and its control on shearing propagation. This integrated process of the spatial analysis allowed identification of an ensemble of variables important for further MPS simulation.

Three primary variables (lithology with associated density, and gold grades) and several secondary variables were created to guide the simulation as well as to assist in describing the non-stationary setting of the deposit.

The investigation into understanding of risk started with uncertainty in the derived conditioning data. 84 sets of pseudo drillholes were extracted from the ‘true’ model. The comparison was made between the dataset statistics against the ‘true’ model available over the simulation grid. The lithology and the mineralised zone proportions were globally very close to the ‘true’ model proportions, within 1-3% difference. However, the 95^{th} percentile of the mean gold grade values in the pseudo grids showed to be higher than those indicated by the 15% error intervals from the ‘true’ model mean: 1.68 vs 1.64 g/t at point support (Figure 4-37c). Performed at a point support, it has no relevance to the classification of a Mineral Resource as required by the

reporting codes, the test in derived data uncertainty allowed to quantify uncertainty in the source data. It indicated that significant risk in the gold grade can be expected from the placement of the 40 mE x 40 mN unbiased sampling grid, prior to committing to any estimation technique.

A detailed description of the simulation process followed next. A hierarchical simulation approach was adopted: the lithology was simulated first, followed by mineralised shear zone simulation. It enabled to perform intermediate validation steps for the two variables which have a direct influence on the resulting mineralised tonnes, prior to commencing with gold simulation. Good results were achieved when validating the realisations of these categorical variables. However, the subsequent simulation of gold grades revealed the importance of reproducing truthful finesse of geomorphological detail for the two controlling variables. Overstated tabularity and bulkiness of lithology and mineralised zone shapes caused the algorithm to retrieve the patterns preferentially from the central core areas of the mineralised morphologies in the training image, resulting in overstated gold.

The gold simulation description focused on the problems encountered and the steps undertaken to overcome them. The validation of the final runs covered visual review, reproduction of the global mean values, variance, boundary analysis and variograms.

While finalising this research, some test work continued in parallel to improve the quality of the gold grade simulation. Good results were produced by resorting to three thresholds for the conditioning multiple-indicator gold variable rather than four in the preceding runs. The improvement is attributable to the presence of more robust geo-bodies' clusters. This simulation run was completed after the final stope design and uncertainty assessment were finalised and was not used in any statistical analysis. There is an opportunity for improving the variogram validation, the uncertainty assessment and appropriateness of the framework for the Mineral Resource classification for these superseding gold grade realisations, which can be considered for future work.

Histogram validation was done in the form of grade-tonnage curves at point support to understand the differences between the simulations, the input conditioning data and the ‘true’ model presented in economic value terms. The simulated tonnes were very well reproduced in comparison to the ‘true’ model and conditioning data. The grade was erring on the conservative side: at the important high-grade cut-off of 2 g/t, the grade is underestimated in comparison to the ‘true’ model and conditioning data by approximately 10%. This is a well-sited property of MPS when the simulated statistics closely reproduce the training image statistics.

An uncertainty assessment concentrated on the global comparison of the grade-tonnage curves at different support sizes of mining units. No spatial risk assessment took place. While of significant value-add for the local Mineral Resource classification, and mine engineering applications, it is considered to be beyond the scope of the current research.

Comparison of the grade-tonnage and metal curves at an open pit SMU support was done to understand the reproduction of the recoverable resource at mineable support, assuming free and independent selection of material is possible. This test was performed at two cut-off values: at an open cut-off of 0.5 g/t and an underground cut-off of 2 g/t at SMU support. For the 0.5 g/t, all three parameters, tonnes, grade and metal are within 15% error from the ‘true’ model 90% of the cases. The MPS approach can be considered suitable for modelling the Mineral Resource of “Inferred” or even “Indicated” category, as defined in the current thesis, since the simulated volume of mineralised material is a few times larger than the required annual production volume. The confidence of reporting decreases towards higher cut-off values and higher demands on mining selectivity. At the cut-off of 2 g/t, while still assuming free and independent selection, it showed close reproduction of tonnes (+6%), grade (-5%) and metal (+4%) with acceptable lower and upper confidence limits on all parameters except upper confidence limit on tonnes (+17%). Raising the cut-off grade above 2 g/t would result in an understatement of grades and a slight overestimation of tonnes producing some overestimation of contained metal (within +5%).

The material would qualify for “Inferred” and, marginally, for the “Indicated” Mineral Resource category, albeit a positive bias on the tonnage present.

Lastly, global uncertainty assessment was done after applying the assumptions of an appropriate underground mining method. Visually, the simulations show repeatability of the stope placement vs the ‘true’ model, where the diagonal waste zone of the ‘true’ model has been correctly reproduced as waste in the simulations. This is mainly attributed to the usage of the lithological contact potential field in the conditioning of the gold simulation, and strong conditioning influence of the ‘true’ model used in reconstructive mode of the DS in the top portion of the simulation grid.

The spread of uncertainty at the underground cut-off of 2 g/t after applying the mining method considerations was relatively wide, beyond the desired limits, specifically at the upper 95th percentile, where the tonnes were 31% higher than in the ‘true’ model. This is believed to be a result of too bulky geometries of the lithologies and the mineralised zone used in conditioning the gold grade simulation.

The chapter concluded with a workflow that can be used in a similar geological setting.

5 FINDINGS AND RECOMMENDATIONS

5.1 Summary of findings

The current research focused on utilising the Direct Sampling multiple-point statistical simulation algorithm as developed by Mariethoz (2009) to create models for a mineral deposit with complex structural architecture, multiple controls on mineralisation and spatial relationships beyond two-point statistics. The main emphasis has been to generate stochastic images to represent a possible “Indicated” or “Inferred” Mineral Resource at widely spaced drilling on a grid exceeding the spacing conventionally used for the mineralisation type under consideration. The reasoning behind choosing a lower confidence category of the Mineral Resource was to test the performance of the algorithm in sparsely informed scenarios when main-stream estimation techniques struggle to provide explicit models of reality required for preliminary engineering studies.

The Geita Mine deposits yield themselves easily to well-behaved performance by most conventional estimation techniques due to highly pronounced logarithmic distribution of gold in relatively narrow shear zones. The difficulty arises in sparsely informed cases and in areas with poorer mineralised splays when realistic representation of mineralisation is required, and engineering decisions are sensitive to inaccuracies in the model. Historically, variogram-based methods have been found to underperform due to their reliance on dense data to model continuity. Gaussian-based methods can result in amorphous shapes of mineralisation and are inefficient in reproducing non-ellipsoidal mineralisation morphology and connectivity of high grade (Journel & Deutsch, 1993; Zinn & Harvey, 2003; Renard & Allard, 2013; Mariethoz & Caers, 2015). These maximise entropy (Journel and Deutsch, 1993) which is described by spatial correlation of extreme values trending to zero.

The findings of the research specific to the Geita Hill orebody modelling as well as generic findings for applicability of MPS in future are described below.

5.1.1 **Modelling process flow**

With regards to this group of findings, a first 3D explicit stochastic lithological model of this deposit has been developed and a workflow to do it in future proposed. The two main lithologies of the deposit, BIF and diorite, have considerably differing densities, and modelling them correctly is of importance for deriving accurate tonnage, mineralised or not. Besides, the contact between the two lithologies has a pronounced control on gold mineralisation and the attitude of the high-grade ore shoots.

Previous attempts at modelling folded lithology explicitly with CAD-based or variogram-based methods have been unsuccessful. The former failed due to complex and abundant intercalated geometries of the lithological boundary, the latter due to the inability of the method to model curvilinear continuity of units beyond two-point statistics (an example is shown in Figure 4-19) even in the presence of dense production data.

While modelling of the lithology in the current study was a trial-and-error exercise, in future it does not have to be, provided sufficient orientation structural measurements are compiled and appropriately declustered to create a representative potential field of ductile deformation.

In the process of deriving the lithological model, it was demonstrated that the frequency and the attitude of the lithological contact between the two main lithologies, BIF and diorite, imposed an important control on the morphology of the ore shoots. The shear splays developed strictly within the S-fold hinge zone of the fold. The ore shoots preferentially occupied the ‘braided’ shears with elevated mineralisation tenor at their intersections and along the lithological contact. The higher tenor of mineralisation was found in the north-west portion of the deposit where the frequency of intrusive bodies is increased.

While the mineralised shear zone imposed the primary control on the distribution of the ore, it was found that using the potential field of the lithological contact as a non-stationarity descriptor and an auxiliary variable to guide simulation of gold improved the outcomes, as it resulted in neighbourhoods being extracted from areas up the plunge of the folded lithologies in the TI.

During the gold grade simulation, it was realised that introducing categorical variables in co-simulation framework improves continuity of reproduced patterns of gold and its statistics. Two variables were used, a binary indicator variable for the mineralised zone and a multiple-indicator gold grade variable. Simulating the mineralised zone binary indicator variable first allowed to arrive at satisfactory reproduction of global statistics of mineralised material, prior to committing to more time-consuming gold grade simulation. Using block conditioning for the mineralised zone proportions helped to achieve acceptable reproduction of the local proportions for block values in the simulation.

The second indicator variable was introduced purely to improve the continuity of gold grade patterns. As a first attempt, four indicator thresholds were used, however, the reproduction of the pattern morphology for low and medium grade categories was not good. Resorting to three indicator thresholds improved the reproduction of patterns. It is suggested that two or three thresholds might be sufficient, and the selection of the cut-off values should ensure a sufficient volume of geobodies to be found in the neighbourhood.

Another factor that improved gold grade simulation was using the same ‘true’ model as a source of patterns for simulation of the mineralised zone. In the first attempt, an approach utilising an elementary training image and a rotational field for the shear zone attitude resulted in morphologies of too bulky an appearance which prompted the overstatement of subsequent gold grade simulation. The same was true for lithologies – while an elementary TI usage resulted in seemingly realistic shapes with good honouring of the

conditioning data and conditioning proportions, these were more regular and continuous than those observed in the exhaustively informed ‘true’ model.

Overall, the work performed in this study demonstrated that in order to successfully approach an MPS simulation in a non-stationarity framework, the training image must be sufficiently large and defined in an appropriate range of values for the variable chosen to describe the non-stationarity setting.

5.1.2 **Validation**

The primary validation means in the study was visual checking. It is an important tool as a trained eye of a specialist serves as an indispensable instrument in picking up complex statistics on an intuitive level, not only between a simulation and an input training image but also between a simulation and an internal library based on experience.

The capability of the visual cortex in validation should not be under-estimated. Quite frequently, good matching between the source and the output images can be obtained with different statistical tools such as histograms or variograms of the datasets, while differences in morphologies and connectivity patterns between the training image and the simulated realisations can be drastic and require complex image-matching algorithms. In complexity, visual validation by a specialist is a many-fold more intricate tool than any validation techniques yet developed.

Simulation of indicator variables (mineralised zone and lithologies) showed very good categorical proportion reproduction, within 3% error vs the ‘true’ model, and this approach can be used as-is in future simulations.

Results of validation for gold simulations was not so direct. Boundary reproduction for transition in mean gold grade values across different interfaces yielded good results: between the mineralised lithologies, waste lithologies, and across the ore-waste boundary.

Variograms of the simulated gold realisations approximated the training image variograms rather than the conditioning data or the 'true' model – a feature normally expected with MPS techniques, as the TI serves as an explicit prior model. The controlling potential field was honoured closely in gold grade simulations, which expressed itself in the reproduction of mineralised zonation along the lithological contact.

It was considered important to understand the performance of global grade-tonnage curves for the simulated volume at point support, albeit for validation only and of no consequence to the classification of the Mineral Resource. No comparison to discrete Gaussian models was done in this study; the conclusions on the performance were drawn by comparing the simulation results to the 'true' model available over the simulation volume at open pit SMU and underground stope selectivity level. They indicated that, at lower grade cut-offs and an assumption of free and independent selection the mineralised material showed low uncertainty and complied to the required annual production 90% confidence 15% error rule of the "Indicated" Mineral Resource. A volume for this analysis was in excess of two-fold of an annual production volume, indicating sufficient confidence; although not explicitly tested at the annual production for the SMU support, the uncertainty is expected to be reduced when increasing the production volume (Verly et al., 2014). The spread of uncertainty widened with an imposed higher cut-off value and a more selective underground mining method. The summary table below demonstrates this.

Table 5-1. Tabulation of differences in tonnes, grade and metal between the simulated models and the 'true' model at different mining scenarios

Simulated statistic	*Difference vs true model*								
	open-pit SMU support, free selection						*underground mining stoping*		
	at 0.5g/t cut-off			*at 2.0g/t cut-off*			*at 2.0g/t cut-off*		
	Tonnes	*Grade*	*Metal*	*Tonnes*	*Grade*	*Metal*	*Tonnes*	*Grade*	*Metal*
Mean	**2.6%**	**-4.2%**	**1.5%**	**6.3%**	**-5.2%**	**4.0%**	**13.7%**	**-13.3%**	**-1.2%**
5th percentile	-1.5%	-10.6%	-0.9%	-6.9%	-10.5%	-7.2%	-3.4%	-16.7%	-19.5%
95th percentile	8.7%	1.5%	4.7%	17.1%	0.4%	12.8%	31.1%	-8.6%	16.2%

It might mean that while in the current research, a low cut-off grade has been adopted for simulating and validating the mineralised zone prior to gold simulation, a better approach for a mining engineering application

would be to simulate and sign-off the mineralised material reproduction at the target cut-off grade to be used further for mine design.

As applicable to precious metal mining, it is important to achieve an acceptable reproduction of total mineralised tonnes and morphology of mineralisation at a specified mining unit resolution and important economic cut-offs early in the modelling process. For this reason, albeit the DS can be run in multivariate mode, a safer approach was adopted - to perform simulation in a hierarchical fashion when the reproduction of total mineralised tonnage is accomplished first in the process and used in further simulations as an exhaustively informing secondary variable.

5.1.3 **General**

Boundary analysis showed that geostatistical domaining for delineation of high-grade core for this deposit is not straight forward and might not be possible by relying on geological logging. The high grade is concentrated along the contact between the BIF and diorite and preferentially located within BIF. Incorporating alteration information and geochemistry would possibly assist the process, however, this was beyond the scope of the current research. Despite this, it was demonstrated that the DS achieved promising results, as shown by boundary analysis and grade tonnage curve validation.

Stope definition in the current research was based on an economic viability decision at a predefined cut-off grade. No geological boundaries were taken cognisance of as there are no sharp geological boundaries for limiting the mineralisation extent. At the same time, engineering decisions rely on the truthfulness of the shapes at the economic cut-off important for mining. The approach has shown to be reasonable in the current implementation of MPS; the arrangement of stope shapes based on simulated realisation was close to those in the ‘true’ model.

A reader is reminded (as mentioned in Section 4.6.3 Gold Simulation) that two streams of tests were carried out during gold grade simulation – one

aimed at continually improving the resulting simulations, while the other one was to finalise through the workflow one of the shortlisted ensembles of realisations, albeit not the best one achieved. Better results were obtained with the first stream, however only late in the research when the bulk of work on validation, stope design and uncertainty assessment has already been finished for the second stream. It would be of interest to accomplish the first stream of test-work, however, this is beyond the scope of the current work. Practicalities of creating a training image were also touched upon in the thesis, such as employing a grade transformation to generate a representative gold grade interpolant in the 'true' model.

The DS presents an algorithm capable of reproducing real-life 3D problems. While many other MPS methods can be similarly successful in simple synthetic cases, to the author's knowledge there is none available at the moment of writing this work, capable of handling complex multi-variate geological environments with an abundance of data.

5.2 Envisaged contribution to new knowledge

The main contribution of this research is to demonstrate the potential of multiple-point simulation and, specifically, the Direct Sampling method, in modelling a complex geological environment. The research provides a practical framework based on a detailed case study to substantiate usage of multiple-point statistics in structurally folded non-stationary geological settings. An application in a multivariate case has been demonstrated using a large 3D dataset.

5.2.1 Explicit 3D model

During the process of TI construction and investigation of the important controlling variables, a first explicit 3D structural and ore-genetic model of Geita Hill deposit originated. It extended the understanding of the structural framework by correlating regional scale folding with the gold mineralisation.

As emphasised by Cowan (2012, 2014), detailed descriptions of structural 3D models are not common in public domain and are normally limited to simplistic 2D schematic representations and cross-sections of deposits, making it difficult to convey a 3D shape of a deposit. Very little information on 3D morphology of ore bodies is available in academic literature to refine theoretical ore genesis models. The topic of incorporating insufficient geological information in domaining has been emphasised by Sterk, de Jong, Partington, Kerkvliet and van de Ven (2019), who conducted a review of 161 public reports on Mineral Resource estimates announced to global markets since January 2017. It was found that less than 50% of estimates use geological information, such as lithology or alteration in addition to pure grade cut-offs for confining mineralised domains.

The S-fold hinge volume was defined by digitising, however, in future it should be possible to use a derivative of the potential field of lithological contact to automate the process.

Building an explicit ‘true’ 3D model of the deposit allowed an insight into the spatial distribution of mineralisation and the extension of the mineralisation at depth. Projection of the lithologies onto the shears and over-imposing potential field of bedding revealed that in the western part of the orebody the plunge of high-grade mineralisation is to the local north-east, and in the eastern poorer mineralised part to the north (Figures 4-10, 4-14, 4-15, 4-16), and it is expected that the grade tenor will increase in that direction. This can be considered for further drilling targeting. The high-grade shoots pattern is more complicated, however; it is a function of the complex sinusoidal intersection of the lithological bedding and shears. It is considered important to incorporate explicit models of the structural architecture in the exploration targeting at the meso-scale.

5.2.2 **Definition of non-stationarity by structural fields**

Stationarity decisions are the most important ones in the Mineral Resource estimation process. They can be quite subjective. An advantage of the

algorithm is that no explicit assumptions with regards to stationarity need to be made. The modelling workflow is akin to a geologist's thinking process - to interpret important controls on mineralisation and represent them in terms of variables. These auxiliary variables are exhaustively known and guide the simulation in the presence of sparsely informed main variables.

In this research, no explicit stationarity domain definition was made. The outcomes show potential for this framework however more rigorous validation tools are required to provide quantified assessment of the simulations.

MPS and the DS, in particular, offer a new paradigm for modelling and estimation - it lifts a need of a strict definition of stationarity by allowing nonstationary modelling and non-explicit description of domains when controlling fields describing important structural variables are provided as input in simulation. Non-stationarity is represented by auxiliary variables and its description and rigorous validation should be an integral and part of the modelling process.

5.2.3 **Lithology modelling**

Where CAD-based techniques for modelling lithologies are not viable or are time-consuming and variogram-based techniques fail, MPS allows an effective alternative approach. Simulation of the lithological interface can be adopted as has been proposed in the workflow. For Geita Hill, it has several advantages:

- On a macro scale, the folded lithological interface is a proxy for the location of brittle shearing;
- Proximity to the lithological contact within the folded hinge zone is a direct influence and controlling variable on the tenor of mineralisation in the follow-up simulation of gold due to competency contrast between the two rock types and sulphidation reaction in the BIF host-rock. The potential field created from the lithological contact serves as a non-stationarity descriptor to guide the gold grade simulation;

- A model of lithologies will enhance the process of the Mineral Resource estimation by an explicit incorporation of densities, with an option for stochastic representation. Density is an important primary variable of the Mineral Resource evaluation process and;
- Risk assessment related to geological uncertainty is made possible.

