Enhanced Geologic Modeling with Data-Driven Training Images for Improved Resources and
Recoverable Reserves

by

Daniel Andrés Silva Maureira

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

Doctor of Philosophy

in

Mining Engineering

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Deterministic geologic modeling methods accurately characterize large-scale continuous features of geological phenomena, but often fail in reproducing their inherent short-scale variability. The opposite is the case with stochastic methods that lack large-scale continuity yet contain reasonable short-scale variability. Both methods are limited in their ability to account for a balanced amount of geologic variability. Integrating both large and short scale geologic elements properly improves the prediction of mineral resources and reserves.

This thesis develops a methodology that improves the characterization of the geologic variability in mineral deposits. The central idea is to combine deterministic and stochastic geologic interpretations and transfer the essential geological features into geostatistical models. The multiple point statistics (MPS) simulation method is suitable for this task. This technique utilizes training images for extracting and then reproducing complicated geomorphological features in the models. The method has been adapted to integrate information from different images. Generally, training images are designed based on conceptual models of the geologic phenomena; in this work, deterministic and stochastic geologic representations are used as data-driven

training images, one comes from modeling the categories by an implicit geologic approach, and another comes from the application of conventional sequential indicator simulation (SIS) method. Such data-driven training images permit a robust inference of spatial structure from reasonably spaced drillhole data.

This work establishes the principles to integrate multiple training images through a scheme of data integration for conditional probabilities known as a linear opinion pool. A methodology for calibrating the contribution of each training image is developed based on the variability at the available drillholes. A measure of multi-point entropy along the drillholes is matched by the combination of the two training images. The resulting calibrated models integrate geologic features from both training images, reproducing the correct underlying continuity and variability of the deposit, and reducing misclassified ore/waste material. Practical implementation of the methodology shows improvement in the predicted profit relative to classical geostatistical approaches.

Preface

This dissertation is an original work by Daniel A. Silva and is submitted for the degree of Doctor of Philosophy in Mining Engineering at the University of Alberta. All material covered was conducted under the supervision of Professor Clayton V. Deutsch in the Department of Civil and Environmental Engineering.

Part of the research documented in this thesis led to the publication of two articles. Most of the published material relates to advancements described in chapter 4 and chapter 6. The first article provides the foundations for integrating multiple data-driven training images for mineral deposits, and it was published as D.A. Silva and C.V. Deutsch, A Multiple Training Image Approach for Spatial Modeling of Geologic Domains, *Mathematical Geosciences*, Volume 46, Issue 7, pp 815-840. The second article examines the practicability of the proposed method using operational data from a mining project; this material may be found as D.A. Silva and C.V. Deutsch, *CIM Journal*, Vol. 6, No. 3, pp 137-148.

To Luz

Acknowledgments

I would like to thank my supervisor, Dr Clayton V. Deutsch, for his guidance, support and help. Without him this would have not been possible. I appreciate the time spent in our meetings, and the kindness with he contagious the knowledge for the theory and practice of geostatistics. More than a professor, he is a mentor who pushed me to go beyond my intellectual limits.

My gratitude also goes to the Government of Chile for funding my studies, and providing the financial support for a quiet and nice stay in Canada. Thank you to the industrial affiliates of the Centre for Computational Geostatistics for supporting the research developed.

I would like also to express my special acknowledgment to Miguel Cuba for welcoming during my first days here, and for those long hours of discussions. I appreciate my conversations with Saina, and the partnership promoted by Jared. Thank you to Jianan, Felipe, Yevgeniy and Diogo for making my stay in the lab enjoyable.

I have to say something about the wonderful people that I met during these four long years, and that made to Luz and me to feel like at home. Thank you to Matias, Vanesa, Memo, Keka, Andres, Paulina, Jose, Paola, Jos and those wonderful little guys Josefina, Renatito and Isabella that made us remember the simple things.

I thank the infinite love of my parents and sister; for being an inexhaustible source of support and wisdom. Finally, I do not have how to appreciate the courage, strength, motivation, and enthusiasm of my beloved Luz; thank you for your love, support, patience and comprehension. This thesis is dedicated to you.

Contents

List of Tables

List of Figures

xvii

List of Symbols, Nomenclature and Abbreviations

$2\gamma(\mathbf{h})$	variogram function
γ	power factor
$\lambda_\beta^{OK}(\mathbf{u}; k)$	ordinary kriging weights
$\lambda_\alpha^{SK}(\mathbf{u}; k)$	simple kriging weights
$\lceil \cdot \rceil$	ceiling function
μ	servosystem correction factor
$\phi(x_1, ..., x_n) = 0$	explicit function expression
A	area
A_0	indicator binary random variable of the state s_k of random variable S
B_{Slope}	combined measure of accuracy, precision and predictive capability
$c(D_{m,n})$	combined number of replicates of data event $D_{m,n}$ in node of search-tree for a set of training images
$c^m(D_{m,n})$	total number of replicates of data event $D_{m,n}$ in node of search-tree of training image m
$C_I(\mathbf{h})$	indicator covariance function
$c_{s_k}^m(D_{m,n})$	node of search-tree with number of replicate of data event $D_{m,n}$ in training image m
$Cov\{Z(\mathbf{u}_1), Z(\mathbf{u}_2)\}$	covariance function between random variables $Z(\mathbf{u}_1)$ and $Z(\mathbf{u}_2)$
D	indicator binary random variable of the joint set of occurrence at n locations

$d(\mathbf{u}_\alpha)$	signed distance function at location $\mathbf{u}_\alpha$
$d^*(\mathbf{u})$	signed distance function ordinary kriging estimate
$d_k(\mathbf{u}_\alpha)$	signed distance vector of k categories
$d_k^*(\mathbf{u})$	signed distance ordinary kriging estimate vector of k categories
D_n	data event of n conditioning data
$D_{m,n}$	data event of n conditioning data in training image m
$data_{spacing}(\mathbf{u})$	data spacing measure
$E\{\cdot\}$	expected value of a random variable
$f^l(\mathbf{u})$	polynomial of degree l
$F_Z(\mathbf{u}; z)$	cumulative distribution function of the random variable $Z(\mathbf{u})$
$F_{Z(\mathbf{u}_1),....,Z(\mathbf{u}_n)}(z_1,....,z_n)$	spatial law of n ramdom variables
$H_{drillholes}$	entropy of drillhole campaign
$H_{TraininImage}$	entropy of training image
$i(\mathbf{u}; k)$	indicator conditioning data at location $\mathbf{u}$ of category k
$I(\mathbf{u}; k)^*$	indicator simple kriging estimate
$I(\mathbf{u}; k)$	indicator random variable at location $\mathbf{u}$ of category k
$i^*(\mathbf{u})$	inside/outside indicator of signed distance estimate
$i_k(\mathbf{u}_\alpha)$	indicator vector of k categories
m	index of training image
$m(\mathbf{u})$	trend
n_d	number of drillholes
$n_{samples}(\mathbf{u})$	number of surrounded samples
n_{sd_α}	number of replicates of drillholes
O	value of objective function
O_m	multiple point statistics component of objective function
O_p	local proportion component of objective function

O_{tp}	two-point transition probability component of objective function
p	symmetric interval
$P(i(\mathbf{u}) = k)^{updated}$	corrected probability after servosystem
$p(k)$	target global proportion
$p_k^c(\mathbf{u})$	current global proportion
$P^{Local}(\mathbf{u}; k)$	local proportion at location $\mathbf{u}$
$P^{MPS}(\mathbf{u}; k$	MPS conditional probability
$p_{i,j}$	probability of template configuration j in drillhole i
$R(\mathbf{u})$	stationary residual
$r(\mathbf{u}_\alpha)$	range field
$S(\mathbf{u})$	indicator random variable of k categories
s_k	categorical states of random variable of S
$S_{sim,TI}$	similarity measure between training images
$sgsim$	sequential Gaussian simulation GSLIB program
V	volume
$Var\{\cdot\}$	variance of random variable
w_i	weight assigned to training image i
$w_{DTI}(\mathbf{u})$	weight assigned to deterministic training image
w_{DTI}^c	calibrated weight assigned to deterministic training image
XY	xy plane
XZ	xz plane
YZ	yz plane
$Z(\mathbf{u})$	random variable at location $\mathbf{u}$
$z(\mathbf{u})$	outcome of the random variable $Z(\mathbf{u})$ at location $\mathbf{u}$
$Z_\lambda(\mathbf{u})$	generalized increment
$caliti$	calibration training image FORTRAN program

ellipsim	ellipsoid simulation GSLIB program
p-field	probability field simulation
snesim	single normal equation simulation
softmax	multiple generalized regression model
ANSI	american national standards institute
B	difference of averaged probability when true and false
BA	baritic rock
C	measure of certainty of probabilistic model
CAD	computer-aided-design methods in computer graphics
CPU	central processing unit
DTI	deterministic training image
FLTI	fluvial training image
ft	feet
GSLIB	geostatistical software library
H	entropy
HISTC	high iron siliceous
HRU	overlying waste rock
LIS	low iron siliceous
LOP	linear opinion pool scheme
M	million of dollars
MPS	multiple point statistics simulation
OF	objective function
RBF	radial basis function
RF	random function
RFLTI	rotated fluvial training image
RV	random variable
SGS	sequential Gaussian simulation
SIS	sequential indicator simulation

SMU	selective mining unit
STI	stochastic training image
STRC	search-tree replicates combination algorithm
TGS	truncated Gaussian simulation
TI	training image
Zn	zinc grade

Chapter 1

Introduction

Economic appraisal of mineral deposits depends on a series of technical decisions. To give engineering support to such decisions, numerical models that characterize and quantify the subsurface geology are required.

Information from the subsurface is limited, and uncertainty in these geological models is inevitable. Geostatistics provides reputable techniques for modeling the geology and quantifying its uncertainty. The characterization of uncertainty is possible thanks to the generation of a set of realizations that honor the subsurface information and the spatial correlation of the geologic phenomenon.

Advances in Geostatistics should aim the development of methodologies that improve the realism of numerical models accounting for complicated morphological geologic features. These advances would help construct enhanced geologic models that lead to improved mineral recoverable resources prediction, better risk assessment and more informed models for decision-making.

1.1 Motivation

The development of mining projects is the product of a series of engineering analyses that aim to determine the economic feasibility of mineral extraction. Figure 1.1 shows the stages involved in the economical appraisal of mining projects in the format of a chain value diagram (Porter, 2008). Seven major stages are identified:

1

(1) exploration, (2) resource estimation, (3) mine planning, (4) mine construction, (5) mine operation, (6) mineral processing and (7) commercialization. Each stage requires the application of distinct fields of engineering and geosciences. The current thesis focuses on the importance of improved geologic modeling methods for the stage of mineral resource estimation, and its effects on economic valuation throughout the mining chain of value.

The resource estimation stage consists of numerical quantification of the mineral grades through the implementation of geostatistics techniques. These models are required at early stages of appraisal for reliable mineral prediction; otherwise, biased results will influence subsequent assessments affecting the basis for economic decision-making perhaps leading to expensive mistakes.
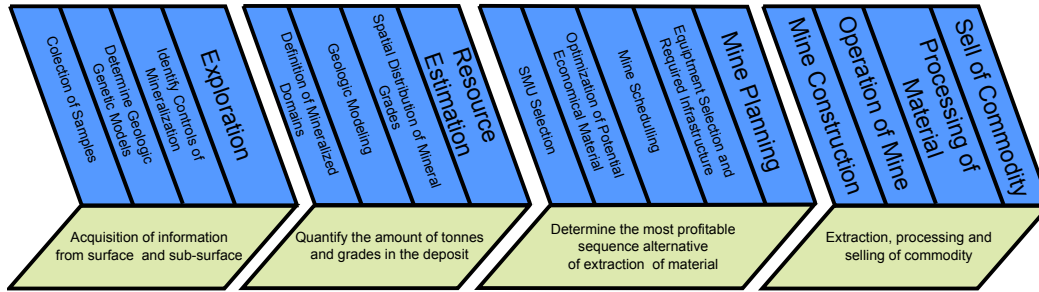
Figure 1.1: Schematic illustration of the chain of value in mining industry (Garcia and Camus, 2011).

Geologic modeling has a significant influence on predicted future profits (Dimitrakopoulos, 1998). The estimation of recoverable resources and reserves is the input upon which mining technical decisions for the extraction and processing of material are made. These engineering definitions are known as mine planning and must establish the selection and design of the mining method, alternatives for processing treatment, metallurgical recovery functions, suitable assumptions on external factors such as metal and energy prices, along with operational costs and capital investment. As a result, a complete year-by-year schedule is constructed to establish the amount of material to be removed, the plan determines the economic mining limits and the optimal sequence of extraction based on the geostatistical models.

1.1.1 Selective Mining Unit (SMU) Size Selection

The link between geostatistical models and mine planning is established through the definition of a selective mining unit (SMU). The SMU corresponds to the minimum volume to practically separate the ore from the waste material. The selection of the size of the SMU will affect predictions of dilution and lost ore by internally mixing grades (internal dilution) and combining the material from distinct domains (external dilution). Figure 1.2 illustrates this using a schematic illustration of a gold vein structure. The coarsest grid corresponds to the SMU and the finest grid to the block model distribution of gold grades within the vein. Usually, the SMU size is at a scale of 5 to 30 meters. Even though the example is based on a centimeter scale, it is intended to illustrate mining conditions and demonstrate the importance of dilution. Internal dilution refers to the mixing of high and low grades within the vein. The reason is that the spatial distribution of gold grade is not homogeneous due to different initial and boundary conditions at the time of deposition. External dilution also known as geologic contact dilution, occurs when dealing with locations near to domain boundaries. It depends on the same parameters as internal dilution, but also on the geometric shape of the boundaries. The type of geologic contact dilution affects the estimation of recoverable resources and reserves.
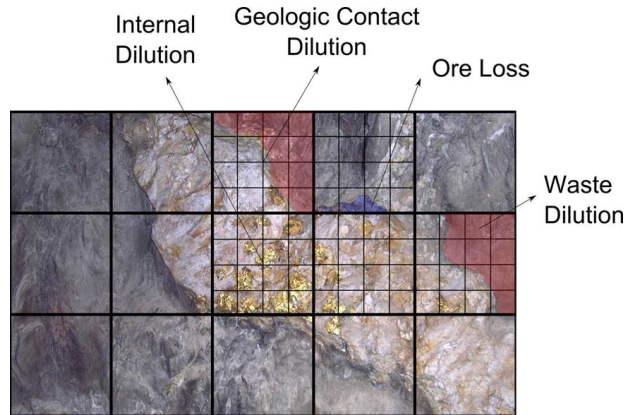
Figure 1.2: Schematic representation of internal and geologic contact dilution.

Dilution and lost ore bring associated costs due to misclassified ore/waste material. As the size of the SMU increases, the cost of dilution and lost ore increases.