5.2.4 Mineral Resource estimation

The current research enabled gold grade simulation, as well as classification and quantification of a Mineral Resource with a novice MPS tool which has not been done for mineral deposits before.

The hypothesis of this research was to quantify an “Inferred” Mineral Resource using MPS simulation to exceed insignificantly the limits of 90% confidence with 15% error for a volume equivalent to annual mining production tested for a specified drill spacing pattern, by a chosen underground mining method. Only 20 realisations were produced, and it can be considered insufficient for conclusive statements. The analysis of the 20 realisations showed a spread beyond the desired 15% error in metal (between -19.5% and 16.2% error for the 90% confidence interval), and more offending in tonnes (between 31.1% and -3.4% error). The grade spread was biased negatively, between -16.7% and -8.6% error. These results fall outside of the requirements commonly adopted in the industry for the “Indicated” Mineral Resource, however, they are sufficiently robust to classify the material into the “Inferred” Mineral Resource. Further work can be suggested to improve the reproduction of the mineralised morphologies while utilising rigorous validation tests.

The MPS approach showed acceptable results which can be specifically useful for defining a lower confidence Mineral Resource where explicit and truthful mineralisation shapes are required for high-level underground mining engineering.

Validation of the simulated categorical variables proportions (lithologies and the mineralised shear zone) produced very narrow spread from the

conditioning data proportions and indicate a potential for broad applicability in the Mineral Resource classification and estimation. Employing the DS for continuous variables simulation and the Mineral Resource estimation, however, requires the development of rigorous validation routines and automated parameter tuning algorithms.

5.2.5 Non-parametric approach

MPS proposes a non-parametric approach to the description of natural phenomenon. This feature allows the algorithm to deal with multi-dimensional multi-variate datasets.

5.2.6 Reducing risk in new areas with similar settings

The methodology suggested in this research can be used in similar Greenstone belt environments.

The study aimed at establishing practical ways of incorporating valuable geological information collected at different stages of the project into a ‘true’ model. A library of such models can be useful for decision making when moving into higher-risk environment with the ultimate goal of reducing uncertainty between a forecast and production.

MPS can be used in regional exploration, when ‘pattern matching’ from a library of ‘true’ models is performed based on the similarity of tectonic and mineralisation environment. This method will be suitable to identify potentially prospective regions due to its ability to propagate patterns or lineations into sparsely informed areas, while having a strong ability for multivariate conditioning, such as geophysical datasets or hyperspectral imagery for mineral assemblages.

The research demonstrated the importance of developing an integrated understanding of a genetic model while making use of abundant data. Rather than being discarded with the depletion of the deposit, the knowledge is assimilated and stored in a library that can be extracted later when similar deposits are believed to have been encountered to explore

possible geological realities. Emphasis is placed on the importance of understanding geology; the establishing of the ‘prior’ model is given squarely into the hands of geologists rather than mathematicians.

While kriging still provides the best unbiased estimate and is appropriate to use with dense data such as production drilling for the parcel destination, it is subject to inaccuracies with sparse data. It cannot adequately populate the modelling space by extrapolating in paucity of data, more so in non-stationary settings.

The advantage of the proposed methodology in presence of inherent geological non-stationarity is that the ‘true’ model can be grown continuously as the orebody is being mined, and the geological understanding of the environment incorporated into the live ‘true’ model.

5.2.7 **Machine Learning**

The workflow proposed in this study can lend itself to the adaptation of modern developments in artificial intelligence and machine learning to mining. The framework is similar to supervised machine learning approach (Dey, 2016; Samson, 2019) – available datasets are used to find predictive relationships by training and testing the quality of prediction vs the ‘true’ model. When the simulation grid covers both, the testing and the ‘unknown’ portions of the deposit, the ensemble of realisations with satisfactory reconciliation against the ‘true’ model that successfully passed the validation routine is retained and can be applied to ‘unknown’ parts of the deposit which satisfy some qualifying rules. With multi-variate relationships established, subjective decisions on stationarity domaining can be minimised, and the method is made automated. The role of a geologist would be to understand structural controls and mineral assemblages and translate the understanding into a digital model – variables and relationships between them.

5.2.8 General

With regards to the gaps identified in the objective of this research the following has been addressed with the MPS workflow:

- Demonstrated better continuity of tabular geometries;
- Reproduction of finely intercalated and folded lithological geomorphology honouring structural potential field. Structural architecture of the deposit in shortage of data can be modelled by utilising an elementary conceptual TI in conjunction with structural data;
- Modelling of complex spatial structures, beyond two-point statistics, which are a result of multiple structural controls on mineralisation, confirmed by boundary analysis and simulated high-grade shoots honouring the potential field zonation;
- Reproduction of multivariate relationships has been tested with regards to boundary analysis and reproduction of gold grade statistics in the two lithologies;
- Good performance of the model with regards to expected confidence levels in the presence of sparse data;
- Reproduction of overall global statistics as indicated by drillhole proportions of categorical variables such as lithologies or mineralised material;
- Partial updates of the model are possible, where informed or previously simulated areas are ‘frozen’ and used as conditioning data and/or a training image for the follow-up simulation. The case study demonstrated a smooth transition between the ‘frozen’ and modelled areas with appropriate honouring of the spatial structures;
- Conditioning by secondary exhaustively known variables has been done to describe non-stationarity: the previously simulated lithologies and potential field of the contact;

- The stochastic framework allowed to do uncertainty assessment. It has, however, been limited to a global valuation within the simulation volume.

The main advantage of the method is a deviation from pure mathematical constructs to model reality which ensures incorporation of a geologist in the complete modelling process to the final product.

Definition of the model at fine resolution allows mining engineers to test different mining scenarios at any cut-off and mining unit size without a need to produce change-of-support models for the specified cut-off grades and mining unit sizes.

5.3 Limitations

The author admits that the simulation results are far from perfection and there is considerable space for improvement. More rigorous validation routines should be applied to test the appropriateness of simulated results. The assessed uncertainty is likely understated as no parameter uncertainty testing took place. The proposed method and parameters might be valid for the orebody under consideration but do not guarantee a successful implementation at another settings. Fine-tuning and expert knowledge are required.

5.4 Ethics

The ethical issues relate to the disclosure of any sensitive information pertaining to AngloGold Ashanti.

5.5 Scope for future work

While many different features of the Direct Sampling and MPS, in general, were covered in this research, there are some that can be explored in future. A few aspects are described below.

5.5.1 **Improvements to the proposed workflow**

Suggestions for improvements to the workflow have been scattered through the thesis and are summarised below:

- An enhancement to the proposed lithological modelling workflow would be to allow explicit conditioning by the introduction of an additional continuous variable, the true thickness of the logged lithological units. An elementary training image would need to contain this variable per lithological unit. A further possible enhancement would be to employ dynamic calculations eliminating a need for an explicit elementary training image;
- Usage of an elementary training image in simulation of mineralised zone and lithology showed to be effective, however, irregularity needs to be introduced into the shapes. This would prevent simulation of tabular and bulky morphologies which inevitably resulted in overstated gold mineralisation;
- To introduce more variability in the simulated realisations employing elementary TIs, simulation with dip and azimuth fields tolerances can be attempted in future. Tolerance can be input as an exhaustively informing grid with the variable value of 0 along the drillhole traces and increasing values away from drillholes;
- During the simulation of the mineralised zone, the potential field of the lithological contact was supplied as an auxiliary variable due to dependence of the intensity of mineralisation on a local scale from the folding architecture. Tolerance of 0.05 of the potential field was used which translates to width of about 10-15 m along GHSZ. More testwork can be undertaken in future on the sensitivity of this value;
- Using a variable representing the distance to “braided” shear splays was attempted in the current research to no success. An attempt can be made in future to further help localise the gold simulation. This variable can be useful for guiding simulations if orientation readings of the brittle structures can be logged reliably in the diamond core. In this case,

another step can be introduced in the workflow after lithology simulation – simulating ‘braided’ shears splays;

- Simulation with connectivity constraints has not been attempted. It is interchangeable with the idea expressed in the previous paragraph when mineralised structures are identifiable from diamond core. The appropriate geologically truthful connectivity paths can be simulated preceding gold mineralisation;
- Block conditioning has been attempted for cuboidal blocks with long axes being aligned north-south. Sharp discontinuities and artefacts in the transition of simulated geometries have been observed. It might be improved by employing non-cuboidal blocks, e.g. cylindrical around drillholes, or overlapping blocks. The block geometry can be rotated locally to make the elongated axes of the cuboidal blocks orthogonal to the plane of the shear zone;
- Simulation with a new approach in the DS, giving more weight to the conditioning nodes can be attempted;
- Further investigation can be beneficial into selection of gold indicator thresholds (the number of categories and the choice of cut-off values) to guide simulation of the continuous gold variable. Important considerations are relevance for mining and sufficient volumes of geobodies of multiple-indicator node clusters to get representative neighbourhoods during simulation;
- It will be beneficial to introduce an automated routine to define non-stationarity. Rather than relying on visual analysis, trend analysis for gold grade average can be used in relation to the potential field values to apply zonation and use it in validation;
- Further attempts can be undertaken to achieve gold simulations with improved morphology of patterns. The most significant parameter is the number of conditioning nodes. In the performed gold simulation, it has been set to 100, and might be insufficient for a 3D dataset;
- The framework can be used to explicitly model planned and unplanned dilution, as well as to assist in deciding on an appropriate cut-off grade;

- There are validation tools proposed by different authors that have not been attempted in this research and there is a scope to undertake it in future. For instance, De Iaco and Maggio (2011) suggested to use physically measurable criteria of simulated categorical geobodies: the number of bodies per unit volume, the global volume fraction, the direction of maximum variability, the mean volume or mean area, and the mean value of the area to volume ratio. Boisvert, Pyrcz and Deutsch (2008b) give examples of utilising metrics beyond the first and second order statistics, such as trends, the multiscale histograms, the multiple point density function, and statistics of the missing bins in the multiple point density function;
- Uncertainty in density has not been evaluated directly, only through stochastic modelling of lithologies. Further work can be done in future in this regard if a sufficient number of density measurements are available;
- The simulation was undertaken in the western portion of the orebody. The eastern part has lower tenor of mineralisation and complex curvilinear morphology of the ore shoots which seem to be less dependent on lithological contact (intercalation is finer here) but rather a function of the intersection of individual shear splays. Another approach would need to be adopted here. Individual shear splays might need to be simulated first and an auxiliary variable representing the distance from shears intersection to be considered when conditioning gold simulations. Morphological scaling property of the DS algorithm can also assist when simulating in areas with similar geological controls but differing size of geobody morphologies;
- The exhaustively known ‘true’ model of the deposit in the current work was successfully created using production drilling. In situations when no abundant 3D data is available to reliably interpolate the important variables, simulations can be performed using 2D training images (Comunian et al, 2011b). Outcrop maps, sectional or level plan interpretations are some examples of such 2D training images. The application can be beneficial when describing complex multivariate

relationships, such as alteration haloes, mineralisation assemblages or complex structurally controlled weathering patterns;

- It can also be valuable if MPS is used when simulating for exploration potential identification. In this case, training images of geophysical data with interpreted structures and known orebodies can be used in input with auxiliary field of geophysical data to guide simulation over SG.
- Using the derivative of the potential field one can 'train' the model to improve the prediction of shearing controlling mineralisation.
- No localised risk assessment took place. It can add a valuable dimension to the applicability of the method in mining engineering. Some examples include 3D maps of the most likely-to-occur lithology type, entropy of lithological simulations, E-type mean maps of gold simulations, probability maps of exceeding a specified grade cut-off within a given error threshold (Deutsch & Journel, 1998; Deutsch, 2013; Rossi & Deutsch, 2015). Possible applications of it for mining engineering are: spatial classification of the Mineral Resource and Ore Reserve, risk assessment of the Ore Reserve tonnes, grade and metal per stope, translating into the mine scheduling.
- The stochastic multivariate 3D model can be utilised in geotechnical and hydrogeological engineering fields. For instance, the difference in angle between the attitude of the shearing and the lithological contact or bedding planes which has been estimated into the 'true' model can be used to delineate different geotechnical domains and locate hazardous areas associated with low angle differences in relation to the openings.

5.5.2 **Applications of the DS not attempted before**

Some suggestions for new utilisations of the DS are summarised below:

- Block data conditioning has been attempted when simulating the mineralised shear zone. It allows more control over the simulated fields and can be used when dealing with data at different support sizes. An example of it is ascribing geometallurgical properties to different

isometrically shaped domains (similar to the treatment of density in the current research);

- In future, it might be worthwhile to attempt using a patch-based modification of the DS when simulating multiple-indicator gold variable. Afterwards, for the continuous gold variable pixel-support simulation can be reverted to due to recorded difficulty of conditioning continuous variables to hard data in patch-based methods;
- The DS can be applied to modelling deposits with strong stratigraphic control on mineralisation, which pose difficulties in cases of complex brittle structural framework complicated by displacement and late diluting intrusions;
- The algorithm can be used in 1D settings while maintaining multivariate framework. Examples are plant recovery performance as a function of feed with different ore type mixtures.

5.5.3 General suggestions for enhancements in the DS

The DS algorithm can benefit from further development work; a few areas are suggested below:

- The DS proposes a wide range of parameters to guide simulations. It can be very time-consuming to arrive at a well-performing set of parameters. Automatic tuning of the DS parameters has been investigated before (Dagasan et al., 2018), however, there is space for further work;
- To be able to impose local anisotropy field for ascribing increased weight to nodes in the vicinity of conditioning data;
- As mentioned by Mariethoz (2009), similarity distances can be adapted to be angle-invariant. They would be computed by scanning the training image with rotated data events. A pattern containing a feature aligned north-south could match a pattern containing a feature oriented east-west. This would present an alternative to the method adopted in the research of utilising the potential field of the lithological contact and can

be extremely useful. It can enable simulation of patterns of folded or rotated geometry without explicit unfolding.

5.5.4 General suggestions for MPS applicability

Some general suggestions for the applicability of MPS in geosciences reflections for further

- Modelling of the shear zone performed by digitising in the current research can be automated by using a derivative of the potential field of the lithological contact to define the S-fold hinge;
- When generating a training image for continuous variables with poor quality data, artefacts can occur that are not desirable for propagation into the simulated models. In this case, it would be good to have an algorithm to minimise such artefacts – isolated clusters of nodes not connected to the mineralisation. A similar approach has been available as part of GSLib software library for post-processing and noise removal in realisations (Deutsch & Journel, 1998);
- MPS framework supports simultaneous usage of multiple training images. Currently, professional expertise and knowledge of the deposit are required to allocate appropriateness of each training image across the volume to be modelled. There is space for designing automated methods that can assist in quantifying such a decision in a geological environment;
- In cases where libraries of historically mined deposits are available, the training images can be categorised by tectonic settings and ore genesis. Should a TI be considered representative of the environment but with different tenor of mineralisation, the simulation can be performed with rescaling of the model;
- The process of choosing the variables relevant to non-stationarity would benefit from supporting visual analysis with more rigorous numerical analysis;
- Visual comparison is still one of the most popular and trusted methods for primary validation of MPS simulations. There is scope for the

development of automated validation routines to mimic the process performed by the visual cortex in identifying the similarity of images beyond two-point statistics;

- The current research focused on the applicability of MPS to modelling a Mineral Resource. Thorough validation tools need to become commercially available before the workflow can be used to its best in the mining industry. It would be beneficial to advance this part of the research in future.

5.6 Conclusions

Multiple-point statistical framework and in particular the Direct Sampling algorithm propose a new paradigm for approaching geological modelling.

Richness of valuable multivariate multi-structural curvilinear information cannot be captured by two-point covariance models. Parametric description quickly becomes intractable with the addition of new variables and underperforms in the presence of structures beyond two points. With the increasing richness and complexity of data, coupled with the advances in storage capacity and data acquisition methods, the dimensionality of data captured about natural phenomena and complexities of engineering decisions are increasing. It calls for advanced non-parametric algorithms to describe the multivariate multi-dimensional complexity, adaptable to machine learning and artificial intelligence applications.

Some of the advantages of the proposed workflow are re-iterated below:

- MPS only relies on a representative ‘true’ model that reflects the spatial variability of a deposit. Considerable time and effort are required to convey specialists’ understanding of a deposit into a truthful TI, in orders of magnitude bigger than a variogram model would require. In mining, we are in a privileged position to sample a deposit exhaustively before extraction and utilise this knowledge to reduce risk when advancing into the unknown. The proposed

framework does not suggest abandoning a classical geostatistical approach and it does not stipulate that we must have access to an exhaustively informed volume before building a resource model. It, rather, reflects on the fact that we frequently have access to abundant production data and we need to learn from it better than what has been done before, specifically when having to make costly engineering decisions, based on plausible morphologies of mineralisation in highly selective mining environment. The end must justify the means;

- Running the DS in reconstructive mode allows to have live models, where sparsely informed portions of the deposit get progressively updated. Data from historically mined parts of a deposit rather than retiring can get added to the 'true' model to be used as a source of patterns for further simulation. Gradually the deposit transforms into the 'true' model. This process forms an integrated approach to machine-learning from the mined-out areas
- Diverse tools to deal with non-stationarity would allow using libraries of 'true' models in environments believed to be similar;
- Definition of planned and unplanned mining dilution becomes more transparent with MPS;
- Multiple geological realisations allow to assess uncertainty globally and locally. They can be used in Mineral Resource classification and in stochastic engineering to define the most risk-free life of mine design and mine schedules by providing better basis for optimisation under different engineering constraints (Rossi and Deutsch, 2015). 3D maps showing the probability of different mining volumes to be mined can be used to guide the mining engineer in scheduling, by combining high risk mining panels or stopes with high metal content variability with high confidence ones. A conditional loss function, which is a mathematical expression that attaches an economic value to each poor decision and the associated misclassification cost, can be established stochastically for each mining unit. An ideal full

assessment of risk from an ensemble of realisations can include a complete mine planning and scheduling for each realisation, however it is still intractable in mining engineering (Deutsch, 2013). It can be substituted with processing of a few extreme and the most representative realisations. The ranking of realisations can be done using the tonnes, grade and metal at a specified cut-off grade, or any other engineering criteria of importance. Areas where additional drilling is required can be outlined based on the high local uncertainty in the ensemble of realisations.

- A possibility of a new paradigm for public reporting of the Mineral Resource, where the average estimated figures are quoted using the mean values for the tonnes, grade and metal content obtained among the final ensemble of realisations, as well as the worst and the best scenarios. Similarly, the Ore Reserves can be given based on the final most risk-free mine design established after processing the ensemble of realisations, with the associated average, the best and the worst-case scenarios for the tonnes, grade and metal.

There are other advantages of using MPS: formulating and testing mathematically a version of reality is not inviting to a geologist. With conventional geostatistical methods, we rely on the performance of statistical parameters in validation. However, one of the main validation criteria is still a human eye as it serves as a proxy for spatial statistical validation. Adoption of this approach is natural with MPS. As noted by Journel (2004:227): “A training image can be refuted by an expert geologist, the same cannot be said about a geologically non-significant variogram model”.

The critical issue in subsurface numerical modelling is the reproduction of geological features and the ability to convey expert knowledge about the phenomenon. Due to the visual nature of a ‘prior’ model definition, the MPS is suitable for the involvement of multi-disciplinary specialists’ teams and

can provide a platform for integration of diverse information in a numerical model.

The DS can be applied to other big data problems. The multi-dimensionality of data, in this case, can be comprised of spatial element as well as temporal.