The second element in the analysis corresponds to the cost of the extraction process. Less selectivity, that is, larger block size reduces the mining cost due to economies of scale. The selection of SMU must be optimized based on these two factors: dilution and lost ore, and total costs. Figure 1.3 shows a schematic representation of the costs as a function of the SMU size. An optimum SUM size minimizes the overall mining cost of the project represented by the green curve which correspond to the sum of the dilution and mining costs. Indeed, the determination of the SMU size is a heuristic decision made by experienced geomodelers and mining engineers.
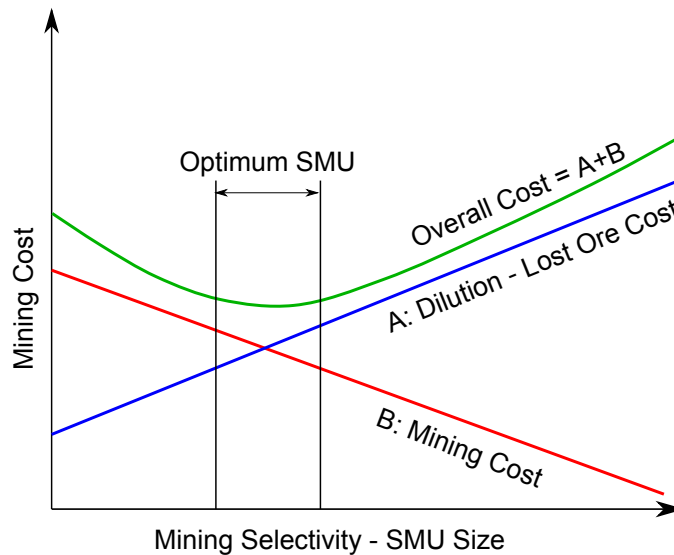
Figure 1.3: Schematic illustration of dilution-ore lost cost and mining cost as function of the SMU size. A minimum total cost criteria establishes its correct size.

The dilution-lost ore curve depends on the characteristic of the modeled geologic contacts and the requirement for mimicking future ore control practices. The first is the result of the geostatistical technique used; whereas the second is required given the fact that the final mining is not performed on SMU scale, but on grade control polygons. These polygonal areas define dig-limits where the mining is developed searching for optimum profitability considering operational criteria. In consequence, the selection of SMU must account for the grade control process in the sense that the estimated tonnage and grades of ore should be close to the future practice of grade control. The cost curve is based on design factors such as the mining extraction

method, the size of equipment, blastholes spacing, mine operating conditions and productivity.

As a result, the size of the SMU along with the variability of geologic contacts and the spatial distribution of grades within these domains influence the estimation of the dilution or lost ore. The character of the contact dilution and the distribution of grades are critical aspect to be considered and reproduced in the geostatistical models.

1.1.2   Mineral Resource Forecasting

Many geostatistical models consider of a two-step hierarchical implementation. First, the boundaries of these domains are spatially established, and secondly, the spatial distribution of grades is modeled within each domain. In practice, the geostatistical models infer the mineral grades on a grid that discretizes the geologic interpretations of mineralized areas, and whose dimensions are determined by the selected SMU size.

Deterministic methods are commonly used in the mining industry for modeling geologic domains (Cowan et al., 2003; Houlding, 1991; Silva and Deutsch, 2012; Wilde and Deutsch, 2011). Despite the fact that the results do not account for the spatial uncertainty, these methods have a large influence on resource/reserves estimation. There are geostatistical methods that represent the geologic phenomena as a stochastic process with short-scale variability and a measure of uncertainty. Such methods include Sequential Indicator Simulation (SIS) (Journel, 1983), pluri-Gaussian simulation (Galli et al., 1994), object-based modeling (Bridge and Leeder, 1979), and more recently, surface-based modeling (Pyrcz et al., 2005). These methods provide equally probable representations of the geologic phenomena. They are mostly limited to relatively simple two-point statistics and do not always capture the complex variability of geologic scenarios. Some geologic depositional environments are not well represented by these techniques. Moreover, the simulated models may appear excessively unstructured, and it is not straightforward to control transitions between domains (Deutsch, 2006). The resulting models may have an increased level

of randomness along the boundaries and may be unrealistic.

After the definition of the geologic model, the inference of the spatial distribution of mineral grades may be carried out by a random function characterized by a histogram and variogram (Journel and Huijbregts, 1978). Then, simulated realizations of both geologic domains and mineral grades offer several realizations of the phenomenon. The set of equally probable realizations allow assessing the local and global uncertainty in models. Transfer functions may be applied to the set of realizations and distinct response variables characterized.

1.1.3 Importance of Geologic Domains Variability in Mineral Resource Forecasting

To understand the influence of the SMU selection and the character of the geologic limits on dilution, a schematic illustration is presented in Figure 1.4. The ordinate axis corresponds to geologic models with different variability along the contacts; each model reflects an increasing randomness on the boundaries yet the models are based on the same data. Two geologic domains are considered (black and white) and the grades estimated at point scale - the black domain represents the spatial extension for high grades and the white (host rock) for low grades. One may be more selective and design the extraction process based on a small SMU, or increase the mine productivity by using a large SMU; the abscissa axis shows different SMU size. The high and low grades are averaged according to the SMU, accounting for internal and external dilution. The colored surface shows the relation between models with increasing contact variability and the influence of the size of the SMU. It is clear that increasing variability and SMU support augment the dilution. In particular, as the geologic variability increases so does the dilution at whatever SMU support. The randomness of the boundaries increases the mixing of high and low grades. For the SMU selection, the dilution increases with the size of the SMU since more material is being mixed.

Characterization of different levels of geologic variability are shown on the abscissa axis of Figure 1.4 may not be obtained by current deterministic and geo-
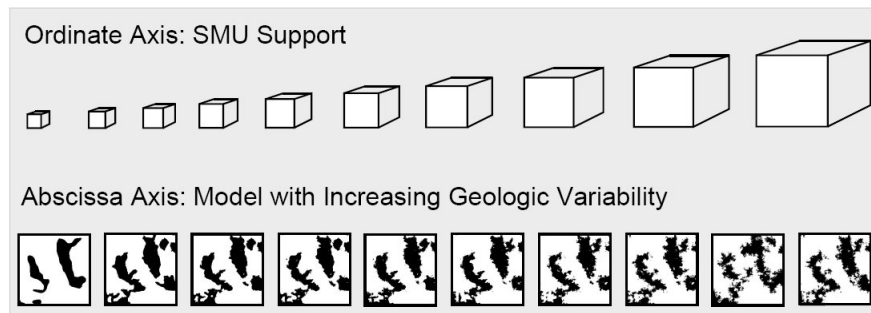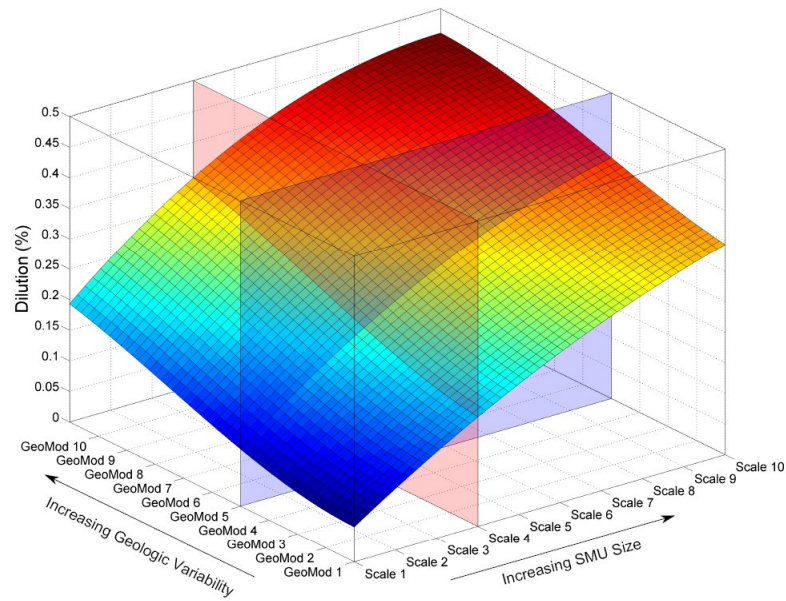
Figure 1.4: Response of the dilution due to the size of the selective mining unit (SMU) and the geologic variability in the models.