Different branches of science experience a move towards image processing routines. Some useful similarities and advances have been learnt from texture synthesis which objective is to produce simulated morphologies similar to a reference image (Arpat, 2005; Arpat & Caers, 2007; Mahmud et al., 2014; Efros & Freeman, 2001; Mariethoz & Lefebvre, 2014).

Parallels can be drawn, and useful features adopted from medical imaging in radiology which has been advancing towards machine learning and AI in giant steps since 1960s (Hosny, Parmar, Quackenbush, Schwartz & Aerts, 2018). Some of the AI applications are image segmentation algorithms which aim at simplification of an image into homogeneous and easier to analyse gatherings of pixels (Pal & Pal, 1993). These tools in medicine are being perfected to automatically recognise complex patterns in imaging data based on quantitative assessments. These algorithms can be useful in MPS simulation validation, identification of exploration targets, and identification of similar geological settings in a library of geological images.

The research posed more questions than answers, however, in view of the current trend towards machine learning and artificial intelligence, MPS and the DS, in particular, provide an applicable framework to automate the process towards geological modelling and the Mineral Resource evaluation. As with all other geostatistical methods, MPS should not be approached as a magical black box. It can only yield good results in the hands of an experienced modeller.

This thesis was being finalised during the global outbreak of COVID-19. As on the 31st of March 2020 there have been 786,854 cases and 37,837 death reported (Worldometers, 2020). While there are different parametric

mathematical models to describe the spread of a virus, there is a lot of ambiguity in applicability of these models as many parameters are at play to project fatality rate or contagion rate (Chan, 2020): country demographics, culture differences, mobility, quality of healthcare, quality of life affecting the immunity, climate, prevailing dietary preferences to name just a few. Further complexity of any projections comes from long incubation periods causing delays in outbreaks. A combination of parametric and non-parametric approaches can be beneficial to categorise and pool the different domains and arrive at projections.

REFERENCES

Abdollahifard, M., Mariethoz, G., Mohammadi, H. (2019) “Investigating extreme scenarios with multiple-point geostatistics and variance maximization”. Stochastic Environmental Research and Risk Assessment, Vol. 33, no11-12.

Alford, C., Brazil, M., Lee, D. (2007) “Optimisation in Underground Mining”. In: Weintraub A., Romero C., Bjørndal T., Epstein R., Miranda J. (eds) Handbook Of Operations Research In Natural Resources. International Series In Operations Research amp; Mana, vol 99. Springer, Boston, MA.

Allard, D., D’Or, D., Froidevaux, R. (2011) “An efficient maximum entropy approach for categorical variable prediction”. European Journal of Soil Science, Vol. 62, no3, pp 381–393.

Armstrong, M., Galli, A. G., Loc’h, G. L., Geoffroy, F. & Eschard, R. (2003) “Plurigaussian simulations in geosciences”, Berlin, Springer.

Arpat, B. (2005) “Sequential simulation with patterns”. PhD dissertation, Stanford University, USA.

Arpat, B., Caers, J. (2007) “Conditional simulations with patterns”. Mathematical Geology, Vol. 39, no2, pp 177–203.

Barnes, J., Gossage, B. (2014) “The do’s and don’ts of geological and grade boundary models and what you can do about it”. In: Mineral Resource and Ore Reserve Estimation, The AusIMM Guide to Good Practice, Second Edition, Monograph 30. Chapter 3, pp 137-144.

Boisvert, J., Leuangthong, O., Ortiz, J., Deutsch, C. (2008a) “A methodology to construct training images for vein-type deposits”. Computers & Geosciences, Vol. 34, pp 491–502.

Boisvert, J., Pyrcz, M., Deutsch, C. (2008b) “Multiple point metrics to assess categorical variable models”. Natural Resources Research, Vol. 19, no3, pp 165-175.

Borg, G., Rittenauer, A (2000) “Syn- and epigenetic sulphides in Archean BIFs of NW-Tanzania and their significance to gold mineralization”. Applied mineralogy. Rotterdam, Balkema. pp 263–266.

Boucher, A. (2007) “Algorithm-driven and representation-driven random function: a new formalism for applied geostatistics”. Stanford Centre for Reservoir Forecasting, Palo Alto, Calif.

Boucher, A. (2009) “Considering complex training images with search tree partitioning”. Computers and Geosciences, Vol. 35, pp 1151–1158.

Brayshaw, M., Nugus, M., Robins, S. (2010) “Reassessment of controls on the gold mineralisation in the Nyankanga deposit – Geita Gold Mines, North-Western Tanzania.” 8th International Mining Geology Conference, Queenstown, New Zealand, August, 2011.

Caers, J., Journel, A. (1998) “Stochastic reservoir simulation using neural networks trained on outcrop data”. SPE Annual Technical Conference and Exhibition, no49026, pp 321–336.

Caers, J. (2000) “Adding local accuracy to direct sequential simulation”. Mathematical Geology, Vol. 32, pp 815–850.

Caers, J., Zhang, T. (2002) “Multiple point geostatistics: a quantitative vehicle for integrating geologic analogues into multiple reservoir models”. Stanford University Press, Stanford Center for Reservoir Forecasting, Stanford.

Caers, J., Strebelle, S., Payrazyan, K. (2003) “Stochastic integration of seismic data and geologic scenarios: a West Africa submarine channel saga”. The Leading Edge, Vol. 22, no3, pp 192-196.

Caers, J. (2011) “Modelling Uncertainty in the Earth Sciences”. John Wiley and Sons.

Carle, S., Fogg, G. (1996) “Transition probability-based indicator geostatistics”. Mathematical Geology, Vol. 28, no4, pp 453-477.

Carr, J., Beatson, R., Cherrie, J., Mitchell, T., Fright W., McCallum, B., Evans, T. (2001) “Reconstruction and representation of 3D objects with Radial Basis Functions” Proceedings of the 2001 SIGGRAPH Conference, Los Angeles, California, USA, pp 67-76.

Chan, C. (2020) Covid-19 statistics: Exactly how deadly is the coronavirus? INTERNET. https://www.thestar.com.my/lifestyle/health/2020/03/12/covid-19-statistics-do-we-really-know-how-deadly-this-virus-is. Cited 31 March 2020.

Chugunova, T., Hu, L. (2008). “Multiple point simulations constrained by continuous auxiliary data”. Mathematical Geosciences, Vol. 40, no2, pp 133–146.

Comunian, A., Renard, P., Straubhaar, J., Bayer, P. (2011a) “Three-dimensional high resolution fluvio-glacial aquifer analog – Part 2: Geostatistical modelling”. Journal of Hydrology, Vol. 405, pp 10-23.

Comunian, A., Renard, P., Straubhaar, J. (2011b) “3D multiple-point statistics simulation using 2D training images”. Computers & Geosciences, Vol. 40, pp 49–65.

Cowan, J., Beatson, R., Ross, H., Fright, W., McLennan, T., Evans, T., Carr, J., Lane, R., Bright, D., Gillman, A., Oshust, P., Titley, M. (2003) “Practical implicit geological modelling”. 5th International Mining Geology Conference Bendigo, Vic, 17 - 19 November 2003.

Cowan, J. (2009). Pers.comm.

Cowan, J. (2012) “The deposit model paradox”. AIG Structural Geology and Resources Symposium, 2012, Kalgoorlie. Australian Institute of Geoscientists Bulletin, Vol. 56, pp 48–49

Cowan, J. (2014) “X-ray plunge projection - understanding structural geology from grade data”. AusIMM Monograph 30: Mineral Resource and Ore Reserve Estimation — The AusIMM Guide to Good Practice, second edition, pp 207-220.

Cowan, E. (2017) “Why can’t we interpret near-mine drilling data effectively?” In: J. Vearncombe, J. (ed.) Mineral Exploration 2017 – Extended Abstracts Volume. Australian Institute of Geoscientists Bulletin.

Cowan, J. (2018). Pers.comm.

Dagasan, Y., Renard, P., Straubhaar, J., Erten, O., Topal, E. (2018) “Automatic Parameter Tuning of Multiple-Point Statistical Simulations for Lateritic Bauxite Deposits”. Minerals, Vol. 8, pp 220.

Dagasan, Y., Erten, O., Renard, P., Straubhaar, J., Topal, E. (2019) “Multiple-point statistical simulation of the ore boundaries for a lateritic bauxite deposit”. Stochastic Environmental Research and Risk Assessment, 2019, Vol. 33, pp 865–878.

Daly, C. (2004) “Higher order models using entropy, Markov random fields and sequential simulation”. In: Leuangthong, O. & Deutsch, C. V. (eds.), Geostatistics Banff 2004, Kluwer Academic, Banff, AB, pp 215–224.

Daly, C., Knudby, C. (2007) “Multipoint statistics in reservoir modelling and in computer vision”. In: Petroleum geostatistics 2007, EAGE, Cascais, Portugal.

Datamine On-line Product Help (2019).

De Iaco, S., Maggio, S. (2011) “Validation techniques for geological patterns simulations based on variogram and multiple-point statistics”. Mathematical Geosciences, Vol. 43, no4, pp 483-500. ·

Deutsch, C., Journel, A. (1998) “GSLIB. Geostatistical Software Library and User's Guide”. 2nd Edition. New York: Oxford University Press.

Deutsch, C. (2013) Lectures on Citation Program in Applied Geostatistics, presented in Johannesburg, RSA. Centre for Computational Geostatistics. University of Alberta, Edmonton, Alberta, Canada.

De Vries, L., Carrera, J., Falivene, O., Gratacos, O., Luit, J. (2009) “Application of multiple-point geostatistics to non-stationary images”. Mathematical Geosciences, Vol. 41, no1, pp 29–42.

Dey, A. (2016) “Machine learning algorithms: A review”. International Journal of Computer Science and Information Technologies, Vol. 7, no3, pp 1174–1179.

Dimitrakopoulos, R., Mustapha, H., Gloaguen, E. (2010) “High-order statistics of spatial random fields: exploring spatial cumulants for modelling complex, non-Gaussian and non-linear phenomena”. Mathematical Geosciences, Vol. 42, no1, pp 65–99.

Dimitrakopoulos, R., Mustapha, H. (2011) “HOSIM: A high-order stochastic simulation algorithm for generating three-dimensional complex geological patterns”. Computers and Geosciences, Vol. 37, pp 1242–1253.

Efros, A., Freeman, W. (2001) “Image quilting for texture synthesis and transfer”. Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, August 2001, pp 341-346.

Glacken, I, Snowden, D. (2001) “Mineral Resource Estimation”. The AusIMM Guide to Good Practice. The Australasian Institute of Mining and Metallurgy, pp 189-198.

Gómez-Hernández, J., Wen, X. (1998) “Numerical modeling of macrodispersion in heterogeneous media: a comparison of multi-Gaussian and non-multi-Gaussian models”. Journal of Contaminant Hydrology, Vol. 30, no1, pp 129-156.

Guardiano, F., Srivastava, M. (1993) “Multivariate geostatistics: Beyond bivariate moments”. In: Soares, A. (ed.) Geostatistics-Troia, 1992, Vol. 1, pp 133-144.

Gumiaux, C., Gapais, D., Brun, J. (2003) “Geostatistics applied to best-fit interpolation of orientation data”. Tectonophysics, Vol. 376, pp 241–259.

Harding, A., Strebelle, S., Levy, M. (2005) “Reservoir facies modeling: New advances in MPS”. Geostatistics Banff, Vol. 2, pp 559–568.

Hashemi, S., Javaherian, A., Ataee-Pour, M., Tahmasebi, P., Khoshdel, H. (2014) “Channel characterization using multiple-point geostatistics, neural network, and modern analogy: a case study from a carbonate reservoir, Southwest Iran”. Journal of Applied Geophysics, Vol. 111, https://doi.org/10.1016/j.jappgeo.2014.09.015.

Honarkhah, M., Caers, J. (2010) “Stochastic simulation of patterns using distance-based pattern modelling”. Mathematical Geosciences, Vol. 42, pp 487–517.

Honarkhah, M., Caers, J. (2012) “Direct Pattern-Based Simulation of Non-stationary Geostatistical Models”. Mathematical Geosciences, Vol. 44, pp 651-672.

Hong J., Oh, S., Cho, S. (2018) “Multiple-point geostatistical simulation for mine evaluation with aeromagnetic data”. Exploration Geophysics, Vol. 49, pp 952–965.

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L., Aerts, H. (2018) “Artificial intelligence in radiology”. Nature Reviews Cancer, Vol. 18, no8, pp 500–510.

Hu, L., Chugunova, T. (2008) “Multiple point geostatistics for modeling subsurface heterogeneity: A comprehensive review”. Water Resources Research, Vol. 44, W11413.

Huysmans, M., Dassargues, A. (2009) “Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium)”. Hydrogeology Journal, Vol. 17, 1901, DOI 10.1007/s10040-009-0495-2.

Journel, A. (1985) “The Deterministic Side of Geostatistics”. Mathematical Geology, Vol. 17, no1, pp 1-15.

Journel, A., Alabert, F. (1990) “New method for reservoir mapping”. Journal of Petroleum Technology, Vol. 42, pp 212–218.

Journel, A., Deutsch, C. (1993) “Entropy and Spatial Disorder”. Mathematical Geology, Vol. 25, no3, pp 329–355.

Journel, A. (1997) “Deterministic geostatistics: a new visit”. Quantitative Geology and Geostatistics. International Geostatistical Congress, 1996, Wollongong, Australia. Kluwer Academic, pp 292-301.

Journel, A. (1999) “Conditioning geostatistical operations to nonlinear volume averages”. Mathematical Geology, Vol. 31, no8, pp 931–953.

Journel, A. (2002) “Combining knowledge from diverse sources: an alternative to traditional data independence hypothesis”. Mathematical Geology, Vol. 34, no5, pp 573–596.

Journel, A. (2004) “Beyond covariance: the advent of multiple-point geostatistics”. In: Leuangthong O., Deutsch C.V. (eds) Geostatistics Banff 2004. Quantitative Geology and Geostatistics, Vol. 14. Springer, Dordrecht

Journel, A., Zhang, T. (2006) “The necessity of a multiple point prior model”. Mathematical Geology, Vol. 38, no5, pp 591–610.

Kabete, J., Groves. D., McNaughton, N., Mruma, A. (2012) “A new tectonic and temporal framework for the Tanzanian Shield: implications for gold metallogeny and undiscovered endowment”. Ore Geology Reviews, Vol. 48, pp 88–124.

Lajaunie, C., Courrioux, G., Manuel, L. (1997). “Foliation fields and 3D cartography in geology: Principles of a method based on potential interpolation”. Mathematical Geology, Vol. 29, pp 571–584.

Lantuejoul, C. (2002) “Geostatistical Simulation: Models and algorithms”. Springer, Berlin.

Legendre Polynomials (2017), in Wikipedia: The Free Encyclopedia, Wikimedia Foundation Inc., viewed 9 March 2017,

https://en.wikipedia.org/wiki/Legendre_polynomials. Page last modified on 17 February 2017, at 13:45 (UCT).

Liu, Y., Harding, A., Gilbert, R., Journel, A. (2005) “Multiple point geostatistical simulation”. In: Leuangthong, O., Deutsch, C.V. (Eds.), Geostatistics Banff 2004. Springer, New York.

Liu, Y., Journel, A. (2009) “A package for geostatistical integration of coarse and fine scale data”. Computer and Geosciences, Vol. 35, no3, pp 527–547.

Mahmud, K., Mariethoz, G., Caers, J., Tahmasebi, P., Baker, A. (2014) “Simulation of Earth textures by conditional image quilting”. Water Resources Research, Vol. 50, no4, pp 3088–3107.

Mariethoz, G. (2009) “Geological stochastic imaging for aquifer characterization”. PhD thesis. University of Neuchâtel.

Mariethoz, G., Renard, P. (2010) “Reconstruction of incomplete data sets or images using Direct Sampling”. Mathematical Geosciences, Vol. 42, pp 245–268.

Mariethoz, G., Renard, P., Straubhaar, J. (2010) “The Direct Sampling method to perform multiple-point geostatistical simulations”. Water Resources Research, Vol. 46, W11536.

Mariethoz, G., Kelly, F. (2011) “Modeling complex geological structures with elementary training images and transform-invariant distances”. Water Resources Research, Vol.47, W07527.

Mariethoz, G., McCabe, M., Renard, P. (2012) “Multivariate spatio-temporal reconstruction of gaps for spatially continuous satellite based Retrievals”, Water Resources Research, Vol.48, W10507.

Mariethoz, G., Lefebvre, S. (2014) “Bridges between multiple point geostatistics and texture synthesis”. Mathematical Geosciences, Vol. 66, pp 66–80.

Mariethoz, G., Caers, J. (2015). “Multiple-point geostatistics: Stochastic modelling with training images”. First Edition. John Wiley & Sons, Ltd.

Mustapha, H., Dimitrakopoulos, R. (2010). “High-order stochastic simulation of complex spatially distributed natural phenomena”. Mathematical Geosciences, Vol. 42, pp 457-485.

Ortiz, J., Deutsch, C. (2004) “Indicator simulation accounting for multiple-point statistics”. Mathematical Geology, Vol. 36, pp 545-566.

Osher, S., Fedkiw, R. (2003) “Signed Distance Functions”. In: Level Set Methods and Dynamic Implicit Surfaces. Applied Mathematical Sciences, Vol. 153. Springer, New York.

Osterholt, V., Dimitrakopoulos, R. (2007) “Simulation of orebody geology with multiple-point geostatistics: application at Yandi channel iron ore deposit, WA and implications for resource uncertainty”. Orebody modelling and strategic mine planning: The Australasian Institute of Mining and Metallurgy, Vol. 14, pp 51–60.

Owusu, S., Dagdelen, K. (20119) “Critical review of Mineral Resource classification techniques in the gold mining industry”. Insights in Mining Science & Technology, Vol. 1, no3, pp 74-80.

Pal, N., Pal, S. (1993) “A review on image segmentation techniques”. Pattern Recognition, Vol. 26, no9, pp 1277-1294.

Paranhos, C. (2008). Pers.comm.

Parra, A., Ortiz, J. (2011) “Adapting a texture synthesis algorithm for conditional multiple-point geostatistical simulation”. Stochastic Environmental Research and Risk Assessment, Vol. 25, pp 1101–1111.

Peredo, O., Ortiz, J. (2011). “Parallel implementation of simulated annealing to reproduce multiple-point statistics”. Computers and Geosciences, Vol. 37, pp 1110–1121.

Pirot, G., Straubhaar, J., Renard, P. (2014) “Simulation of braided river elevation model time series with multiple-point statistics”. Geomorphology, Vol.214, pp 148-156.

Renard, P., Straubhaar, J., Caers, J., Mariethoz, G. (2011) “Conditioning Facies Simulations with Connectivity Data”. Mathematical Geosciences, Vol. 43, no8, pp 879–903

Renard, P., Allard, D. (2013) “Connectivity metrics for subsurface flow and transport”. Advances in Water Resources, Vol. 51, pp 168-196.

Renard, P., Straubhaar, J. (2018) Training in MPS simulation, presented in Johannesburg, RSA. University of Neuchâtel, Switzerland.

Rezaee, H., Mariethoz, G., Koneshloo, M., Asghari, O. (2013) “Multiple-point geostatistical simulation using the bunch-pasting direct sampling method”. Computers & Geosciences, Vol. 54, pp 293-308.