statistical approaches. To expand on this point, consider the example based on a vein structure shown in Figure 1.5. To simplify the analysis, an image processing filtering technique allows generating a binary scale map of the vein and appreciating the nature of the geologic variability of the contact (image in the middle). Both deterministic and stochastic modeling alternatives were implemented to illustrate the shortcomings related to reproduction of geologic contact variability. Neither the deterministic method nor recent developments in geostatistical simulation offer a realistic representation of the geologic domains if compared to the real vein.
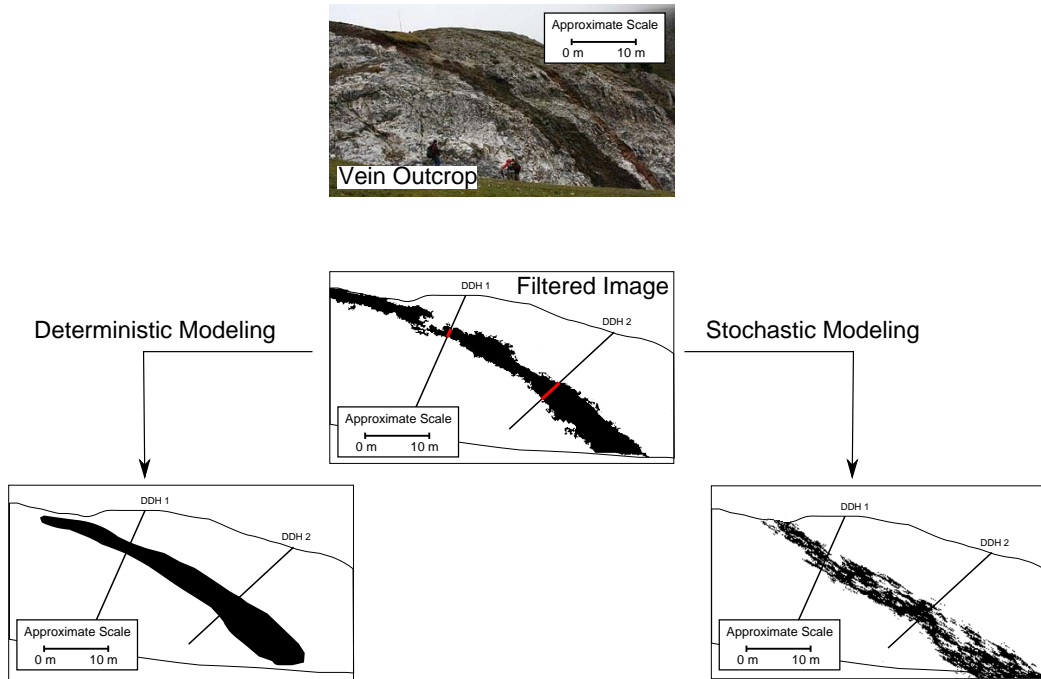
Figure 1.5: A vein outcrop and its filtered binary scale image at the top. Interpreted vein models after applying deterministic and stochastic geostatistical techniques.

The deterministic model is overly continuous, whereas the geostatistical one has an exaggerated random component.

The economic impact on the evaluation of mineral resources by the an inadequate modeling of the geologic variability may be substantial (Richards and Sides, 1991; Srivastava, 2005) leading to biased predictions in the mine plan, unexpected cost in mining operations and deviations from expected revenues . Figure 1.6 shows a schematic illustration of the effect on the distribution of grades contained within the vein when using both approaches, the deterministic model (represented in green) and the indicator stochastic approach (red model). At the top, the colored geologic models compared to the true vein. In the middle, an illustration of the material being mixed for a fixed selective mining unit (SMU) support, that is, the nature of the contact within the block determine the degree of dilution or lost ore. At the bottom, the distribution of mineral grade at SMU support for the geologic models and the true vein. The painted area below the curves represent the proportion of material above a relatively high cutoff grade. When considering the deterministic

8
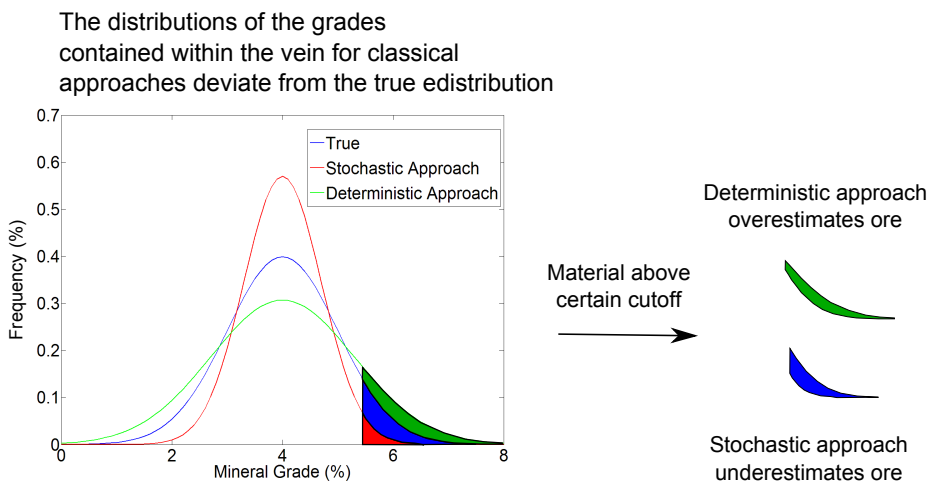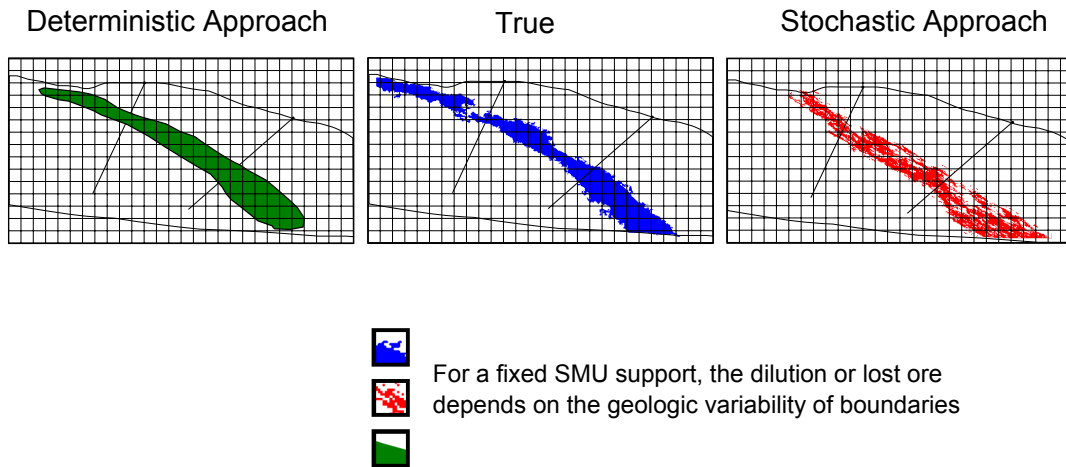
Figure 1.6: Illustration of the consequences for mineral resource estimates due to the lack of geologic realism in the models.

geologic model, the green distribution of grades results more variable than the true model (blue) since the mixing of high and low grade is less pronounced. Thus, the green area below the curve corresponds to the material misclassified as ore above the economic cutoff. The opposite happens for the stochastic geologic model, the distribution has less variance than the true model due to excessive mixing of high and low grades. The distribution based on the stochastic geologic model would overestimate the dilution reporting lower quantities of ore. The blue area corresponds to the lost ore material that was misclassified as waste. Differences make be significant depending on SMU size and the modeling method used.

1.1.4 Problem Statement

Geostatistics has been introduced as a tool for geologic modeling and uncertainty characterization of mineral deposits. The need for geostatistical methods that account for the correct level of geologic variability has been motivated. The thesis addresses the problem of incorporating reasonable large/short scale variability into numerical geological models. Reproducing the correct variability in the models would improve engineering decision-making, specifically:

> **Thesis statement:** *Development of an approach for capturing realistic large and short scale spatial variability of geologic domains will improve the prediction of resources and reserves.*

1.2 The Approach

The thesis proposes the development of a methodology that improves the realism of the geologic contact variability in mineral deposits. Large and short scale spatial structures may be reproduced by incorporating two or more feasible geologic interpretations that portray different contact continuity. The treatment of the correct reproduction of geologic variability calls for the use of geostatistical techniques able to integrate information of different sources beyond conventional geomodeling methods.

The solution to this problem requires modeling complex geological features. In the past, simulation techniques based on linear relationships and the benefits of the Gaussian distribution were used (Deutsch and Wang, 1996; Goovaerts, 1997; Journel and Huijbregts, 1978). Nowadays, modeling complex features and patterns goes beyond linear relationships. Multiple point statistics (MPS) simulation (Guardiano and Srivastava, 1993; Strebelle and Journel, 2000) is an emerging technique developed to overcome limitations of conventional variogram-based simulation. High order statistics are used to introduce complicated features and patterns into the simulated realizations.

These algorithms utilize training images for extracting and transferring essential geological patterns to the simulated realizations. The training image replaces the spatial tools based on two-point statistics such as variogram and covariance functions and relies on the design of an explicit conceptual model of the geologic phenomena.

The multiple point statistics concept is utilized as the platform for the integration of two or more conceptual geologic models. What differentiates the algorithm developed in this thesis from the standard multiple point applications is the inclusion and combination of data-driven training images to balance the large scale structures of continuity adding the correct random short scale resolution in mineral deposits. The overall methodology involves extensive numerical modeling with validation and appraisal based on a real case study with exhaustive data. Advantages and drawbacks are discussed.

1.3   Dissertation Outline

Chapter 2 discusses the theoretical and practical background of geologic modeling methods. Deterministic and geostatistical approaches that support this thesis are reviewed. Explicit and implicit deterministic alternatives are described. Background on classical indicator-based and truncated Gaussian methods provide a perspective of the limitations of current geostatistical techniques. Discussion on MPS demonstrates its potential for being used in mineral deposits. The use of data-driven training image requires discussion on aspects such as stationary, ergodicity and definition of stochastic random functions. The foundations of the stochastic processes provides the theoretical background for their implementation.

In **Chapter 3**, a novel implicit deterministic geologic modeling technique is proposed and implemented. The method consists of an extension of the binary signed distance function. The objective is to devise a geologic modeling method to quickly and efficiently construct geological models without the need for time consuming manual interpretation approaches. The distance function methodology is based on the interpolation of a distance measure. The proposed extension may

11

deal with a large number of domains and introduce a heuristic measure to assess the related uncertainty. A practical implementation to correct global proportions is also developed. The accuracy of the methodology is assessed by a case study and examples.

Chapter 4 presents the theoretical support for combining training images based on a weighted linear data integration scheme. Advantages of this additive aggregation method for integrating data-driven training images are discussed. Classical multiplicative aggregation methods do not work well with probabilities from multiple training images. Integrating the essential features of the deterministic and stochastic training image linearly provides a useful framework to obtain more realistic models. A workflow based on an entropy measure from drillholes establishes a procedure to calibrate the geologic variability from the two end member models.

Chapter 5 presents alternatives for integrating data-driven training images using the linear aggregative scheme for a non-stationary framework. Modifications to the original algorithm allow obtaining models that extend the spectrum of features from the training image. The modifications generate patterns not contained in the training images. The geologic variability of the models may be extended without losing the essential features of the training images.

Chapter 6 shows a case study based on data from an actual mining project. Verification and confirmation of the methodology is evaluated using exhaustive information from the producing mine. Finally, **Chapter 7** discusses advantages and drawbacks of the proposed methodology, as well as conclusions and future avenues of research for multiple point statistics techniques in the field of mineral deposit modeling.

Chapter 2

Geologic Modeling: Concepts and Algorithms

Prior to the application of geostatistical techniques, the geological controls of a deposit - commonly referred to as domains - must be understood and characterized in terms of their physical, chemical and mineral properties. Geologic characterization of domains is based on the interpretation of a series of geologic events identified by mineralizing solution, rock type composition, alteration, and structural constraints. A conceptual genetic geologic model supports the interpretation process by providing a context for the local conditions of how the deposit was generated. Interpreting the spatial continuity of these domains categories constitutes a critical step in the resource estimation process: the extent, shape, volume and uncertainty will have a large influence on the mineral resources.

2.1   Characterization of Domains

Mineral resources estimation is improved with a sound knowledge of the mineral deposit formation. Geologic conceptual models are developed to identify and describe potential mineralized zones. The interpretative process aims to focus the analysis and modeling efforts on certain domains within the deposit. The conceptual models helps to translate the geologic mineralizing principles into geostatistical models.

Areas of high and low grades commonly arise due to different geologic events and may be considered separate populations from a statistical point of view. Geologic domains correspond to a simplification of the mineralizing process by pooling data into a number of statistical populations. Their spatial influence must be modeled by considering the conceptual model and the distribution of data over the mineral deposit. Domains must be large enough to contain a meaningful number of data, but should not be so large as to mask important features.

2.2 Deterministic Geologic Modeling

Advances in computer graphics, information technology and computer processing have allowed the implementation of 3-D computer-aided-design (CAD) for deterministic geologic modeling. The first CAD software packages, traced back to the early 80s, allowed generating wireframe curves and surfaces of objects based upon parametric methods (Bezier, 1967, 1974), Karonen and Akatemia (1985) and Bak and Mil (1989). Since then, mining software packages have improved in terms of wireframing and displaying of 3-D geologic models. These methods are sometimes called explicit geologic modeling.

2.2.1 Explicit Modeling

The classical CAD approach starts by digitizing 2-D polygons on cross-sectional and plan views to define the shape of the geometric components of the 3-D model. The digitalization process is time-consuming since all the polygons must be similar on all sections. Then, the wireframing process of connecting polygons generates 3-D objects. This step is critical because the resulting 3-D object should have a consistent geometry in all directions.