Rezaee, H., Asghari, O., Koneshloo, M., Ortiz, J. (2014) “Multiple-point geostatistical simulation of dykes: application at Sungun porphyry copper system, Iran”. Stochastic Environmental Research and Risk Assessment, Vol. 28, pp 1913–1927.

Rezaee, H., Marcotte, D. (2016) “Integration of multiple soft data sets in MPS through multinomial logistic regression: a case study of gas hydrates”. Stochastic Environmental Research and Risk Assessment, Vol. 31, pp 1727-1745.

Rossi, M., Deutsch, C. (2015) “Mineral Resource Estimation”. Springer Science+Business Media Dordrecht.

Rue, H., Martino, S., Chopin, N. (2009) “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. Journal of the Royal Statistical Society. Series B: Statistical Methodology, Vol. 71, pp 319–392.

SAMREC, 2016.

Samson, M. (2019) “Mineral Resource Estimates with Machine Learning and Geostatistics”. Master of Science Thesis, University of Alberta. December 2019.

Sanislav, I., Brayshaw, M., Kolling, S., Dirks, P., Cook, Y, Blenkinsop, T. (2017) “The structural history and mineralization controls of the world-class Geita Hill gold deposit, Geita Greenstone Belt, Tanzania”, Mineralium Deposita, Vol. 52, no2, pp 257–279.

Sanchez-Vila, X., Carrera, J., Girardi, J. (1996) “Scale effects in transmissivity”. Journal of Hydrology, Vol. 186, pp 1– 22.

Scheidt, C., Caers, J. (2009) “Representing spatial uncertainty using distances and kernels”. Mathematical Geosciences, Vol. 41, pp 397–419.

Silva, D. A. (2015). “Signed Distance Function Modeling with Multiple Categories”. In J. L. Deutsch (Ed.), Geostatistics Lessons. Retrieved from http://www.geostatisticslessons.com/lessons/signeddistancefunctions on the 4th of November 2019.

Soares, A. (2001) “Direct sequential simulation and cosimulation.” Mathematical Geology, Vol. 33, pp 911–926.

Srivastava, M. (2018) “The Origins of the Multiple-Point Statistics (MPS) Algorithm”. In: Daya Sagar B., Cheng Q., Agterberg F. (eds) Handbook of Mathematical Geosciences. Springer, Cham.

Sterk, R., de Jong, K., Partington, G., Kerkvliet, S., van de Ven, M. (2019) “Domaining in Mineral Resource Estimation: A Stock-Take of 2019 Common Practice”. Special report: RSC Mineral Intelligence, RIU Sydney Resources Round-up 2019. An analysis of public reporting on Australasian projects from Q1 2019.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., Besson, O. (2011) “An improved parallel multiple-point algorithm using a list approach”. Mathematical Geosciences, Vol. 43, no3, pp 305–328.

Straubhaar, J., Walgenwitz, A., Renard, P. (2013) “Parallel multiple-point statistics algorithm based on list and tree structures”. Mathematical Geosciences, Vol. 45, pp 131–147.

Straubhaar, J., Renard, P., Mariethoz, G. (2016) “Conditioning multiple-point statistics simulations to block data”. Spatial Statistics, Vol. 16, pp 53-71.

Straubhaar, J. (2019) “DeeSse User’s Guide”. The Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel.

Strebelle, S., Journel, A. (2001) “Reservoir modeling using multiple-point statistics”. In: SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 30 September–3 October 2001.

Strebelle, S. (2002) “Conditional simulation of complex geological structures using multiple-point statistics”. Mathematical Geology, Vol.34, pp 1–22.

Strebelle, S. (2003) “New multiple-point statistics simulation implementation to reduce memory and CPU-demand”. In: Proceedings to the IAMG 2003, Portsmouth, UK, pp 7–12.

Tahmasebi, P., Hezarkhani, A., Sahimi, M. (2012) “Multiple-point geostatistical modeling based on the cross-correlation functions”. Computational Geosciences, Vol. 16, pp 779–797.

Tahmasebi, P. (2018) “Multiple Point Statistics: A Review”. In: Daya Sagar B., Cheng Q., Agterberg F. (eds) Handbook of Mathematical Geosciences. Springer, Cham, pp 613-643.

Tjelmeland, H., Eidsvik, J. (2004). “Directional Metropolis: Hastings updates for posteriors with nonlinear likelihoods”. In: Geostatistics Banff 2004. Springer, Berlin, pp 95–104.

Tran, T. (1994) “Improving variogram reproduction on dense simulation grids”. Computers and Geosciences, Vol. 20, no7, pp 1161-1168.

Van der Grijp, Y., Minnitt, R. (2015) “Application of Direct Sampling multi-point statistic and sequential Gaussian simulation algorithms for modelling uncertainty in gold deposits”. SAIMM Journal, Vol. 115, pp 73-85.

Vargas, H., Caetano, H., Mata-Lima, H. (2008) “A new parallelization approach for sequential simulation”. In: Soares A, Pereira MJ, Dimitrakopoulos R (eds) geoENV VI geostatistics for environmental applications. Springer, Berlin, pp 489–496.

Verly, G., Postolski, T., Parker, H. (2014) “Assessing uncertainty with drill hole spacing studies: applications to mineral resources”. In: Proceedings Orebody Modelling and Strategic Mine Planning Symposium. AusIMM Melbourne, 2014, pp 109-118.

Wallis, J., Miller, T., Lerner, C., Kleerup, E. (1989) “Three-dimensional display in nuclear medicine”. IEEE Transactions on Medical Imaging, Vol. 8, no4, pp 297–303.

Worldometers, (2020) INTERNET. https://www.worldometers.info/coronavirus/. Cited 31 March 2020.

Zinn, B., Harvey, C. (2003) “When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields”, Water Resources Research, Vol. 39, no3, WR001146.

Zhang, T., Switzer, P., Journel, A. (2006) “Filter-based classification of training image patterns for spatial simulation”. Mathematical Geology, Vol. 38, pp 63–80.

Zhang, T., Bombarde, S., Strebelle, S., Oatney, E. (2006) “3D Porosity Modeling of a Carbonate Reservoir Using Continuous Multiple-Point Statistics Simulation”. Society of Petroleum Engineers Journal, September 2006, pp375-379.

BIBLIOGRAPHY

Mariethoz, G. (2009) “Geological stochastic imaging for aquifer characterization”. PhD thesis. University of Neuchâtel.

Mariethoz, G., Caers, J. (2015). “Multiple-Point Geostatistics: Stochastic Modelling with Training Images”. First Edition. John Wiley & Sons, Ltd.

Rossi, M., Deutsch, C. (2015) “Mineral Resource Estimation”. Springer Science+Business Media Dordrecht.

APPENDICES

5.7 Appendix A: Simulated lithology statistics

Datatype	*Cell count*		*Proportion*		*Difference vs conditioning data*	*Difference vs 'true' model*
	BIF	*Diorite*	*BIF*	*Diorite*	*BIF*	*BIF*
Training image	**615,847**	**187,542**	**76.66**	**23.34**		
Conditioning data	**519**	**505**	**50.68**	**49.32**		
'True' model	**127,261**	**137,714**	**48.03**	**51.97**		
Realisation 0	131,127	133,848	49.49	50.51	-2.4%	5.2%
Realisation 1	133,320	131,655	50.31	49.69	-0.7%	3.5%
Realisation 2	131,426	133,549	49.60	50.40	-2.1%	4.9%
Realisation 3	130,473	134,502	49.24	50.76	-2.8%	5.7%
Realisation 4	130,062	134,913	49.08	50.92	-3.2%	6.0%
Realisation 5	128,639	136,336	48.55	51.45	-4.2%	7.1%
Realisation 6	127,470	137,505	48.11	51.89	-5.1%	8.0%
Realisation 7	132,744	132,231	50.10	49.90	-1.2%	3.9%
Realisation 8	128,854	136,121	48.63	51.37	-4.1%	7.0%
Realisation 9	131,018	133,957	49.45	50.55	-2.4%	5.3%
Realisation 10	133,706	131,269	50.46	49.54	-0.4%	3.1%
Realisation 11	130,031	134,944	49.07	50.93	-3.2%	6.0%
Realisation 12	130,293	134,682	49.17	50.83	-3.0%	5.8%
Realisation 13	134,612	130,363	50.80	49.20	0.2%	2.4%
Realisation 14	133,317	131,658	50.31	49.69	-0.7%	3.5%
Realisation 15	132,813	132,162	50.12	49.88	-1.1%	3.9%
Realisation 16	133,051	131,924	50.21	49.79	-0.9%	3.7%
Realisation 17	132,764	132,211	50.10	49.90	-1.1%	3.9%
Realisation 18	133,609	131,366	50.42	49.58	-0.5%	3.2%
Realisation 19	134,537	130,438	50.77	49.23	0.2%	2.5%
Spread of uncertainty in simulations						
Mean	131,693	133,282	49.70	50.30	-1.9%	4.7%
5th percentile	128,581	130,434	48.53	49.23	-4.3%	2.5%
95th percentile	134,541	136,394	50.77	51.47	0.2%	7.2%

5.8 Appendix B: Simulated mineralised zone statistics

Datatype	*Cell count*		*Proportion*		*Difference vs conditioning data*	*Difference vs 'true' model*
	Waste	*Ore*	*Waste*	*Ore*	*Ore*	*Ore*
Training image	**595,100**	**208,289**	**74.07**	**25.93**		
Conditioning data	**662**	**362**	**64.65**	**35.35**		
'True' model	**173,723**	**91,252**	**65.56**	**34.44**		
Realisation 0	176,727	88,248	66.70	33.30	-5.8%	-3.3%
Realisation 1	176,226	88,749	66.51	33.49	-5.3%	-2.7%
Realisation 2	178,285	86,690	67.28	32.72	-7.5%	-5.0%
Realisation 3	179,728	85,247	67.83	32.17	-9.0%	-6.6%
Realisation 4	177,037	87,938	66.81	33.19	-6.1%	-3.6%
Realisation 5	176,921	88,054	66.77	33.23	-6.0%	-3.5%
Realisation 6	178,712	86,263	67.44	32.56	-7.9%	-5.5%
Realisation 7	175,671	89,304	66.30	33.70	-4.7%	-2.1%
Realisation 8	175,843	89,132	66.36	33.64	-4.8%	-2.3%
Realisation 9	177,655	87,320	67.05	32.95	-6.8%	-4.3%
Realisation 10	176,213	88,762	66.50	33.50	-5.2%	-2.7%
Realisation 11	176,251	88,724	66.52	33.48	-5.3%	-2.8%
Realisation 12	175,017	89,958	66.05	33.95	-4.0%	-1.4%
Realisation 13	175,937	89,038	66.40	33.60	-4.9%	-2.4%
Realisation 14	176,348	88,627	66.55	33.45	-5.4%	-2.9%
Realisation 15	176,673	88,302	66.68	33.32	-5.7%	-3.2%
Realisation 16	175,969	89,006	66.41	33.59	-5.0%	-2.5%
Realisation 17	176,017	88,958	66.43	33.57	-5.0%	-2.5%
Realisation 18	176,319	88,656	66.54	33.46	-5.4%	-2.8%
Realisation 19	176,479	88,496	66.60	33.40	-5.5%	-3.0%
Spread of uncertainty in simulations						
Mean	176,701	88,274	66.69	33.31	-5.8%	-3.3%
5th percentile	175,638	86,212	66.28	32.54	-8.0%	-5.5%
95th percentile	178,763	89,337	67.46	33.72	-4.6%	-2.1%

5.9 Appendix C: Simulated gold grade statistics in BIF within the mineralised zone

Datatype	*Count*	*Mean*	*Std. Dev.*	*Variance*	*Minimum*	*1st Quartile*	*2nd Quartile*	*3rd Quartile*	*Maximum*
Training image	**167592**	**1.47**	**1.90**	**3.62**	**0.30**	**0.48**	**0.86**	**1.71**	**71.32**
Conditioning data	**166**	**2.06**	**2.61**	**6.82**	**0.30**	**0.48**	**1.05**	**2.86**	**19.25**
True model	**42354**	**1.94**	**2.72**	**7.41**	**0.30**	**0.51**	**0.98**	**2.28**	**172.59**
Realisation 0	46541	2.08	2.30	5.28	0.00	0.52	1.22	2.96	67.13
Realisation 1	47299	1.97	2.32	5.39	0.00	0.50	1.06	2.68	67.13
Realisation 2	42420	2.15	2.21	4.88	0.00	0.62	1.35	3.01	35.24
Realisation 3	40992	2.23	2.25	5.05	0.00	0.63	1.42	3.19	35.24
Realisation 4	44767	1.89	2.06	4.26	0.00	0.50	1.12	2.61	35.24
Realisation 5	43748	1.77	1.93	3.72	0.00	0.49	1.03	2.41	35.24
Realisation 6	42716	2.19	2.20	4.84	0.00	0.63	1.42	3.12	35.24
Realisation 7	47956	2.04	2.29	5.26	0.00	0.52	1.14	2.88	67.13
Realisation 8	44826	2.06	2.32	5.37	0.00	0.50	1.09	2.93	35.24
Realisation 9	45493	1.98	2.18	4.74	0.00	0.51	1.13	2.77	49.71
Realisation 10	46289	1.76	1.93	3.71	0.00	0.50	1.04	2.37	41.51
Realisation 11	45004	1.88	2.13	4.54	0.00	0.48	1.04	2.62	67.13
Realisation 12	47032	1.90	2.14	4.56	0.00	0.51	1.06	2.63	67.13
Realisation 13	44057	2.26	2.13	4.54	0.00	0.69	1.59	3.25	49.06
Realisation 14	47521	1.87	2.02	4.10	0.00	0.51	1.11	2.59	38.34
Realisation 15	47213	1.93	2.05	4.19	0.00	0.51	1.16	2.77	49.71
Realisation 16	47016	1.88	1.99	3.98	0.00	0.52	1.14	2.59	38.34
Realisation 17	46563	1.84	1.95	3.81	0.00	0.51	1.11	2.55	35.24
Realisation 18	44897	1.81	1.91	3.66	0.00	0.49	1.07	2.51	32.88
Realisation 19	46878	1.92	2.05	4.22	0.00	0.51	1.14	2.71	36.27
Mean of realisations		**1.97**	**2.12**	**4.51**	**0.00**	**0.53**	**1.17**	**2.76**	**45.91**

5.10 Appendix D: Simulated gold grade statistics in diorite within the mineralised zone

Datatype	*Count*	*Mean*	*Std. Dev.*	*Variance*	*Minimum*	*1st Quartile*	*2nd Quartile*	*3rd Quartile*	*Maximum*
Training image	**40697**	**0.94**	**1.32**	**1.75**	**0.30**	**0.40**	**0.58**	**0.96**	**57.64**
Conditioning data	**196**	**1.04**	**1.47**	**2.17**	**0.30**	**0.38**	**0.52**	**0.91**	**11.04**
True model	**48898**	**1.03**	**1.44**	**2.08**	**0.30**	**0.40**	**0.59**	**1.06**	**49.13**
Realisation 0	41707	0.85	1.09	1.20	0.00	0.40	0.57	0.87	40.23
Realisation 1	41450	0.79	0.82	0.68	0.00	0.40	0.57	0.86	23.71
Realisation 2	44270	0.77	0.70	0.49	0.00	0.43	0.59	0.86	21.50
Realisation 3	44255	0.81	0.83	0.69	0.00	0.43	0.59	0.87	21.50
Realisation 4	43171	0.79	0.86	0.73	0.00	0.40	0.57	0.85	21.50
Realisation 5	44306	0.77	0.81	0.66	0.00	0.40	0.56	0.84	29.51
Realisation 6	43547	0.76	0.69	0.48	0.00	0.42	0.59	0.86	21.50
Realisation 7	41348	0.80	0.94	0.89	0.00	0.40	0.57	0.86	42.66
Realisation 8	44306	0.80	0.89	0.80	0.00	0.41	0.58	0.86	25.83
Realisation 9	41827	0.78	0.80	0.64	0.00	0.40	0.57	0.86	21.50
Realisation 10	42473	0.78	0.80	0.65	0.00	0.41	0.57	0.86	16.71
Realisation 11	43720	0.77	0.83	0.68	0.00	0.40	0.56	0.83	29.51
Realisation 12	42926	0.75	0.78	0.61	0.00	0.40	0.57	0.83	25.83
Realisation 13	44981	0.82	0.82	0.68	0.00	0.44	0.62	0.92	25.83
Realisation 14	41106	0.82	0.88	0.78	0.00	0.41	0.58	0.89	25.83
Realisation 15	41089	0.83	0.93	0.87	0.00	0.40	0.58	0.89	42.66
Realisation 16	41990	0.85	1.09	1.18	0.00	0.40	0.57	0.89	40.23
Realisation 17	42395	0.85	0.93	0.86	0.00	0.41	0.59	0.91	29.51
Realisation 18	43759	0.80	0.79	0.62	0.00	0.41	0.59	0.89	21.50
Realisation 19	41618	0.80	0.84	0.70	0.00	0.41	0.58	0.87	25.83
Mean of realisations		**0.80**	**0.86**	**0.74**	**0.00**	**0.41**	**0.58**	**0.87**	**27.64**

5.11 Appendix E: Simulated gold grade statistics in the mineralised zone, lithologies undifferentiated

Datatype	*Count*	*Mean*	*Std. Dev.*	*Variance*	*Minimum*	*1st Quartile*	*2nd Quartile*	*3rd Quartile*	*Maximum*
Training image	**208289**	**1.36**	**1.82**	**3.30**	**0.30**	**0.46**	**0.78**	**1.55**	**71.32**
Conditioning data	**362**	**1.51**	**2.13**	**4.55**	**0.30**	**0.41**	**0.64**	**1.57**	**19.25**
True model	**91252**	**1.45**	**2.18**	**4.76**	**0.30**	**0.44**	**0.71**	**1.53**	**172.59**
Realisation 0	88248	1.50	1.93	3.73	0.00	0.44	0.74	1.71	67.13
Realisation 1	88749	1.42	1.88	3.53	0.00	0.44	0.72	1.51	67.13
Realisation 2	86690	1.44	1.77	3.12	0.00	0.48	0.77	1.59	35.24
Realisation 3	85247	1.49	1.82	3.30	0.00	0.47	0.77	1.68	35.24
Realisation 4	87938	1.35	1.68	2.83	0.00	0.43	0.70	1.50	35.24
Realisation 5	88054	1.26	1.56	2.43	0.00	0.43	0.69	1.37	35.24
Realisation 6	86263	1.47	1.77	3.15	0.00	0.47	0.77	1.66	35.24
Realisation 7	89304	1.46	1.90	3.62	0.00	0.44	0.72	1.63	67.13
Realisation 8	89132	1.43	1.87	3.49	0.00	0.44	0.71	1.50	35.24
Realisation 9	87320	1.40	1.77	3.14	0.00	0.44	0.72	1.54	49.71
Realisation 10	88762	1.30	1.58	2.49	0.00	0.44	0.71	1.43	41.51
Realisation 11	88724	1.33	1.72	2.95	0.00	0.43	0.68	1.43	67.13
Realisation 12	89958	1.35	1.73	3.00	0.00	0.44	0.70	1.44	67.13
Realisation 13	89038	1.53	1.76	3.10	0.00	0.50	0.83	1.85	49.06
Realisation 14	88627	1.38	1.68	2.84	0.00	0.45	0.74	1.56	38.34
Realisation 15	88302	1.42	1.72	2.94	0.00	0.44	0.74	1.66	49.71
Realisation 16	89006	1.39	1.71	2.92	0.00	0.44	0.74	1.60	40.23
Realisation 17	88958	1.37	1.63	2.65	0.00	0.45	0.74	1.58	35.24
Realisation 18	88656	1.31	1.55	2.42	0.00	0.44	0.72	1.45	32.88
Realisation 19	88496	1.39	1.70	2.88	0.00	0.45	0.73	1.58	36.27
Mean of realisations		**1.40**	**1.74**	**3.03**	**0.00**	**0.45**	**0.73**	**1.56**	**46.00**