Figure 2.1 illustrates the modeling workflow; a simple binary case is considered. The set of cross-section and plan views in A and B describe the spatial extent of the domain. These sections and plans are triangulated to generate a reasonable geologic object (image C). In the case that multiple domains are being modeled, the process
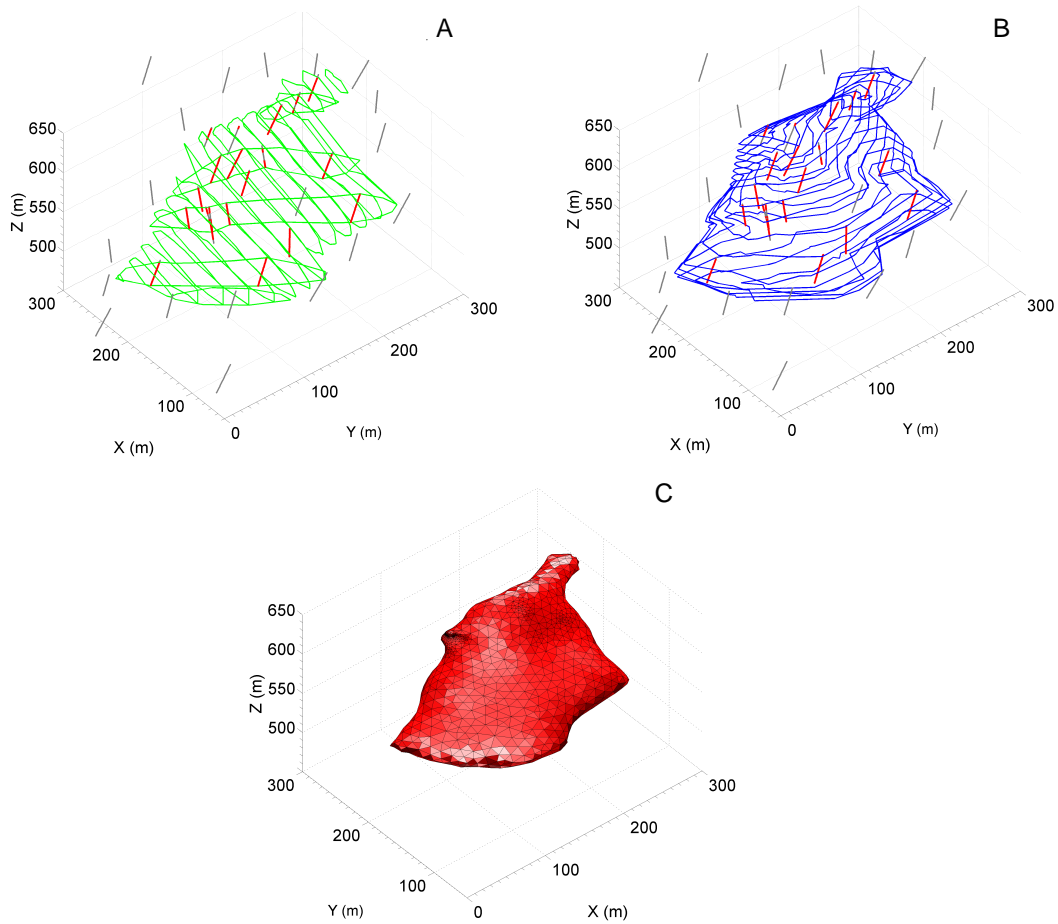
Figure 2.1: Representation of the stages involved in the explicit geologic modeling workflow. In A, the interpreted polygons on cross sections are presented. The plan views are illustrated in B, while the resulting wireframed of the geologic domain is presented in C.

would be repeated for each domain in hierarchical order. One of the disadvantages of manual digitalization of polygons is that increases the chances to overlapping and generates inconsistencies between different objects causing problems when triangulating. Other disadvantages include (1) significant professional interpretation time required, and (2) subjectivity and non-repeatability.

2.2.2   Implicit Modeling

Mining software packages have provided computational tools to display drillhole data and to speed the manual digitalization of cross sections and plan views. De-

15

spite these advances, explicit modeling is still a tedious process. Furthermore, the explicit approach lacks flexibility, reproducibility and is difficult to audit by external personnel. Recently, novel techniques, referred to as implicit modeling, provide algorithms that reduce the level of subjectivity by replacing the digitalization process with some form of automatic estimation.

Volume Function

Cowan et al. (2003) introduced a novel grid-free approach to interpolate a volume function for geologic modeling. The method is based on the work of Savchenko et al. (1995) for modeling objects using interpolated volume functions. The volume function defines the volume of 3-D objects and is constructed by implementing a fast scattered interpolator method known as Radial Basis Function (RBF) (Franke, 1982; Hardy, 1990). The interpolator is represented as the linear combination of basic functions similarly as dual kriging (Journel, 1988). However, these basic functions are standard and not inferred from the data as the case of the dual kriging. Thus, the volume function is defined throughout the space at any resolution. The triangulation of the function constructs the surfaces of the objects.

The advantage of implicit modeling is offering rapid 3-D interpretation of geologic domains due to the dual formalism of the interpolator. It is not necessary to digitalize a complete set of polygons. Also, the parameterized volume function reduces subjectivity. A sensitivity analysis could be rapidly performed with different parameters. However, there are still issues with the results: 1) the interpolated domains surfaces are smooth, 2) not reproduce trends and particular patterns that trained professional might understand, and 3) the method does not account for uncertainty in the boundaries.

Signed Distance Functions

The distance function methodology is an implicit modeling technique based on the interpolation of signed distance functions among the data locations (McLennan and Deutsch, 2006; McLennan, 2007; Hosseini, 2009). The signed distance function

measures the separation of domains with respect to each other. Anisotropic distance may be applied depending on the orientation, geologic shape and extension of the domains. Positive and negative distances distinguish between inside/outside domains. Interpolation methods including kriging could be used to estimate the signed distance function values at unsampled locations; then, a zero cutoff rule sets the boundaries between domains. The method offers repeatable geologic models with smooth boundaries. A global search strategy with kriging avoids artifacts due to relatively few data. Figure 2.2 illustrates the algorithm with a 3-D example. On the left, the red samples within a particular domain; on the right, a triangulated surface at zero cutoff.
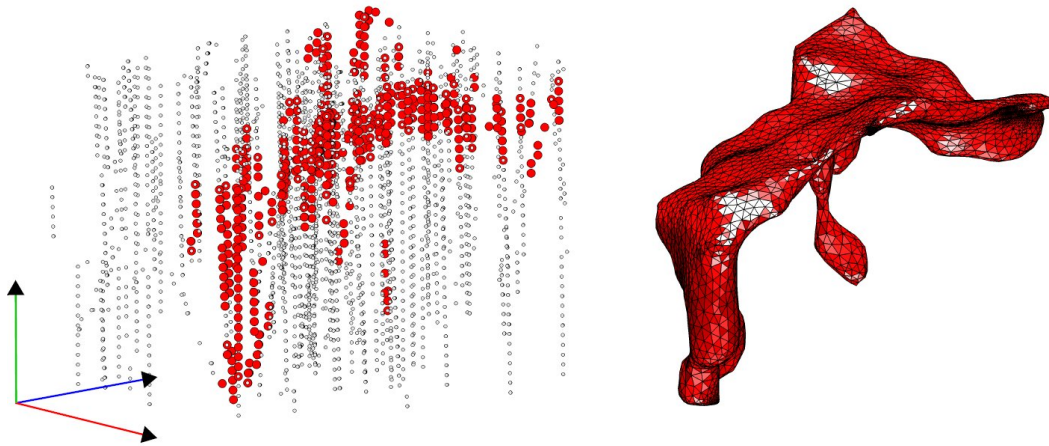
Figure 2.2: An example of a geologic model based on the interpolation of signed distance function values. On the left, the spatial configuration of samples; on the right, the resulting wireframed model after interpolation is performed.

Among the disadvantages of the method, it does not directly account for volume uncertainty, does not work well if there are short scale variations not intersected by drillholes, and is not suited to more than two domains. Extended details of the algorithm will be reviewed in Section 3.2.

Explicit and implicit modeling tend to converge to similar geologic representation when closely spaced drillhole information is available. Significant differences might occur at an early stage of exploration when the data is scarce and the assumptions made by the modeler become the most important element for domain

characterization. Further details on the practical application to multiple coal seams may be found in Deutsch and Wilde (2013).

2.2.3 Disadvantages of Deterministic Geologic Modeling

Deterministic geologic models do not quantify the uncertainty, that is, a single deposit model for mineral resources estimation is provided. There may be other plausible interpretations that reproduce the same data and lead to different resources.