5.12 Appendix F: The Direct Sampling algorithm parameter file for lithology simulation

```
/* SIMULATION GRID (SG) */
81    94     102    // size    in each direction
  2.5    2.5     2.5  // spacing in each direction
52382.5  9196.25  1096.25  // origin

/* SIMULATION VARIABLES */
/* Number of simulation variable(s), and for each variable:
      variable name, output flag (0 / 1), and if output flag is 1: format string (as passed to fprintf).
   Write DEFAULT_FORMAT for format string to use default format.
   Example (with 3 variables):
      3
      varName1  1  %10.5E
      varName2  0
      varName3  1  DEFAULT_FORMAT
*/
1
LITH      1  DEFAULT_FORMAT

/* OUTPUT SETTINGS FOR SIMULATION */
/* Key word and required name(s) or prefix, for output of the realizations:
      - OUTPUT_SIM_NO_FILE:
           no file in output,
      - OUTPUT_SIM_ALL_IN_ONE_FILE:
           one file in output,
           requires one file name
      - OUTPUT_SIM_ONE_FILE_PER_VARIABLE:
           one file per variable in output (flagged as 1 above),
           requires as many file name(s) as variable(s) flagged as 1 above
      - OUTPUT_SIM_ONE_FILE_PER_REALIZATION:
           one file per realization,
           requires one prefix (for file name) */
OUTPUT_SIM_ALL_IN_ONE_FILE
sim83_10.gslib


/* OUTPUT REPORT */
/* Flag (0 / 1), and if 1, output report file. */
1
sim83_10_report.txt

/* TRAINING IMAGE */
/* Number of training image(s) (nTI >= 1), followed by nTI file(s)
   (a file can be replaced by the string "_DATA_" which means that the
   simulation grid itself is taken as training image), and
   if nTI > 1, one pdf image file (for training images, nTI variables). */
1
ti2_lith.gslib

/* DATA IMAGE FILE FOR SG */
/* Number of image file(s) (n >= 0), followed by n file(s). */
1
bm_83_cond40.gslib

/* DATA POINT SET FILE FOR SG */
/* Number of point set file(s) (n >= 0), followed by n file(s). */
0

/* MASK IMAGE */
/* Flag (0: mask not used / 1: mask used) and if 1, mask image file
   (this image contains one variable on the simulation grid: flag (0 / 1)
   for each node of the simulation grid that indicates if the variable(s)
   will be simulated at the corresponding node (flag 1) or not (flag 0). */
1
bm_83_mask.gslib

/* HOMOTHETY */
/* 1. Homothety usage, integer (homothetyUsage):
        - 0: no homothety
        - 1: homothety without tolerance
        - 2: homothety with tolerance
   2a. If homothetyUsage == 1,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
```

```
   2b. If homothetyUsage == 2,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
0

/* ROTATION */
/* 1. Rotation usage, integer (rotationUsage):
        - 0: no rotation
        - 1: rotation without tolerance
        - 2: rotation with tolerance
   2a. If rotationUsage == 1,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
   2b. If rotationUsage == 2,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
1
1 bm_83_codipdir.gslib
1 bm_83_codip.gslib
0 0

/* CONSISTENCY OF CONDITIONING DATA (TOLERANCE RELATIVELY TO THE RANGE OF TRAINING VALUES) */
/* Maximal accepted expansion in both extremities of the range of values in training images
   for covering the conditioning data values; e.g. if this number is set to 0.05,
   the conditioning data values can be beyond the range of the values in the training images
   (in both extremities) of at most 5%; this separately applies to all variables for which
   the distance is absolute (not relative, see "relative distance flag" below) and the
   distance type is not 0 (see "distance type" below). */
0.05

/* NORMALIZATION TYPE (FOR VARIABLES FOR WHICH DISTANCE TYPE IS NOT 0 AND DISTANCE IS ABSOLUTE) */
/* Available types:
      - NORMALIZING_LINEAR
      - NORMALIZING_UNIFORM
      - NORMALIZING_NORMAL */
NORMALIZING_LINEAR

/* SEARCH NEIGHBORHOOD PARAMETERS */
/* A search neighborhood is a 3D ellipsoid, defined by:
      - search radii (in number of nodes), for each direction
        (-1.0 for a default value automatically computed)
      - anisotropy ratios, for each direction, i.e. numbers of nodes corresponding
        to a distance of one, in each direction; for example (1.0, 1.0, 2.0) means
        that the distance to the central node is the Euclidean distance where
       the unit (distance=1) corresponds to 1, 1 and 2 nodes for the 1st, 2nd and
       3rd direction respectively.
     - angles (azimuth, dip and plunge) defining the rotation that sends the coordinates
       system xyz onto the coordinates system x'y'z' in which the search radii
       and the anisotropy ratios are given
     - power at which the distance is elevated for computing the weight of each
       node in the search neighborhood
   Note that
      - the search neighborhood is delimited by the search radii and the angles
      - the anisotropy ratios are used only for computing the distance to the central
        node, from each node in the search neighborhood
      - the nodes inside the search neighborhood are sorted according to their
        distance to the central node, from the closest one to the furthest one */
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
 -1.0   -1.0   -1.0   // search radius in each direction
  1.0    1.0    1.0   // anisotropy ratio in each direction
  0.0    0.0    0.0   // angles (azimuth, dip, plunge in degrees) for rotation
  0.0                 // power for computing weight according to distance

/* MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
30

/* MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
   as number of variable(s)) */
1.0

/* RELATIVE DISTANCE FLAG FOR EACH VARIABLE (as many flag(s) (0 / 1) as number of variable(s)) */
0
```

```
/* DISTANCE TYPE FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* Available distance (between data events):
      - 0: non-matching nodes (typically for categorical variable)
      - 1: L-1 distance
      - 2: L-2 distance
      - 3: L-p distance, requires the real positive parameter p
      - 4: L-infinity distance */
0

/* WEIGHT FACTOR FOR CONDITIONING DATA, FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* For the computation of distance between data events, if a value is a conditioning
   data, its corresponding contribution is multiplied by the factor given here. */
2.0

/* SIMULATION AND PATH PARAMETERS */
/* Key word for simulation type:
      - SIM_ONE_BY_ONE:
           successive simulation of one variable at one node in the simulation grid (4D path)
      - SIM_VARIABLE_VECTOR:
           successive simulation of all variable(s) at one node in the simulation grid (3D path) */
SIM_ONE_BY_ONE

/* Key word for path type:
      - PATH_RANDOM:
           random path, for simulation type:
              - SIM_ONE_BY_ONE     : path visiting all nodes and variables in a random order
              - SIM_VARIABLE_VECTOR: path visiting all nodes in a random order
      - PATH_UNILATERAL:
           unilateral path, for simulation type:
              - SIM_ONE_BY_ONE: requires a vector of size 4.
                   Example: u = (0, -2, 1, 0) means that the path will visit all nodes:
                   randomly in xv-sections, with increasing z-coordinate, and then decreasing y-coordinate.
              - SIM_VARIABLE_VECTOR: requires a vector of size 3.
                   Example: u = (-1, 0, 2) means that the path will visit all nodes:
                   randomly in y-sections, with decreasing x-coordinate, and then increasing z-coordinate.
           This vector must be given after the key word PATH_UNILATERAL. */
PATH_RANDOM

/* DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
0.01

/* PROBABILITY CONSTRAINTS */
/* FOR EACH VARIABLE:
   1. Probability constraint usage, integer (probabilityConstraintUsage):
        - 0: no probability constraint
        - 1: global probability constraint
        - 2: local probability constraint

   2. If probabilityConstraintUsage > 0, then the classes of values (for which the
      probability constraints will be given) have to be defined; a class of values
      is given by a union of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[;
      Here are given:
         - nclass: number of classes of values
         - for i in 1,..., nclass: definition of the i-th class of values:
              - ninterval: number of interval(s)
              - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                   these values should satisfy inf_i < sup_i

   3a. If probabilityConstraintUsage == 1, then
       - global probability for each class (defined in 2. above), i.e.
         nclass numbers in [0,1] of sum 1
   3b. If probabilityConstraintUsage == 2, then
       - one pdf image file (for every class, nclass variables)
         (image of same dimensions as the simulation grid)
       - support radius for probability maps (i.e. distance according to
         the unit defined in the search neighborhood parameters for the
         considered variable)
       - method for computing the current pdf (in the simulation grid),
          integer (localCurrentPdfComputation):
            - 0: "COMPLETE" mode: all the informed node in the search neighborhood
                 for the considered variable, and within the support are taken into account
            - 1: "APPROXIMATE" mode: only the neighboring nodes (used for the
                 search in the TI) within the support are taken into account

   4. If probabilityConstraintUsage > 0, then
      method for comparing pdf's, integer (comparingPdfMethod):
         - 0: MAE (Mean Absolute Error)
         - 1: RMSE (Root Mean Squared Error)
         - 2: KLD (Kullback Leibler Divergence)
         - 3: JSD (JSD (Jensen-Shannon Divergence)
         - 4: MLikRsym (Mean Likelihood Ratio (over each class indicator, symmetric target interval))
         - 5: MLikRopt (Mean Likelihood Ratio (over each class indicator, optimal target interval))

   5. If probabilityConstraintUsage > 0, then
       - deactivation distance, i.e. one positive number
         (the probability constraint is deactivated if the distance between
```

```
          the current simulated node and the last node in its neighbors (used
          for the search in the TI) (distance computed according to the corresponding
          search neighborhood parameters) is below the given deactivation distance)

   6. If probabilityConstraintUsage > 0, then
      - threshold type for pdf's comparison, integer (probabilityConstraintThresholdType)
          - 0: constant threshold
          - 1: dynamic threshold
        note: if comparingPdfMethod is set to 4 or 5, probabilityConstraintThresholdType must be set to 0
      6.1a If probabilityConstraintThresholdType == 0, then
             - threshold value
      6.1b If probabilityConstraintThresholdType == 1, then the 7 parameters:
             - M1 M2 M3
             - T1 T2 T3
             - W
          These parameters should satisfy:
             0 <= M1 <= M2 < M3,
             T1 >= T2 >= T3,
             w != 0.
          The threshold value t is defined as a function of the number M
          of nodes used for computing the current pdf (in the simulation grid)
          including the candidate (i.e. current simulated) node by:
             t(M) = T1, if M < M1
             t(M) = T2, if M1 <= M < M2
             t(M) = (T3 - T2) / (M3^W - M2^W) * (M^W - M2^W) + T2, if M2 <= M < M3
             t(M) = T3, if M3 <= M
*/
/* PROBABILITY CONSTRAINTS FOR VARIABLE #0 */
0

/* BLOCK DATA */
/* FOR EACH VARIABLE:
   1. Block data usage, integer (blockDataUsage):
         - 0: no block data
         - 1: use of block data (mean value)

   2. If blockDataUsage == 1, then
         - block data file name
*/
/* BLOCK DATA FOR VARIABLE #0 */
0

/* MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s)) */
0.5

/* TOLERANCE */
/* Tolerance t on the threshold value for flagging nodes:
   let d(i) be the distance between the data event in the simulation grid and in the training
   image for the i-th variable and s(i) be the distance threshold for the i-th variable, and let
   e(i) = max(0, (d(i)-s(i))/s(i)) be the relative error for the i-th variable, i.e. the relative
   part of the distance d(i) beyond the threshold s(i); during the scan of the training image, a node
   that minimizes e = sum (e(i)) is retained (the scan is stopped if e = 0); if e is greater than the
   tolerance t (given here), then the current simulated node and all non-conditioning nodes of the
   data events (one per variable) in the simulation grid are flagged for resimulation (post-processing).
   Note that if probability constraints is used, a similar error as e(i) is computed from the comparison
   of the pdf's, and will contributed in the sum defining the error e. */
0.0

/* POST-PROCESSING */
/* 1. Maximal number of path(s) (npostProcessingPathMax)
   2. If npostProcessingPathMax > 0:
      key word for post-processing parameters (i. e. number of neighboring nodes, distance threshold,
      maximal scan fraction, and tolerance):
         - POST_PROCESSING_PARAMETERS_DEFAULT: for default parameters
         - POST_PROCESSING_PARAMETERS_SAME   : for same parameters as given above
         - POST_PROCESSING_PARAMETERS_MANUAL : for manual settings
   3. If npostProcessingPathMax > 0 and POST_PROCESSING_PARAMETERS_MANUAL:
         MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
            as number of variable(s))
         DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s))
         TOLERANCE
*/
1
POST_PROCESSING_PARAMETERS_DEFAULT

/* PYRAMIDS */
/* I. PYRAMID GENERAL PARAMETERS:
      I.1. Number of pyramid level(s) (in addition to original simulation grid, i.e. number of
           reduction operations), integer (npyramidLevel):
              - = 0: no use of pyramids
              - > 0: use pyramids, npyramidLevel should be equal to the max of "nlevel" entries
                     in pyramid parameters for every variable (point II.1 below);
                     pyramid levels are indexed from fine to coarse:
                        * index 0            : original simulation grid
                        * index npyramidLevel: coarsest level
```

```
      If npyramidLevel > 0:
         I.2. for each level, i.e. for i = 1,..., npyramidLevel:
                 - kx, ky, kz (3 integer): reduction step along x,y,z-direction for the i-th reduction:
                      k[x|y|z] = 0: nothing is done, same dimension in output
                      k[x|y|z] = 1: same dimension in output (with weighted average over 3 nodes)
                      k[x|y|z] = 2: classical gaussian pyramid
                      k[x|y|z] > 2: generalized gaussian pyramid
         I.3. pyramid simulation mode, key work (pyramidSimulationMode):
                 - PYRAMID_SIM_HIERARCHICAL:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the level one rank coarser
                 - PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the gaussian expansion of the level one rank coarser
                 - PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                      co-simulation of all levels, simulation done at one level at a time
         I.4. Factors to adpat the maximal number of neighboring nodes:
              I.4.1. Setting mode, key word (factorNneighboringNodeSettingMode):
                        - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                        - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
              If factorNneighboringNodeSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                 I.4.2. The factors, depending on pyramid simulation mode:
                    - if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL
                      or PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                         - faCond[0], faSim[0], fbCond[0], fbSim[0],
                          ...,
                          faCond[n-1], faSim[n-1], fbCond[n-1], fbSim[n-1],
                          fbSim[n]:
                             I.e. (4*n+1) positive numbers where n = npyramidLevel, with the following
                             meaning. The maximal number of neighboring nodes (according to each variable)
                             is multiplied by
                               (a) faCond[j] and faSim[j] for the conditioning level (level j)
                                   and the simulated level (level j+1) resp. during step (a) above
                               (b) fbCond[j] and fbSim[j] for the conditioning level (level j+1) (expanded
                                   if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION)
                                   and the simulated level (level j) resp. during step (b) above
                   - if pyramidSimulationMode == PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                        - f[0],...,f[npyramidLevel-1],f[npyramidLevel]:
                             I.e. (npyramidLevel + 1) positive numbers, with the following meaning. The
                             maximal number of neighboring nodes (according to each variable) is
                             multiplied by f[j] for the j-th pyramid level.
         I.5. Factors to adpat the distance threshold (similar to I.4):
              I.5.1. Setting mode, key word (factorDistanceThresholdSettingMode):
                        - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                        - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
              If factorDistanceThresholdSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                 I.5.2. The factors, depending on pyramid simulation mode (similar to I.4.2).

         II. PYRAMID PARAMETERS FOR EACH VARIABLE:

         II.1. nlevel: number of pyramid level(s) (number of reduction operations)
                          - = 0: no use of pyramid for the considered variable
                          - > 0: use pyramids for the considered variable, with nlevel level

         If nlevel > 0:
            II.2. Pyramid type, key word (pyramidType):
                      - PYRAMID_CONTINUOUS        : pyramid applied to continuous variable (direct)
                      - PYRAMID_CATEGORICAL_AUTO  : pyramid for categorical variable
                                                       - pyramid for indicator variable of each category
                                                         except one (one pyramid per indicator variable)
                      - PYRAMID_CATEGORICAL_CUSTOM: pyramid for categorical variable
                                                       - pyramid for indicator variable of each class
                                                         of values given explicitly (one pyramid per
                                                         indicator variable)

            If pyramidType == PYRAMID_CATEGORICAL_CUSTOM:
               II.3.  The classes of values (for which the indicator variables are
                  considered for pyramids) have to be defined; a class of values is given by a union
                  of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[.
                  Here are given:
                     - nclass: number of classes of values
                     - for i in 1,..., nclass: definition of the i-th class of values:
                          - ninterval: number of interval(s)
                          - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                               these values should satisfy inf_i < sup_i
                  Then, for each class, the number of pyramid levels (number of reduction operations) is
                     - nlevel_i (for i in 1,..., nclass)
*/
/* PYRAMID GENERAL PARAMETERS */
0

/* SEED NUMBER AND SEED INCREMENT */
```

```
20190210
1

/* NUMBER OF REALIZATIONS */
10

END
```

5.13 Appendix G: The Direct Sampling algorithm parameter file for mineralised zone simulation

```
/* SIMULATION GRID (SG) */
81    94     102    // size    in each direction
 2.5   2.5     2.5  // spacing in each direction
52382.5  9196.25  1096.25  // origin

/* SIMULATION VARIABLES */
/* Number of simulation variable(s), and for each variable:
      variable name, output flag (0 / 1), and if output flag is 1: format string (as passed to fprintf).
   Write DEFAULT_FORMAT for format string to use default format.
   Example (with 3 variables):
      3
      varName1  1  %10.5E
      varName2  0
      varName3  1  DEFAULT_FORMAT
*/
3
POTENCO   0
LITH      0
SHSPLAYS  1  DEFAULT_FORMAT

/* OUTPUT SETTINGS FOR SIMULATION */
/* Key word and required name(s) or prefix, for output of the realizations:
      - OUTPUT_SIM_NO_FILE:
           no file in output
      - OUTPUT_SIM_ALL_IN_ONE_FILE
           one file in output (every variable in output (flagged as 1 above),
           for all realizations will be written),
           - requires one file name
      - OUTPUT_SIM_ONE_FILE_PER_VARIABLE:
            one file per variable in output (flagged as 1 above),
            - requires as many file name(s) as variable(s) flagged as 1 above
      - OUTPUT_SIM_ONE_FILE_PER_REALIZATION:
           one file per realization,
           - requires one prefix (for file name, the string "_real#####.gslib"
             will be appended where "######" is the realization index)
*/
OUTPUT_SIM_ALL_IN_ONE_FILE
sim99_0.gslib

/* OUTPUT: ADDITIONAL MAPS */
/* Additional maps (images) can be retrieved in output (one file per realization).
      - Flag ( 0 / 1) for path index map (index in the simulation path is attached
        to each simulation grid node), and if 1, one prefix (for file name)
      - Flag ( 0 / 1) for error map (error e, see TOLERANCE below) obtained
        during the simulation (excluding possible post-processing paths),
        and if 1, one prefix (for file name)
   Note: the string "_real#####.gslib" will be appended to these prefix, where
   "######" is the realization index. */
1 // path index map
sim99indx
1 // error map
sim99err

/* OUTPUT REPORT */
/* Flag (0 / 1), and if 1, output report file. */
1
sim99_0_report.txt

/* TRAINING IMAGE */
/* Number of training image(s) (nTI >= 1), followed by nTI file(s)
   (a file can be replaced by the string "_DATA_" which means that the
   simulation grid itself is taken as training image), and
   if nTI > 1, one pdf image file (for training images, nTI variables). */
1
bm_97_ti.gslib

/* DATA IMAGE FILE FOR SG */
/* Number of image file(s) (n >= 0), followed by n file(s). */
1
bm_97_cond40_0.gslib

/* DATA POINT SET FILE FOR SG */
/* Number of point set file(s) (n >= 0), followed by n file(s). */
0

```

```
/* MASK IMAGE */
/* Flag (0: mask not used / 1: mask used) and if 1, mask image file
   (this image contains one variable on the simulation grid: flag (0 / 1)
   for each node of the simulation grid that indicates if the variable(s)
   will be simulated at the corresponding node (flag 1) or not (flag 0). */
1
bm_85_mask.gslib

/* HOMOTHETY */
/* 1. Homothety usage, integer (homothetyUsage):
        - 0: no homothety
        - 1: homothety without tolerance
        - 2: homothety with tolerance
   2a. If homothetyUsage == 1,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
   2b. If homothetyUsage == 2,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
0

/* ROTATION */
/* 1. Rotation usage, integer (rotationUsage):
        - 0: no rotation
        - 1: rotation without tolerance
        - 2: rotation with tolerance
   2a. If rotationUsage == 1,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
   2b. If rotationUsage == 2,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
0

/* CONSISTENCY OF CONDITIONING DATA (TOLERANCE RELATIVELY TO THE RANGE OF TRAINING VALUES) */
/* Maximal expansion (expMax): real number (negative to not check consistency).
   The following is applied for each variable separetely:
      - For variable with distance type set to 0 (see below):
           * expMax >= 0:
                if a conditioning data value is not in the set of training image values,
                an error occurs
           * expMax < 0:
                if a conditioning data value is not in the set of training image values,
                a warning is displayed (no error occurs)
      - For variable with distance type not set to 0 (see below): if relative distance
        flag is set to 1 (see below), nothing is done, else:
           * expMax >= 0: maximal accepted expansion of the range of the training image values
                for covering the conditioning data values:
                - if conditioning data values are within the range of the training image values:
                     nothing is done
                - if a conditioning data value is out of the range of the training image values:
                     let
                        new_min_ti = min ( min_cd, min_ti )
                        new_max_ti = max ( max_cd, max_ti )
                     with
                        min_cd, max_cd, the min and max of the conditioning values,
                        min_ti, max_ti, the min and max of the training imges values.
                     If new_max_ti - new_min_ti <= (1 + expMax) * (ti_max - ti_min), then
                     the training image values are linearly rescaled from [ti_min, ti_max] to
                     [new_ti_min, new_ti_max], and a warning is displayed (no error occurs).
                     Otherwise, an error occurs.
           * expMax < 0: if a conditioning data value is out of the range of the training image
                values, a warning is displayed (no error occurs), the training image values are
                not modified.
*/
0.02
```

```
/* NORMALIZATION TYPE (FOR VARIABLES FOR WHICH DISTANCE TYPE IS NOT 0 AND DISTANCE IS ABSOLUTE) */
/* Available types:
      - NORMALIZING_LINEAR
      - NORMALIZING_UNIFORM
      - NORMALIZING_NORMAL */
NORMALIZING_LINEAR

/* SEARCH NEIGHBORHOOD PARAMETERS */
/* A search neighborhood is a 3D ellipsoid, defined by:
      - (rx, ry, rz): search radius (in number of nodes) for each direction
      - (ax, ay, az): anisotropy ratio or inverse distance unit for each direction;
        the distance to the central node is computed as the Euclidean distance by
        considering the nodes with the unit 1/ax x 1/ay x 1/az;
      - (angle1, angle2, angle3): angles (azimuth, dip and plunge) defining the
        rotation that sends the coordinates system xyz onto the coordinates
        system x'y'z' in which the search radii and the anisotropy ratios are given.
      - power: at which the distance is elevated for computing the weight of each
        node in the search neighborhood.
   Note that
     - the search neighborhood is delimited by the search radii and the angles
     - the anisotropy ratios are used only for computing the distance to the central
       node, from each node in the search neighborhood
     - the nodes inside the search neighborhood are sorted according to their
       distance to the central node, from the closest one to the furthest one

   FOR EACH VARIABLE:
      1. search radii, available specifications (format: key word [parameters]):
            SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_DEFAULT
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XY
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XZ
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_YZ
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XYZ
            SEARCHNEIGHBORHOOD_RADIUS_MANUAL rx ry rz
         with the following meaning, where tx, ty, tz denote the ranges of the
         variogram of the TI(s) in x-, y-, z-directions:
            - SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT:
                 large radii set according to the size of the SG and the TI(s),
                 and the use of homothethy and/or rotation for the simulation
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_DEFAULT:
                 search radii set according to the TI(s) variogram ranges,
                 one of the 5 next modes SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE* will be
                 used according to the use of homothethy and/or rotation for the
                 simulation
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE:
                 rx prop. to tx, ry prop. to ty, rz prop. to tz,
                 independently in each direction
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XY:
                 rx = ry  prop. to max(tx, ty), independently from rz prop. to tz
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XZ:
                 rx = rz  prop. to max(tx, tz), independently from ry prop. to ty
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_YZ:
                 ry = rz  prop. to max(ty, tz), independently from rx prop. to tx
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XYZ:
                 rx = ry = rz prop. to max(tx, ty, tz)
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_MANUAL:
                 search radii are explicitly given

      2. anisotropy ratios, available specifications (format: key word [parameters]):
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XY
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_YZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XYZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_MANUAL ax ay az
         with the following meaning:
            - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE:
                 ax = ay = az = 1
                 (isotropic distance:
                 maximal distance for search neighborhood nodes will be equal to the
                 maximum of the search radii)
            - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS:
```

```
                ax = rx, ay = ry, az = rz
                (anisotropic distance:
                nodes at distance one on the border of the search neighborhood,
                maximal distance for search neighborhood nodes will be 1)
             - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XY:
                ax = ay = max(rx, ry), az = rz
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
             - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XZ:
                ax = az = max(rx, rz), ay = ry
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
             - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_YZ:
                ay = az = max(ry, rz), ax = rx
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
             - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XYZ:
                ax = ay = az = max(rx, ry, rz)
                (isotropic distance:
                maximal distance for search neighborhood nodes will be 1)
             - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_MANUAL:
                anisotropy ratios are explicitly given

      3. rotation, available specifications (format: key word [parameters]):
            SEARCHNEIGHBORHOOD_ROTATION_OFF
            SEARCHNEIGHBORHOOD_ROTATION_ON angle1 angle2 angle3
         with the following meaning:
            - SEARCHNEIGHBORHOOD_ROTATION_OFF:
                search neighborhood is not rotated
            - SEARCHNEIGHBORHOOD_ROTATION_ON:
                search neighborhood is rotated, the rotation is described by
                angle1 (azimuth), angle2 (dip) and angle3 (plunge)

      4. power (real number)
*/
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance

/* MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
40 40 80

/* MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
   as number of variable(s)) */
1.0 1.0 1.0

/* RESCALING MODE FOR EACH VARIABLE (as many specification as number of variable(s))*/
/* Available specifications (format: key word [parameters]):
      RESCALING_NONE
      RESCALING_MIN_MAX min_value max_value
      RESCALING_MEAN_LENGTH mean_value length_value
   with the following meaning:
      - RESCALING_NONE:
           target interval for simulated values = interval of training image(s) value(s)
           (standard mode)
      - RESCALING_MIN_MAX:
           target interval for simulated values given by min_value and max_value
      - RESCALING_MEAN_LENGTH:
           target interval for simulated values given by mean_value and length_value
   For the two last modes, a linear correspondence is set between the target interval
   for simulated values and the interval of the training image(s):
      - for mode RESCALING_MIN_MAX, min_value and max_value are set in correspondence
        with the min and max training image(s) values respectively
      - for mode RESCALING_MEAN_LENGTH, mean_value is set in correspondence
        with the mean training image(s) value, and length_value is the length
        of the target interval.
   Note that:
      - conditioning data values (if any) must be in the target interval.
      - the two last modes are not compatible with distance type 0
        (non-matching nodes), or with relative distance mode (see below)
      - class of values (for other options such probability constraints or
```

```
        connectivity constraints) are given according to the target interval.
*/
/* RESCALING FOR VARIABLE #0 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #1 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #2 */
RESCALING_NONE

/* RELATIVE DISTANCE FLAG FOR EACH VARIABLE (as many flag(s) (0 / 1) as number of variable(s)) */
0 0 0

/* DISTANCE TYPE FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* Available distance (between data events):
      - 0: non-matching nodes (typically for categorical variable)
      - 1: L-1 distance
      - 2: L-2 distance
      - 3: L-p distance, requires the real positive parameter p
      - 4: L-infinity distance */
1 0 0

/* WEIGHT FACTOR FOR CONDITIONING DATA, FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* For the computation of distance between data events, if a value is a conditioning
   data, its corresponding contribution is multiplied by the factor given here. */
1 1 1

/* SIMULATION TYPE */
/* Key word:
      - SIM_ONE_BY_ONE:
           successive simulation of one variable at one node in the simulation grid (4D path)
      - SIM_VARIABLE_VECTOR:
           successive simulation of all variable(s) at one node in the simulation grid (3D path) */
SIM_VARIABLE_VECTOR

/* SIMULATION PATH */
/* Available paths (format: key word [parameters]):
      - PATH_RANDOM
      - PATH_RANDOM_HD_DISTANCE_PDF s
      - PATH_RANDOM_HD_DISTANCE_SORT s
      - PATH_RANDOM_HD_DISTANCE_SUM_PDF p s
      - PATH_RANDOM_HD_DISTANCE_SUM_SORT p s
      - PATH_UNILATERAL u (vector)
   with the following meaning:
      - PATH_RANDOM:
           Random path, for simulation type:
              - SIM_ONE_BY_ONE     : path visiting all nodes and variables in a random order
              - SIM_VARIABLE_VECTOR: path visiting all nodes in a random order
      - PATH_RANDOM_HD_DISTANCE_PDF:
           This path preferentially visits first the SG nodes close to the conditioning nodes.
           For any uninformed node, the distance d to the set of conditioning nodes is computed,
           and a weight proportional to a^d is attached, with 0 < a = 1 + s*(a0-1) <= 1, a0
           being the minimal value for a.
           The path is then built by successively drawing nodes according to their weight.
           The required parameter s in [0,1] controls the strength of the distance
           (s = 0: random, s = 1: most guided by the distance).
           Notes: 1) in absence of conditioning node, one random node in the path is considered
           to compute the distances, 2) if simulation type is SIM_VARIABLE_VECTOR, variables
           exhaustively informed are not considered as conditioning data for computing distance.
      - PATH_RANDOM_HD_DISTANCE_SORT:
           This path preferentially visits first the SG nodes close to the conditioning nodes.
           For any uninformed node, the distance d to the set of conditioning nodes is computed,
           and the normalized distance d' in [0,1] is combined with a random number r in [0,1]
           to define a weight w = s*d' + (1-s)*r.
           The path is then built by sorting the nodes such that the weight are in increasing order.
           The required parameter s in [0,1] controls the strength of the distance
           (s = 0: random, s = 1: most guided by the distance - quasi deterministic).
           Notes: 1) in absence of conditioning node, one random node in the path is considered
           to compute the distances, 2) if simulation type is SIM_VARIABLE_VECTOR, variables
           exhaustively informed are not considered as conditioning data for computing distance.
      - PATH_RANDOM_HD_DISTANCE_SUM_PDF:
           The path is built as for PATH_RANDOM_HD_DISTANCE_PDF, but the distance to the set of
           conditioning nodes is replaced by the sum of the distances to every conditioning node,
           each distance being raised to a power p.
           Two parameters are required: p > 0 controls the computation of the distances,
           and s in [0,1] controls the strength of the distance
           (s = 0: random, s = 1: most guided by the distance).
           Note: 1) in absence of conditioning node, random path is considered.
      - PATH_RANDOM_HD_DISTANCE_SUM_SORT:
           The path is built as for PATH_RANDOM_HD_DISTANCE_SORT, but the distance to the set of
           conditioning nodes is replaced by the sum of the distances to every conditioning node,
           each distance being raised to a power p.
           Two parameters are required: p > 0 controls the computation of the distances,
           and s in [0,1] controls the strength of the distance
```

```
          (s = 0: random, s = 1: most guided by the distance - quasi deterministic).
          Note: 1) in absence of conditioning node, random path is considered.
      - PATH_UNILATERAL:
          It requires a vector u of parameters of size n=4 (resp. n=3) for simulation type set to
          SIM_ONE_BY_ONE (resp. SIM_VARIABLE_VECTOR), given as ux uy uz uv (resp. ux uy uz).
          Some components of u can be equal to 0, and the other ones must be the integer 1,...,m,
          with sign +/-. For
              u[j(1)] = ... = u[j(n-m)] = 0 and (u[i(1)], ..., u[i(m)]) = (+/-1,...,+/-m),
          the path visits all nodes in sections of coordinates j(1),...,j(n-m) randomly,
          and makes varying the i(1)-th coordinate first (most rapidly), ..., the i(m)-th coordinate
          last, each one in increasing or decreasing order according to the sign (+ or - resp.).
          Examples:
            - for simulation type SIM_ONE_BY_ONE, and u = (0, -2, 1, 0), then the path will visit
              all nodes: randomly in xv-sections, with increasing z-coordinate first (most rapidly),
              and decreasing y-coordinate.
            - for simulation type SIM_VARIABLE_VECTOR, and u = (0, 2, 1), then the path will visit
              all nodes: randomly in x-direction, with increasing z-coordinate first (most rapidly),
              and increasing y-coordinate.
*/
PATH_RANDOM_HD_DISTANCE_PDF 1

/* DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
0.05 0.01 0.01

/* PROBABILITY CONSTRAINTS */
/* FOR EACH VARIABLE:
   1. Probability constraint usage, integer (probabilityConstraintUsage):
        - 0: no probability constraint
        - 1: global probability constraint
        - 2: local probability constraint

   2. If probabilityConstraintUsage > 0, then the classes of values (for which the
       probability constraints will be given) have to be defined; a class of values
       is given by a union of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[;
       Here are given:
          - nclass: number of classes of values
          - for i in 1,..., nclass: definition of the i-th class of values:
               - ninterval: number of interval(s)
               - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                    these values should satisfy inf_i < sup_i

   3a. If probabilityConstraintUsage == 1, then
         - global probability for each class (defined in 2. above), i.e.
           nclass numbers in [0,1] of sum 1
   3b. If probabilityConstraintUsage == 2, then
         - one pdf image file (for every class, nclass variables)
           (image of same dimensions as the simulation grid)
         - support radius for probability maps (i.e. distance according to
           the unit defined in the search neighborhood parameters for the
           considered variable)
         - method for computing the current pdf (in the simulation grid),
            integer (localCurrentPdfComputation):
              - 0: "COMPLETE" mode: all the informed node in the search neighborhood
                   for the considered variable, and within the support are taken into account
              - 1: "APPROXIMATE" mode: only the neighboring nodes (used for the
                   search in the TI) within the support are taken into account

   4. If probabilityConstraintUsage > 0, then
       method for comparing pdf's, integer (comparingPdfMethod):
          - 0: MAE (Mean Absolute Error)
          - 1: RMSE (Root Mean Squared Error)
          - 2: KLD (Kullback Leibler Divergence)
          - 3: JSD (JSD (Jensen-Shannon Divergence)
          - 4: MLikRsym (Mean Likelihood Ratio (over each class indicator, symmetric target interval))
          - 5: MLikRopt (Mean Likelihood Ratio (over each class indicator, optimal target interval))

   5. If probabilityConstraintUsage > 0, then
       - deactivation distance, i.e. one positive number
         (the probability constraint is deactivated if the distance between
         the current simulated node and the last node in its neighbors (used
         for the search in the TI) (distance computed according to the corresponding
         search neighborhood parameters) is below the given deactivation distance)

   6. If probabilityConstraintUsage > 0, then
       - threshold type for pdf's comparison, integer (probabilityConstraintThresholdType)
            - 0: constant threshold
            - 1: dynamic threshold
         note: if comparingPdfMethod is set to 4 or 5, probabilityConstraintThresholdType must be set to 0
       6.1a If probabilityConstraintThresholdType == 0, then
              - threshold value
       6.1b If probabilityConstraintThresholdType == 1, then the 7 parameters:
              - M1 M2 M3
              - T1 T2 T3
              - W
```

```
492                These parameters should satisfy:
493                   0 <= M1 <= M2 < M3,
494                   T1 >= T2 >= T3,
495                   w != 0.
496                The threshold value t is defined as a function of the number M
497                of nodes used for computing the current pdf (in the simulation grid)
498                including the candidate (i.e. current simulated) node by:
499                   t(M) = T1, if M < M1
500                   t(M) = T2, if M1 <= M < M2
501                   t(M) = (T3 - T2) / (M3^W - M2^W) * (M^W - M2^W) + T2, if M2 <= M < M3
502                   t(M) = T3, if M3 <= M
503 */
504 /* PROBABILITY CONSTRAINTS FOR VARIABLE #0 */
505 0
506 /* PROBABILITY CONSTRAINTS FOR VARIABLE #1 */
507 0
508 /* PROBABILITY CONSTRAINTS FOR VARIABLE #2 */
509 0
510
511
512 /* CONNECTIVITY CONSTRAINTS */
513 /* FOR EACH VARIABLE:
514    1. Connectivity constraint usage, integer (connectivityConstraintUsage):
515          - 0: no connectivity constraint
516          - 1: set connecting paths before simulation by successively
517               binding the nodes to be connected in a random order
518          - 2: set connecting paths before simulation by successively
519               binding the nodes to be connected beginning with
520               the pair of nodes with the smallest distance and then
521               the remaining nodes in increasing order according to
522               their distance to the set of nodes already connected.
523          - 3: check connectivity pattern during simulation
524    2. If connectivityConstraintUsage > 0, then key word for type of connectivity:
525          - CONNECT_FACE             : 6-neighbors connected (by face)
526          - CONNECT_FACE_EDGE        : 18-neighbors connected (by face or edge)
527          - CONNECT_FACE_EDGE_CORNER : 26-neigbhors connected (by face, edge or corner)
528    3. If connectivityConstraintUsage > 0, then the classes of values (that can
529          be considered in the same connected components) have to be defined; a
530          class of values is given by a union of interval(s):
531             [inf_1,sup_1[ U ... U [inf_n,sup_n[;
532          Here are given:
533             - nclass: number of classes of values
534             - for i in 1,...,nclass: definition of the i-th class of values:
535                  - ninterval: number of interval(s)
536                  - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
537                       these values should satisfy inf_i < sup_i
538    4. If connectivityConstraintUsage > 0, then:
539          - variable name for connected component label
540            (included in data image / data point set above)
541            (Note: label negative or zero means no connectivity constraint)
542    5a. If connectivityConstraintUsage == 1 or connectivityConstraintUsage == 2, then:
543          - name of the image file and name of the variable for the search of
544            connected paths (set string "_TI_" instead for searching in the (first)
545            training image and the corresponding variable index)
546    5b. If connectivityConstraintUsage == 3, then:
547          - deactivation distance, i.e. one positive number
548            (the connectivity constraint is deactivated if the distance between
549            the current simulated node and the last node in its neighbors (used
550            for the search in the TI) (distance computed according to the corresponding
551            search neighborhood parameters) is below the given deactivation distance)
552          - threshold: threshold value for acceptation of connectivity pattern
553 */
554 /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #0 */
555 0
556 /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #1 */
557 0
558 /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #2 */
559 0
560
561 /* BLOCK DATA */
562 /* FOR EACH VARIABLE:
563    1. Block data usage, integer (blockDataUsage):
564          - 0: no block data
565          - 1: use of block data (mean value)
566
567    2. If blockDataUsage == 1, then
568          - block data file name
569 */
570 /* BLOCK DATA FOR VARIABLE #0 */
571 0
572 /* BLOCK DATA FOR VARIABLE #1 */
573 0
```

```
/* BLOCK DATA FOR VARIABLE #2 */
1
bm_99_block.txt

/* MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s)) */
0.25

/* PYRAMIDS */
/* I. PYRAMID GENERAL PARAMETERS:
      I.1. Number of pyramid level(s) (in addition to original simulation grid, i.e. number of
           reduction operations), integer (npyramidLevel):
              - = 0: no use of pyramids
              - > 0: use pyramids, npyramidLevel should be equal to the max of "nlevel" entries
                     in pyramid parameters for every variable (point II.1 below);
                     pyramid levels are indexed from fine to coarse:
                        * index 0            : original simulation grid
                        * index npyramidLevel: coarsest level
      If npyramidLevel > 0:
         I.2. for each level, i.e. for i = 1,..., npyramidLevel:
                 - kx, ky, kz (3 integer): reduction step along x,y,z-direction for the i-th reduction:
                      k[x|y|z] = 0: nothing is done, same dimension in output
                      k[x|y|z] = 1: same dimension in output (with weighted average over 3 nodes)
                      k[x|y|z] = 2: classical gaussian pyramid
                      k[x|y|z] > 2: generalized gaussian pyramid
         I.3. pyramid simulation mode, key work (pyramidSimulationMode):
                 - PYRAMID_SIM_HIERARCHICAL:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the level one rank coarser
                 - PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the gaussian expansion of the level one rank coarser
                 - PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                      co-simulation of all levels, simulation done at one level at a time
         I.4. Factors to adapt the maximal number of neighboring nodes:
              I.4.1. Setting mode, key word (factorNneighboringNodeSettingMode):
                        - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                        - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
              If factorNneighboringNodeSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                 I.4.2. The factors, depending on pyramid simulation mode:
                    - if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL
                      or PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                         - faCond[0], faSim[0], fbCond[0], fbSim[0],
                           ...,
                           faCond[n-1], faSim[n-1], fbCond[n-1], fbSim[n-1],
                           fbSim[n]:
                              I.e. (4*n+1) positive numbers where n = npyramidLevel, with the following
                              meaning. The maximal number of neighboring nodes (according to each variable)
                              is multiplied by
                                (a) faCond[j] and faSim[j] for the conditioning level (level j)
                                    and the simulated level (level j+1) resp. during step (a) above
                                (b) fbCond[j] and fbSim[j] for the conditioning level (level j+1) (expanded
                                    if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION)
                                    and the simulated level (level j) resp. during step (b) above
                    - if pyramidSimulationMode == PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                         - f[0],...,f[npyramidLevel-1],f[npyramidLevel]:
                              I.e. (npyramidLevel + 1) positive numbers, with the following meaning. The
                              maximal number of neighboring nodes (according to each variable) is
                              multiplied by f[j] for the j-th pyramid level.
         I.5. Factors to adapt the distance threshold (similar to I.4):
              I.5.1. Setting mode, key word (factorDistanceThresholdSettingMode):
                        - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                        - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
              If factorDistanceThresholdSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                 I.5.2. The factors, depending on pyramid simulation mode (similar to I.4.2).

         II. PYRAMID PARAMETERS FOR EACH VARIABLE:

         II.1. nlevel: number of pyramid level(s) (number of reduction operations)
                        - = 0: no use of pyramid for the considered variable
                        - > 0: use pyramids for the considered variable, with nlevel level

         If nlevel > 0:
            II.2. Pyramid type, key word (pyramidType):
                      - PYRAMID_CONTINUOUS        : pyramid applied to continuous variable (direct)
                      - PYRAMID_CATEGORICAL_AUTO  : pyramid for categorical variable
                                                        - pyramid for indicator variable of each category
                                                          except one (one pyramid per indicator variable)
```

```
                    - PYRAMID_CATEGORICAL_CUSTOM: pyramid for categorical variable
                                                    - pyramid for indicator variable of each class
                                                      of values given explicitly (one pyramid per
                                                      indicator variable)

            If pyramidType == PYRAMID_CATEGORICAL_CUSTOM:
               II.3.  The classes of values (for which the indicator variables are
                  considered for pyramids) have to be defined; a class of values is given by a union
                  of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[.
                  Here are given:
                     - nclass: number of classes of values
                     - for i in 1,..., nclass: definition of the i-th class of values:
                          - ninterval: number of interval(s)
                          - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                               these values should satisfy inf_i < sup_i
                  Then, for each class, the number of pyramid levels (number of reduction operations) is
                     - nlevel_i (for i in 1,..., nclass)
*/
/* PYRAMID GENERAL PARAMETERS */
0

/* TOLERANCE */
/* Tolerance t on the threshold value for flagging nodes (for post-processing):
   let d(i) be the distance between the data event in the simulation grid and in the training
   image for the i-th variable and s(i) be the distance threshold for the i-th variable, and let
   e(i) = max(0, (d(i)-s(i))/s(i)) be the relative error for the i-th variable, i.e. the relative
   part of the distance d(i) beyond the threshold s(i); during the scan of the training image, a node
   that minimizes e = sum (e(i)) is retained (the scan is stopped if e = 0); if e is greater than the
   tolerance t (given here), then the current simulated node and all non-conditioning nodes of the
   data events (one per variable) in the simulation grid are flagged for resimulation (post-processing).
   Note that if probability / connectivity / block data constraints used a similar error as e(i) that
   contributes in the sum defining the error e.
*/
0.01

/* POST-PROCESSING */
/* 1. Maximal number of path(s) (npostProcessingPathMax)
   2. If npostProcessingPathMax > 0:
      key word for post-processing parameters (i. e. number of neighboring nodes, distance threshold,
      maximal scan fraction, and tolerance):
         - POST_PROCESSING_PARAMETERS_DEFAULT: for default parameters
         - POST_PROCESSING_PARAMETERS_SAME   : for same parameters as given above
         - POST_PROCESSING_PARAMETERS_MANUAL : for manual settings
   3. If npostProcessingPathMax > 0 and POST_PROCESSING_PARAMETERS_MANUAL:
         MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
            as number of variable(s))
         DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s))
         TOLERANCE
*/
1
POST_PROCESSING_PARAMETERS_DEFAULT

/* SEED NUMBER AND SEED INCREMENT */
190514
1

/* NUMBER OF REALIZATION(S) */
1

END
```