The estimation of recoverable resources depends on the extension and spatial arrangements of geologic domains Rossi and Parker (1994). It is especially important when small scale structural elements such as veins or faults control the location of high and low grades. A manual interpolation of domain is always subjective. Suppose that three teams propose distinct interpretations using the same four drillholes in Figure 2.3. The red, yellow and green interpretations will report an increasing quantity of metal due to extents of each model.
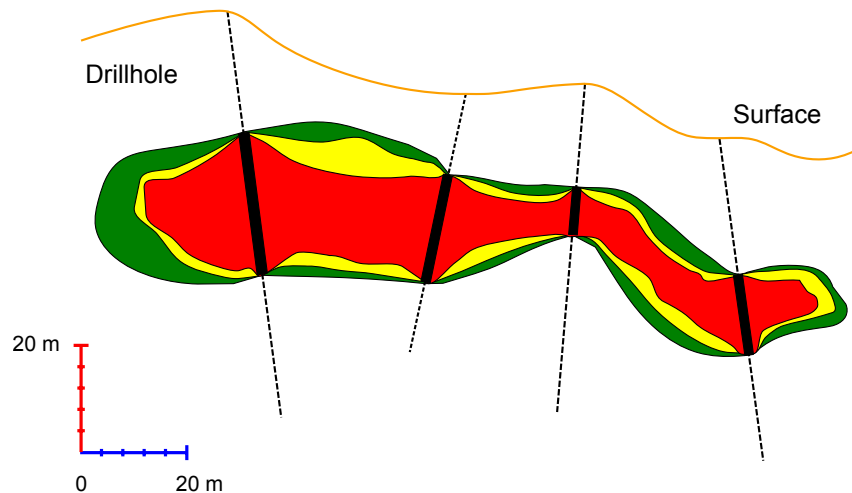
Figure 2.3: Three subjective geologic interpretations of a lens-shaped orebody. The same amount of drillhole information is considered. The red, yellow and green orebodies can be seen as the result of conservative to optimistic interpretation that depends on the assumption taken by the modeler.

At early stages of evaluation, there is relatively wide spacing between drillholes in the deposit. This limited amount of data leads to smooth modeling of the domains with little information on short scale variations for both explicit and implicit

18

methods. In general, these type of models result in too optimistic and biased mineral resource estimates (Srivastava, 2005) due to the misrepresentation of the variability of contacts of domains, that is, the reality is more variable and would incur more dilution and lost ore.

Updating of the geologic model is a continuous process fed by the addition of more information. Richards and Sides (1991) shows an illustrative case study of the evolution of the geologic modeling by the acquisition of additional data. The study showed that the deterministic modeling was correct in representing the large-scale geologic features; but, did not represent the short-scale features, leading to biased resource estimates.

2.3   Spatial Stochastic Geologic Modeling

The large scale macro features of domains could be represented by the techniques mentioned above. Yet, there is micro shorter scale element that appears random with widely spaced data. The macro-micro duality (Christakos, 2012) reflects the fact that in earth sciences phenomena there exist variability at all scales. Concepts in probability theory, in particular in the area of stochastic processes, provide models for this macro-micro duality. Doob (1953) defined a stochastic process as any process in nature whose evolution may be characterized in terms of probability theory.

Modern probability theory rests on the concept of random variables. Geostatistics was largely influenced by the developments around this idea and applied to spatial prediction of natural variables. The presence of the macro-micro duality in geologic deposits strongly suggests a treatment based on structured spatial stochastic process. An illustrative example is showed in Figure 2.4 that recalls the vein-binary image in Figure 1.5. A small portion of the contact is shown in detail. On the right, three stochastic realizations provide plausible interpretations of the real contact shown in the middle. The spatial prediction of the contact based on the fundamentals of random variables provides an accurate description of the uncertainty in the contact.
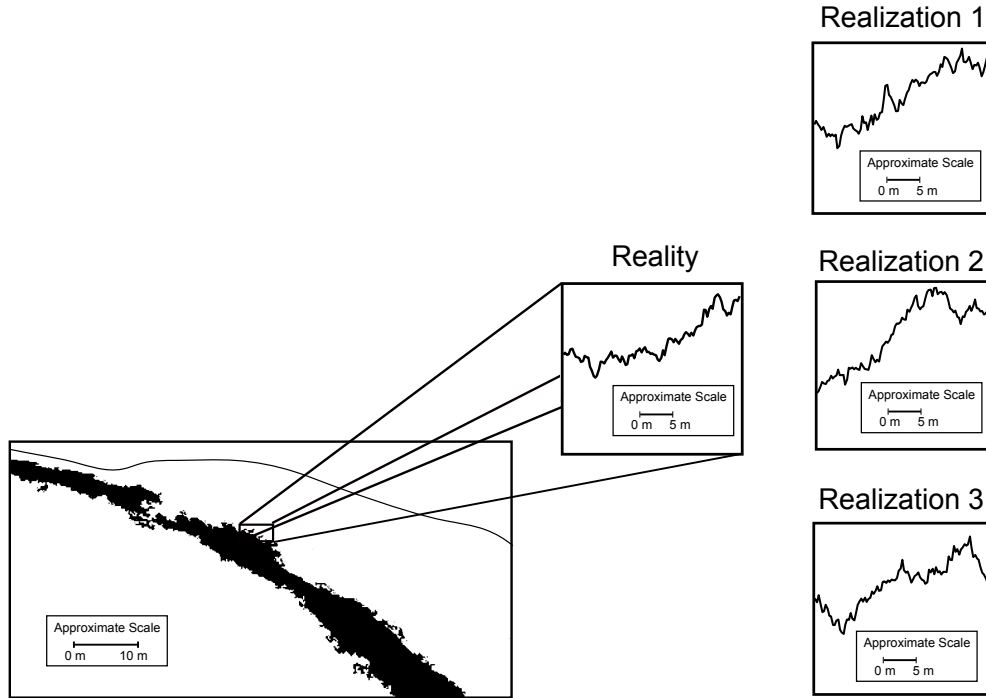
Figure 2.4: Macro-micro duality phenomenon observed on earth sciences field leads to implement algorithms and techniques based on the probability theory. The illustration on the right shows that the evolution of the vein-contact may be characterized by a stochastic process.

2.3.1 Background: Overview of Geostatistics

The roots of geostatistics may be traced back to the pioneering work of D. G. Krige in the late 1950's faced with the problem of estimating grades in the Witwatersrand gold fields in South Africa (Sichel, 1952; Krige, 1951). Matheron (1962, 1965) formalized the field of geostatistics by providing theoretical support for the estimation of unbiased mineral recoverable resources . The development of the estimation technique known as kriging led to an improvement in the estimation at unsampled locations using nearby correlated data. Kriging has the advantage over other estimation techniques since it accounts for information about the correlation of the data, their configuration in space, and a measure of the error of estimation. However, despite its optimality, the estimates show a stronger spatial correlation than the data, and a smaller variance than the global distribution

The introduction of simulation techniques (Matheron, 1973; Journel, 1974; David,

[1977](#)) changed the original paradigm of geostatistics from the optimal estimation at unsampled locations to (1) reproduction of the spatial structure of data and (2) uncertainty characterization. The use of simulation provides equally probable realizations of the grade distributions. Each location has a set of possible values conditioned to surrounding data. The realizations reproduce the histogram as the conditioning data and the variogram or correlation structure. Simulation methods are not restricted to grades, but also provide techniques for geologic domain modeling (Journel, 1983) through the developments of discrete geostatistical analysis.

Geostatistics is defined as the branch of applied statistics that provides methods to characterize the spatial distribution of natural variables and to assess their underlying uncertainty. The concept of a *regionalized variable* was developed by Matheron (1965) and constitutes the foundational element of geostatistics. It corresponds to a function defined in space that reflects macro correlated structures with micro random behavior. Geostatistics models do not intend to understand the complex natural system, but rather describes numerically the spatial behavior in a statistical sense.