5.14 Appendix H: The Direct Sampling algorithm parameter file for gold grade simulation

```
/* SIMULATION GRID (SG) */
   81 94 102       // size    in each direction
    2.5  2.5     2.5  // spacing in each direction
52382.5 9196.25 1096.25  // origin

/* SIMULATION VARIABLES */
/* Number of simulation variable(s), and for each variable:
      variable name, output flag (0 / 1), and if output flag is 1: format string (as passed to fprintf).
   Write DEFAULT_FORMAT for format string to use default format.
   Example (with 3 variables):
      3
      varName1  1  %10.5E
      varName2  0
      varName3  1  DEFAULT_FORMAT
*/
5
POTENCO   0
LITH      0
SHSPLAYS  0
AUCAT     1  DEFAULT_FORMAT
AU        1  DEFAULT_FORMAT

/* OUTPUT SETTINGS FOR SIMULATION */
/* Key word and required name(s) or prefix, for output of the realizations:
      - OUTPUT_SIM_NO_FILE:
           no file in output
      - OUTPUT_SIM_ALL_IN_ONE_FILE
           one file in output (every variable in output (flagged as 1 above),
           for all realizations will be written),
           - requires one file name
      - OUTPUT_SIM_ONE_FILE_PER_VARIABLE:
            one file per variable in output (flagged as 1 above),
            - requires as many file name(s) as variable(s) flagged as 1 above
      - OUTPUT_SIM_ONE_FILE_PER_REALIZATION:
           one file per realization,
           - requires one prefix (for file name, the string "_real#####.gslib"
             will be appended where "######" is the realization index)
*/
OUTPUT_SIM_ALL_IN_ONE_FILE
sim109_0.gslib

/* OUTPUT: ADDITIONAL MAPS */
/* Additional maps (images) can be retrieved in output (one file per realization).
      - Flag ( 0 / 1) for path index map (index in the simulation path is attached
        to each simulation grid node), and if 1, one prefix (for file name)
      - Flag ( 0 / 1) for error map (error e, see TOLERANCE below) obtained
        during the simulation (excluding possible post-processing paths),
        and if 1, one prefix (for file name)
   Note: the string "_real#####.gslib" will be appended to these prefix, where
   "######" is the realization index. */
0 // path index map
0 // error map

/* OUTPUT REPORT */
/* Flag (0 / 1), and if 1, output report file. */
1
sim109_0_report.txt

/* TRAINING IMAGE */
/* Number of training image(s) (nTI >= 1), followed by nTI file(s)
   (a file can be replaced by the string "_DATA_" which means that the
   simulation grid itself is taken as training image), and
   if nTI > 1, one pdf image file (for training images, nTI variables). */
1
bm_95_ti.gslib

/* DATA IMAGE FILE FOR SG */
/* Number of image file(s) (n >= 0), followed by n file(s). */
1
bm_109_cond40_0.gslib

/* DATA POINT SET FILE FOR SG */
/* Number of point set file(s) (n >= 0), followed by n file(s). */
0

```

```
/* MASK IMAGE */
/* Flag (0: mask not used / 1: mask used) and if 1, mask image file
   (this image contains one variable on the simulation grid: flag (0 / 1)
   for each node of the simulation grid that indicates if the variable(s)
   will be simulated at the corresponding node (flag 1) or not (flag 0). */
1
bm_85_mask.gslib

/* HOMOTHETY */
/* 1. Homothety usage, integer (homothetyUsage):
        - 0: no homothety
        - 1: homothety without tolerance
        - 2: homothety with tolerance
   2a. If homothetyUsage == 1,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
   2b. If homothetyUsage == 2,
          then for homothety ratio in each direction,
          first for x, then for y, and then for z-axis direction:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
0

/* ROTATION */
/* 1. Rotation usage, integer (rotationUsage):
        - 0: no rotation
        - 1: rotation without tolerance
        - 2: rotation with tolerance
   2a. If rotationUsage == 1,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - one value (real) if flag is 0
                  - name of the image file (one variable) if flag is 1
   2b. If rotationUsage == 2,
          then for each angle,
          first for azimuth, then for dip, and then for plunge:
             - Flag (0 / 1) indicating if given in an image file,
               followed by
                  - two values (lower and upper bounds) (real) if flag is 0
                  - name of the image file (two variables) if flag is 1
*/
0

/* CONSISTENCY OF CONDITIONING DATA (TOLERANCE RELATIVELY TO THE RANGE OF TRAINING VALUES) */
/* Maximal expansion (expMax): real number (negative to not check consistency).
   The following is applied for each variable separetely:
      - For variable with distance type set to 0 (see below):
           * expMax >= 0:
                if a conditioning data value is not in the set of training image values,
                an error occurs
           * expMax < 0:
                if a conditioning data value is not in the set of training image values,
                a warning is displayed (no error occurs)
      - For variable with distance type not set to 0 (see below): if relative distance
        flag is set to 1 (see below), nothing is done, else:
           * expMax >= 0: maximal accepted expansion of the range of the training image values
                for covering the conditioning data values:
                - if conditioning data values are within the range of the training image values:
                     nothing is done
                - if a conditioning data value is out of the range of the training image values:
                     let
                        new_min_ti = min ( min_cd, min_ti )
                        new_max_ti = max ( max_cd, max_ti )
                     with
                        min_cd, max_cd, the min and max of the conditioning values,
                        min_ti, max_ti, the min and max of the training imges values.
                     If new_max_ti - new_min_ti <= (1 + expMax) * (ti_max - ti_min), then
                     the training image values are linearly rescaled from [ti_min, ti_max] to
                     [new_ti_min, new_ti_max], and a warning is displayed (no error occurs).
                     Otherwise, an error occurs.
           * expMax < 0: if a conditioning data value is out of the range of the training image
                values, a warning is displayed (no error occurs), the training image values are
                not modified.
```

```
*/
0.02

/* NORMALIZATION TYPE (FOR VARIABLES FOR WHICH DISTANCE TYPE IS NOT 0 AND DISTANCE IS ABSOLUTE) */
/* Available types:
      - NORMALIZING_LINEAR
      - NORMALIZING_UNIFORM
      - NORMALIZING_NORMAL */
NORMALIZING_LINEAR

/* SEARCH NEIGHBORHOOD PARAMETERS */
/* A search neighborhood is a 3D ellipsoid, defined by:
      - (rx, ry, rz): search radius (in number of nodes) for each direction
      - (ax, ay, az): anisotropy ratio or inverse distance unit for each direction;
        the distance to the central node is computed as the Euclidean distance by
        considering the nodes with the unit 1/ax x 1/ay x 1/az;
      - (angle1, angle2, angle3): angles (azimuth, dip and plunge) defining the
        rotation that sends the coordinates system xyz onto the coordinates
        system x'y'z' in which the search radii and the anisotropy ratios are given.
      - power: at which the distance is elevated for computing the weight of each
        node in the search neighborhood.
   Note that
     - the search neighborhood is delimited by the search radii and the angles
     - the anisotropy ratios are used only for computing the distance to the central
       node, from each node in the search neighborhood
     - the nodes inside the search neighborhood are sorted according to their
       distance to the central node, from the closest one to the furthest one

   FOR EACH VARIABLE:
      1. search radii, available specifications (format: key word [parameters]):
            SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_DEFAULT
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XY
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XZ
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_YZ
            SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XYZ
            SEARCHNEIGHBORHOOD_RADIUS_MANUAL rx ry rz
         with the following meaning, where tx, ty, tz denote the ranges of the
         variogram of the TI(s) in x-, y-, z-directions:
            - SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT:
                 large radii set according to the size of the SG and the TI(s),
                 and the use of homothethy and/or rotation for the simulation
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_DEFAULT:
                 search radii set according to the TI(s) variogram ranges,
                 one of the 5 next modes SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE* will be
                 used according to the use of homothethy and/or rotation for the
                 simulation
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE:
                 rx prop. to tx, ry prop. to ty, rz prop. to tz,
                 independently in each direction
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XY:
                 rx = ry  prop. to max(tx, ty), independently from rz prop. to tz
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XZ:
                 rx = rz  prop. to max(tx, tz), independently from ry prop. to ty
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_YZ:
                 ry = rz  prop. to max(ty, tz), independently from rx prop. to tx
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_TI_RANGE_XYZ:
                 rx = ry = rz prop. to max(tx, ty, tz)
                 (automatically computed)
            - SEARCHNEIGHBORHOOD_RADIUS_MANUAL:
                 search radii are explicitly given

      2. anisotropy ratios, available specifications (format: key word [parameters]):
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XY
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_YZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XYZ
            SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_MANUAL ax ay az
         with the following meaning:
            - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE:
                 ax = ay = az = 1
                 (isotropic distance:
                 maximal distance for search neighborhood nodes will be equal to the
                 maximum of the search radii)
            - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS:
```

```
                ax = rx, ay = ry, az = rz
                (anisotropic distance:
                nodes at distance one on the border of the search neighborhood,
                maximal distance for search neighborhood nodes will be 1)
           - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XY:
                ax = ay = max(rx, ry), az = rz
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
           - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XZ:
                ax = az = max(rx, rz), ay = ry
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
           - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_YZ:
                ay = az = max(ry, rz), ax = rx
                (anisotropic distance:
                maximal distance for search neighborhood nodes will be 1)
           - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_RADIUS_XYZ:
                ax = ay = az = max(rx, ry, rz)
                (isotropic distance:
                maximal distance for search neighborhood nodes will be 1)
           - SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_MANUAL:
                anisotropy ratios are explicitly given

      3. rotation, available specifications (format: key word [parameters]):
           SEARCHNEIGHBORHOOD_ROTATION_OFF
           SEARCHNEIGHBORHOOD_ROTATION_ON angle1 angle2 angle3
         with the following meaning:
           - SEARCHNEIGHBORHOOD_ROTATION_OFF:
                search neighborhood is not rotated
           - SEARCHNEIGHBORHOOD_ROTATION_ON:
                search neighborhood is rotated, the rotation is described by
                angle1 (azimuth), angle2 (dip) and angle3 (plunge)

      4. power (real number)
*/
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance
/* SEARCH NEIGHBORHOOD PARAMETERS FOR VARIABLE #0 */
SEARCHNEIGHBORHOOD_RADIUS_LARGE_DEFAULT          // search radii
SEARCHNEIGHBORHOOD_ANISOTROPY_RATIO_ONE          // anisotropy ratios
SEARCHNEIGHBORHOOD_ROTATION_OFF                  // rotation
0.0                                              // power for computing weight according to distance

/* MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
40 100 100 100 100

/* MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
   as number of variable(s)) */
1.0 1.0 1.0 1.0 1.0

/* RESCALING MODE FOR EACH VARIABLE (as many specification as number of variable(s))*/
/* Available specifications (format: key word [parameters]):
      RESCALING_NONE
      RESCALING_MIN_MAX min_value max_value
      RESCALING_MEAN_LENGTH mean_value length_value
   with the following meaning:
      - RESCALING_NONE:
           target interval for simulated values = interval of training image(s) value(s)
           (standard mode)
      - RESCALING_MIN_MAX:
           target interval for simulated values given by min_value and max_value
      - RESCALING_MEAN_LENGTH:
           target interval for simulated values given by mean_value and length_value
   For the two last modes, a linear correspondence is set between the target interval
   for simulated values and the interval of the training image(s):
      - for mode RESCALING_MIN_MAX, min_value and max_value are set in correspondence
```

```
        with the min and max training image(s) values respectively
      - for mode RESCALING_MEAN_LENGTH, mean_value is set in correspondence
        with the mean training image(s) value, and length_value is the length
        of the target interval.
   Note that:
      - conditioning data values (if any) must be in the target interval.
      - the two last modes are not compatible with distance type 0
        (non-matching nodes), or with relative distance mode (see below)
      - class of values (for other options such probability constraints or
        connectivity constraints) are given according to the target interval.
*/
/* RESCALING FOR VARIABLE #0 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #1 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #2 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #3 */
RESCALING_NONE
/* RESCALING FOR VARIABLE #4 */
RESCALING_NONE

/* RELATIVE DISTANCE FLAG FOR EACH VARIABLE (as many flag(s) (0 / 1) as number of variable(s)) */
0 0 0 0 0

/* DISTANCE TYPE FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* Available distance (between data events):
      - 0: non-matching nodes (typically for categorical variable)
      - 1: L-1 distance
      - 2: L-2 distance
      - 3: L-p distance, requires the real positive parameter p
      - 4: L-infinity distance */
1 0 0 0 1

/* WEIGHT FACTOR FOR CONDITIONING DATA, FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
/* For the computation of distance between data events, if a value is a conditioning
   data, its corresponding contribution is multiplied by the factor given here. */
1 1 1 2 2

/* SIMULATION TYPE */
/* Key word:
      - SIM_ONE_BY_ONE:
           successive simulation of one variable at one node in the simulation grid (4D path)
      - SIM_VARIABLE_VECTOR:
           successive simulation of all variable(s) at one node in the simulation grid (3D path) */
SIM_VARIABLE_VECTOR

/* SIMULATION PATH */
/* Available paths (format: key word [parameters]):
      - PATH_RANDOM
      - PATH_RANDOM_HD_DISTANCE_PDF s
      - PATH_RANDOM_HD_DISTANCE_SORT s
      - PATH_RANDOM_HD_DISTANCE_SUM_PDF p s
      - PATH_RANDOM_HD_DISTANCE_SUM_SORT p s
      - PATH_UNILATERAL u (vector)
   with the following meaning:
      - PATH_RANDOM:
           Random path, for simulation type:
              - SIM_ONE_BY_ONE     : path visiting all nodes and variables in a random order
              - SIM_VARIABLE_VECTOR: path visiting all nodes in a random order
      - PATH_RANDOM_HD_DISTANCE_PDF:
           This path preferentially visits first the SG nodes close to the conditioning nodes.
           For any uninformed node, the distance d to the set of conditioning nodes is computed,
           and a weight proportional to a^d is attached, with 0 < a = 1 + s*(a0-1) <= 1, a0
           being the minimal value for a.
           The path is then built by successively drawing nodes according to their weight.
           The required parameter s in [0,1] controls the strength of the distance
           (s = 0: random, s = 1: most guided by the distance).
           Notes: 1) in absence of conditioning node, one random node in the path is considered
           to compute the distances, 2) if simulation type is SIM_VARIABLE_VECTOR, variables
           exhaustively informed are not considered as conditioning data for computing distance.
     - PATH_RANDOM_HD_DISTANCE_SORT:
           This path preferentially visits first the SG nodes close to the conditioning nodes.
           For any uninformed node, the distance d to the set of conditioning nodes is computed,
           and the normalized distance d' in [0,1] is combined with a random number r in [0,1]
           to define a weight w = s*d' + (1-s)*r.
           The path is then built by sorting the nodes such that the weight are in increasing order.
           The required parameter s in [0,1] controls the strength of the distance
           (s = 0: random, s = 1: most guided by the distance - quasi deterministic).
           Notes: 1) in absence of conditioning node, one random node in the path is considered
           to compute the distances, 2) if simulation type is SIM_VARIABLE_VECTOR, variables
           exhaustively informed are not considered as conditioning data for computing distance.
      - PATH_RANDOM_HD_DISTANCE_SUM_PDF:
           The path is built as for PATH_RANDOM_HD_DISTANCE_PDF, but the distance to the set of
```

```
          conditioning nodes is replaced by the sum of the distances to every conditioning node,
          each distance being raised to a power p.
          Two parameters are required: p > 0 controls the computation of the distances,
          and s in [0,1] controls the strength of the distance
          (s = 0: random, s = 1: most guided by the distance).
          Note: 1) in absence of conditioning node, random path is considered.
      - PATH_RANDOM_HD_DISTANCE_SUM_SORT:
          The path is built as for PATH_RANDOM_HD_DISTANCE_SORT, but the distance to the set of
          conditioning nodes is replaced by the sum of the distances to every conditioning node,
          each distance being raised to a power p.
          Two parameters are required: p > 0 controls the computation of the distances,
          and s in [0,1] controls the strength of the distance
          (s = 0: random, s = 1: most guided by the distance - quasi deterministic).
          Note: 1) in absence of conditioning node, random path is considered.
      - PATH_UNILATERAL:
          It requires a vector u of parameters of size n=4 (resp. n=3) for simulation type set to
          SIM_ONE_BY_ONE (resp. SIM_VARIABLE_VECTOR), given as ux uy uz uv (resp. ux uy uz).
          Some components of u can be equal to 0, and the other ones must be the integer 1,...,m,
          with sign +/-. For
              u[j(1)] = ... = u[j(n-m)] = 0 and (u[i(1)], ..., u[i(m)]) = (+/-1,...,+/-m),
          the path visits all nodes in sections of coordinates j(1),...,j(n-m) randomly,
          and makes varying the i(1)-th coordinate first (most rapidly), ..., the i(m)-th coordinate
          last, each one in increasing or decreasing order according to the sign (+ or - resp.).
          Examples:
            - for simulation type SIM_ONE_BY_ONE, and u = (0, -2, 1, 0), then the path will visit
              all nodes: randomly in xv-sections, with increasing z-coordinate first (most rapidly),
              and decreasing y-coordinate.
            - for simulation type SIM_VARIABLE_VECTOR, and u = (0, 2, 1), then the path will visit
              all nodes: randomly in x-direction, with increasing z-coordinate first (most rapidly),
              and increasing y-coordinate.
*/
PATH_RANDOM_HD_DISTANCE_PDF 1

/* DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s)) */
0.1 0.01 0.01 0.01 0.01

/* PROBABILITY CONSTRAINTS */
/* FOR EACH VARIABLE:
   1. Probability constraint usage, integer (probabilityConstraintUsage):
        - 0: no probability constraint
        - 1: global probability constraint
        - 2: local probability constraint

   2. If probabilityConstraintUsage > 0, then the classes of values (for which the
        probability constraints will be given) have to be defined; a class of values
        is given by a union of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[;
        Here are given:
           - nclass: number of classes of values
           - for i in 1,..., nclass: definition of the i-th class of values:
                - ninterval: number of interval(s)
                - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                    these values should satisfy inf_i < sup_i

   3a. If probabilityConstraintUsage == 1, then
         - global probability for each class (defined in 2. above), i.e.
           nclass numbers in [0,1] of sum 1
   3b. If probabilityConstraintUsage == 2, then
         - one pdf image file (for every class, nclass variables)
           (image of same dimensions as the simulation grid)
         - support radius for probability maps (i.e. distance according to
           the unit defined in the search neighborhood parameters for the
           considered variable)
         - method for computing the current pdf (in the simulation grid),
            integer (localCurrentPdfComputation):
              - 0: "COMPLETE" mode: all the informed node in the search neighborhood
                   for the considered variable, and within the support are taken into account
              - 1: "APPROXIMATE" mode: only the neighboring nodes (used for the
                   search in the TI) within the support are taken into account

   4. If probabilityConstraintUsage > 0, then
        method for comparing pdf's, integer (comparingPdfMethod):
           - 0: MAE (Mean Absolute Error)
           - 1: RMSE (Root Mean Squared Error)
           - 2: KLD (Kullback Leibler Divergence)
           - 3: JSD (JSD (Jensen-Shannon Divergence)
           - 4: MLikRsym (Mean Likelihood Ratio (over each class indicator, symmetric target interval))
           - 5: MLikRopt (Mean Likelihood Ratio (over each class indicator, optimal target interval))

   5. If probabilityConstraintUsage > 0, then
        - deactivation distance, i.e. one positive number
          (the probability constraint is deactivated if the distance between
          the current simulated node and the last node in its neighbors (used
          for the search in the TI) (distance computed according to the corresponding
          search neighborhood parameters) is below the given deactivation distance)
```

```
494
495     6. If probabilityConstraintUsage > 0, then
496           - threshold type for pdf's comparison, integer (probabilityConstraintThresholdType)
497                - 0: constant threshold
498                - 1: dynamic threshold
499             note: if comparingPdfMethod is set to 4 or 5, probabilityConstraintThresholdType must be set to 0
500           6.1a If probabilityConstraintThresholdType == 0, then
501                - threshold value
502           6.1b If probabilityConstraintThresholdType == 1, then the 7 parameters:
503                - M1 M2 M3
504                - T1 T2 T3
505                - W
506              These parameters should satisfy:
507                 0 <= M1 <= M2 < M3,
508                 T1 >= T2 >= T3,
509                 w != 0.
510              The threshold value t is defined as a function of the number M
511              of nodes used for computing the current pdf (in the simulation grid)
512              including the candidate (i.e. current simulated) node by:
513                 t(M) = T1, if M < M1
514                 t(M) = T2, if M1 <= M < M2
515                 t(M) = (T3 - T2) / (M3^W - M2^W) * (M^W - M2^W) + T2, if M2 <= M < M3
516                 t(M) = T3, if M3 <= M
517  */
518  /* PROBABILITY CONSTRAINTS FOR VARIABLE #0 */
519  0
520  /* PROBABILITY CONSTRAINTS FOR VARIABLE #1 */
521  0
522  /* PROBABILITY CONSTRAINTS FOR VARIABLE #2 */
523  0
524  /* PROBABILITY CONSTRAINTS FOR VARIABLE #3 */
525  0
526  /* PROBABILITY CONSTRAINTS FOR VARIABLE #4 */
527  0
528
529  /* CONNECTIVITY CONSTRAINTS */
530  /* FOR EACH VARIABLE:
531     1. Connectivity constraint usage, integer (connectivityConstraintUsage):
532           - 0: no connectivity constraint
533           - 1: set connecting paths before simulation by successively
534                binding the nodes to be connected in a random order
535           - 2: set connecting paths before simulation by successively
536                binding the nodes to be connected beginning with
537                the pair of nodes with the smallest distance and then
538                the remaining nodes in increasing order according to
539                their distance to the set of nodes already connected.
540           - 3: check connectivity pattern during simulation
541     2. If connectivityConstraintUsage > 0, then key word for type of connectivity:
542           - CONNECT_FACE             : 6-neighbors connected (by face)
543           - CONNECT_FACE_EDGE        : 18-neighbors connected (by face or edge)
544           - CONNECT_FACE_EDGE_CORNER : 26-neigbhors connected (by face, edge or corner)
545     3. If connectivityConstraintUsage > 0, then the classes of values (that can
546           be considered in the same connected components) have to be defined; a
547           class of values is given by a union of interval(s):
548              [inf_1,sup_1[ U ... U [inf_n,sup_n[;
549           Here are given:
550              - nclass: number of classes of values
551              - for i in 1,..., nclass: definition of the i-th class of values:
552                   - ninterval: number of interval(s)
553                   - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
554                        these values should satisfy inf_i < sup_i
555     4. If connectivityConstraintUsage > 0, then:
556           - variable name for connected component label
557             (included in data image / data point set above)
558             (Note: label negative or zero means no connectivity constraint)
559     5a. If connectivityConstraintUsage == 1 or connectivityConstraintUsage == 2, then:
560           - name of the image file and name of the variable for the search of
561             connected paths (set string "_TI_" instead for searching in the (first)
562             training image and the corresponding variable index)
563     5b. If connectivityConstraintUsage == 3, then:
564           - deactivation distance, i.e. one positive number
565             (the connectivity constraint is deactivated if the distance between
566             the current simulated node and the last node in its neighbors (used
567             for the search in the TI) (distance computed according to the corresponding
568             search neighborhood parameters) is below the given deactivation distance)
569           - threshold: threshold value for acceptation of connectivity pattern
570  */
571  /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #0 */
572  0
573  /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #1 */
574  0
575  /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #2 */
576  0
577  /* CONNECTIVITY CONSTRAINTS FOR VARIABLE #3 */
```

```
0
/* CONNECTIVITY CONSTRAINTS FOR VARIABLE #4 */
0

/* BLOCK DATA */
/* FOR EACH VARIABLE:
   1. Block data usage, integer (blockDataUsage):
         - 0: no block data
         - 1: use of block data (mean value)

   2. If blockDataUsage == 1, then
         - block data file name
*/
/* BLOCK DATA FOR VARIABLE #0 */
0
/* BLOCK DATA FOR VARIABLE #1 */
0
/* BLOCK DATA FOR VARIABLE #2 */
0
/* BLOCK DATA FOR VARIABLE #3 */
0
/* BLOCK DATA FOR VARIABLE #4 */
0

/* MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s)) */
0.25

/* PYRAMIDS */
/* I. PYRAMID GENERAL PARAMETERS:
      I.1. Number of pyramid level(s) (in addition to original simulation grid, i.e. number of
           reduction operations), integer (npyramidLevel):
              - = 0: no use of pyramids
              - > 0: use pyramids, npyramidLevel should be equal to the max of "nlevel" entries
                     in pyramid parameters for every variable (point II.1 below);
                     pyramid levels are indexed from fine to coarse:
                        * index 0            : original simulation grid
                        * index npyramidLevel: coarsest level
      If npyramidLevel > 0:
         I.2. for each level, i.e. for i = 1,..., npyramidLevel:
                 - kx, ky, kz (3 integer): reduction step along x,y,z-direction for the i-th reduction:
                      k[x|y|z] = 0: nothing is done, same dimension in output
                      k[x|y|z] = 1: same dimension in output (with weighted average over 3 nodes)
                      k[x|y|z] = 2: classical gaussian pyramid
                      k[x|y|z] > 2: generalized gaussian pyramid
         I.3. pyramid simulation mode, key work (pyramidSimulationMode):
                 - PYRAMID_SIM_HIERARCHICAL:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the level one rank coarser
                 - PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                      (a) spreading conditioning data through the pyramid by simulation at each
                          level, from fine to coarse, conditioned to the level one rank finer
                      (b) simulation at the coarsest level, then simulation of each level, from coarse
                          to fine, conditioned to the gaussian expansion of the level one rank coarser
                 - PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                      co-simulation of all levels, simulation done at one level at a time
          I.4. Factors to adapt the maximal number of neighboring nodes:
              I.4.1. Setting mode, key word (factorNneighboringNodeSettingMode):
                       - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                       - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
              If factorNneighboringNodeSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                 I.4.2. The factors, depending on pyramid simulation mode:
                    - if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL
                      or PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION:
                         - faCond[0], faSim[0], fbCond[0], fbSim[0],
                           ...,
                           faCond[n-1], faSim[n-1], fbCond[n-1], fbSim[n-1],
                           fbSim[n]:
                              I.e. (4*n+1) positive numbers where n = npyramidLevel, with the following
                              meaning. The maximal number of neighboring nodes (according to each variable)
                              is multiplied by
                                (a) faCond[j] and faSim[j] for the conditioning level (level j)
                                    and the simulated level (level j+1) resp. during step (a) above
                                (b) fbCond[j] and fbSim[j] for the conditioning level (level j+1) (expanded
                                    if pyramidSimulationMode == PYRAMID_SIM_HIERARCHICAL_USING_EXPANSION)
                                    and the simulated level (level j) resp. during step (b) above
                    - if pyramidSimulationMode == PYRAMID_SIM_ALL_LEVEL_ONE_BY_ONE:
                         - f[0],...,f[npyramidLevel-1],f[npyramidLevel]:
                              I.e. (npyramidLevel + 1) positive numbers, with the following meaning. The
                              maximal number of neighboring nodes (according to each variable) is
                              multiplied by f[j] for the j-th pyramid level.
          I.5. Factors to adapt the distance threshold (similar to I.4):
              I.5.1. Setting mode, key word (factorDistanceThresholdSettingMode):
```

```
                        - PYRAMID_ADAPTING_FACTOR_DEFAULT: set by default
                        - PYRAMID_ADAPTING_FACTOR_MANUAL : read in input
                If factorDistanceThresholdSettingMode == PYRAMID_ADAPTING_FACTOR_MANUAL:
                   I.5.2. The factors, depending on pyramid simulation mode (similar to I.4.2).

         II. PYRAMID PARAMETERS FOR EACH VARIABLE:

         II.1. nlevel: number of pyramid level(s) (number of reduction operations)
                         - = 0: no use of pyramid for the considered variable
                         - > 0: use pyramids for the considered variable, with nlevel level

         If nlevel > 0:
            II.2. Pyramid type, key word (pyramidType):
                      - PYRAMID_CONTINUOUS        : pyramid applied to continuous variable (direct)
                      - PYRAMID_CATEGORICAL_AUTO  : pyramid for categorical variable
                                                      - pyramid for indicator variable of each category
                                                        except one (one pyramid per indicator variable)
                      - PYRAMID_CATEGORICAL_CUSTOM: pyramid for categorical variable
                                                      - pyramid for indicator variable of each class
                                                        of values given explicitly (one pyramid per
                                                        indicator variable)

            If pyramidType == PYRAMID_CATEGORICAL_CUSTOM:
               II.3.  The classes of values (for which the indicator variables are
                  considered for pyramids) have to be defined; a class of values is given by a union
                  of interval(s): [inf_1,sup_1[ U ... U [inf_n,sup_n[.
                  Here are given:
                     - nclass: number of classes of values
                     - for i in 1,..., nclass: definition of the i-th class of values:
                          - ninterval: number of interval(s)
                          - inf_1 sup_1 ... inf_ninterval sup_ninterval: inf and sup for each interval
                               these values should satisfy inf_i < sup_i
                  Then, for each class, the number of pyramid levels (number of reduction operations) is
                     - nlevel_i (for i in 1,..., nclass)
*/
/* PYRAMID GENERAL PARAMETERS */
0

/* TOLERANCE */
/* Tolerance t on the threshold value for flagging nodes (for post-processing):
   let d(i) be the distance between the data event in the simulation grid and in the training
   image for the i-th variable and s(i) be the distance threshold for the i-th variable, and let
   e(i) = max(0, (d(i)-s(i))/s(i)) be the relative error for the i-th variable, i.e. the relative
   part of the distance d(i) beyond the threshold s(i); during the scan of the training image, a node
   that minimizes e = sum (e(i)) is retained (the scan is stopped if e = 0); if e is greater than the
   tolerance t (given here), then the current simulated node and all non-conditioning nodes of the
   data events (one per variable) in the simulation grid are flagged for resimulation (post-processing).
   Note that if probability / connectivity / block data constraints used a similar error as e(i) that
   contributes in the sum defining the error e.
*/
0.01

/* POST-PROCESSING */
/* 1. Maximal number of path(s) (npostProcessingPathMax)
   2. If npostProcessingPathMax > 0:
      key word for post-processing parameters (i. e. number of neighboring nodes, distance threshold,
      maximal scan fraction, and tolerance):
         - POST_PROCESSING_PARAMETERS_DEFAULT: for default parameters
         - POST_PROCESSING_PARAMETERS_SAME   : for same parameters as given above
         - POST_PROCESSING_PARAMETERS_MANUAL : for manual settings
   3. If npostProcessingPathMax > 0 and POST_PROCESSING_PARAMETERS_MANUAL:
         MAXIMAL NUMBER OF NEIGHBORING NODES FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL DENSITY OF NEIGHBORING NODES IN SEARCH NEIGHBORHOOD FOR EACH VARIABLE (as many number(s)
            as number of variable(s))
         DISTANCE THRESHOLD FOR EACH VARIABLE (as many number(s) as number of variable(s))
         MAXIMAL SCAN FRACTION FOR EACH TI (as many number(s) as number of training image(s))
         TOLERANCE
*/
1
POST_PROCESSING_PARAMETERS_DEFAULT

/* SEED NUMBER AND SEED INCREMENT */
4565541
1

/* NUMBER OF REALIZATION(S) */
1

END
```