The Concept of Random Functions

Matheron (1962) established the theoretical formalism of geostatistics aimed at unifying the observed spatial structure of natural variables with stochastic processes. The concept of a random variable (RV) is introduced. A RV may take a set of possible outcomes. The set of values is characterized by a probability distribution function. In geostatistics, the RV depends on location within space, commonly a domain $D \subset \mathbb{R}^3$. Typically, a random variable is denoted in upper case $Z(\mathbf{u})$ at location $\mathbf{u}$ and its outcome value as lower case $z(\mathbf{u})$, where $\mathbf{u} \in D$ relates the location of the RV in the space. Its distribution function is represented in terms of cumulative probability:

$$F_Z(\mathbf{u}; z) = Prob\{Z(\mathbf{u}) \leq z\}$$

A random function (RF) is defined as a family of correlated random variables within the study area D such as $\{Z(\mathbf{u}), \mathbf{u} \in D\}$. The cumulative probability distribution function considering all locations simultaneously is known as the **spatial law**:

$$F_{Z(\mathbf{u}_1),....,Z(\mathbf{u}_n)}(z_1, ..., z_n) = P\{Z(\mathbf{u}_1) \leq z_1, ..., Z(\mathbf{u}_n) \leq z_n\}$$

$$\forall \mathbf{u}_\alpha \in D, \alpha = 1, ..., n \qquad (2.1)$$

Random functions work under the same probability principles as random variables, except that they require the probability of many locations simultaneously.

Moments of Random Functions

Important features related to the behavior of the probability distribution functions may be characterized by its moments. In geostatistics, the first two moments are often considered enough to characterize the spatial law. The first-order moment correspond to the expected mean:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}), \mathbf{u} \in D \qquad (2.2)$$

The second-order moment is known as the variance:

$$Var\{Z(\mathbf{u})\} = E\{(Z(\mathbf{u}) - m(\mathbf{u}))^2\}, \mathbf{u} \in D \qquad (2.3)$$

in general, the expected mean and variance depend on the location $\mathbf{u}$. The second-order moment can also be defined between two locations, in this case the moment is called the covariance:

$$Cov\{Z(\mathbf{u}_1), Z(\mathbf{u}_2)\} = E\{(Z(\mathbf{u}_1) - m(\mathbf{u}_1))(Z(\mathbf{u}_2) - m(\mathbf{u}_2))\}, \mathbf{u}_1, \mathbf{u}_2 \in D \qquad (2.4)$$

Another second-order moment widely implemented in geostatistical models is

the variogram defined as:

$$2\gamma(\mathbf{h}) = Var\{Z(\mathbf{u}_1) - Z(\mathbf{u}_2)\}, \mathbf{u}_1, \mathbf{u}_2 \in D \tag{2.5}$$

The spatial structure of a wide variety of natural phenomena in earth sciences may be modeled using random functions described by second-order moments. These statistics consider the two-point configuration of data, which is the simplest geometric arrangement possible.

Stationarity

To infer the spatial multivariate distribution of natural variables requires repetitions. The problem consists of determining the probability distribution of the random variables at each unsampled location conditioned and correlated to the surrounding data. The uniqueness of natural phenomena precludes the inference of the distribution function due to a single observation at each location. A weaker hypotheses must be considered. The concept of stationarity allows the data to be considered as replicates of the same random variable over the domain. The inference of the first two-order moments then depends on all of the data taken together. The order of moments are considered invariant under translation. Further discussion around this hypothesis is developed when dealing with the concept of stationarity of training images for multiple point statistics techniques in Subsection 2.5.1.

Ergodicity

The concept of ergodicity allows the inference of the random functions when dealing with a unique phenomena. By definition, a random function is ergodic in the mean if its average converges to the expected mean as long as the extension of the domain tends to infinite:

$$\lim_{D\to\infty} \frac{1}{|D|} \int_D Z(\mathbf{u}) \, \mathrm{d}u = E\{Z(\mathbf{u})\}, \mathbf{u} \in D \tag{2.6}$$

In practice, deviations from the expected mean are observed due to the finite

23

character of a practical domain. The concepts of stationarity and ergodicity are fundamental assumptions that make possible the inference and prediction of spatially distributed variables in space.

Different assumptions may be made for the random function. The models are classified based on the nature of their multivariate distribution. The spatial distribution may follow a parametric form as in the case of Disjunctive Kriging (Matheron, 1976) and multivariate Gaussian kriging (Verly, 1983). In early 1980s, a nonparametric approach indicator formalism including Sequential Indicator Simulation (SIS) was proposed by (Alabert, 1987; Journel, 1989). In this method, the conditional probability functions are inferred directly from the data without any strong assumption about the entire multivariate spatial distribution. Further background on indicator-based method is presented here given its importance for the development of the thesis.

2.3.2 Indicator-Based Geostatistical Approach

An intuitive approach to modeling the probability of occurrence of a value k at an unsampled location consists of estimating the proportions of k according to the information provided by surrounding data. This idea leads to consider indicator random variables, that is, to codify the data as probabilities values for a set of categories $k = 1, ..., K$:

$$i(\mathbf{u}; k) = Prob(z(\mathbf{u}) \in k) \tag{2.7}$$

The indicator-based approach, proposed by Journel (1983), allows the direct estimation of the conditional probability distribution at an unsampled location $\mathbf{u}$ by determining the indicator estimate conditioned to n indicator values. The methodology is based on the use of kriging to estimate the conditional probabilities for each category. The kriging involves the construction of a linear estimator based on conditioning data and the indicator mean (Eq 2.8). Two assumptions may be made about the knowledge of the mean. In this first case, the mean is assumed stationary

at all location $\mathbf{u} \in D$, that is, $E[I(\mathbf{u}; k)] = F(\mathbf{u}; k)$. This form of kriging is referred to as simple kriging.

$$I(\mathbf{u}; k)^* = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(\mathbf{u}; k) i(\mathbf{u}_{\alpha}; k) + F(\mathbf{u}; k) \left(1 - \sum_{\alpha=1}^{n} \lambda_{\alpha}^{SK}(\mathbf{u}; k) \right) \qquad (2.8)$$

The weights $\lambda_{\alpha}^{SK}(\mathbf{u}; k)$ assigned to each indicator datum $i(\mathbf{u}_{\alpha}; k)$ and the mean $F(\mathbf{u}; k)$ are found by solving the system of kriging equations in Eq 2.9 for each category $k = 1, ..., K$. The weights obtained are optimal in the sense that they minimize the expected square error of the estimator $I(\mathbf{u}; k)^*$.

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{SK}(\mathbf{u}; k) C_I(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}; k) = C_I(\mathbf{u} - \mathbf{u}_{\alpha}; k), \alpha = 1, ..., n, \qquad (2.9)$$

The indicator estimate accounts for the relations between the unsampled location and the conditioning data locations, and the relation among the surrounding the data itself through the characterization of the indicator covariance function $C_I(\mathbf{h})$.

A second approach considers the mean as unknown and subjected to local surrounding data. This second version of kriging is referred to as ordinary kriging. As the mean is unknown, the estimate of the indicator variable of Eq 2.8 becomes:

$$I^*(\mathbf{u}; k) = \sum_{\alpha=1}^{n} \lambda_{\alpha}^{OK}(\mathbf{u}; k) i(\mathbf{u}_{\alpha}; k) \qquad (2.10)$$

and the constraint $\sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(\mathbf{u}; k) = 1$ is added to the minimization problem. As consequence, the system has now $n + 1$ unknowns, the n weights $\lambda_{\beta}^{OK}(\mathbf{u}; k)$ for each datum and the Lagrange parameter for the added constraint. Thus, the system of kriging equations for each category or threshold $k = 1, ..., K$ leaves:

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(\mathbf{u}; k) C_I(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}; k) + \mu_{OK}(\mathbf{u}; k) = C_I(\mathbf{u} - \mathbf{u}_{\alpha}; k), \alpha = 1, ..., n$$

$$\sum_{\beta=1}^{n} \lambda_{\beta}^{OK}(\mathbf{u}; k) = 1 \qquad (2.11)$$

The probability for each category is estimated individually. This may lead to estimated probabilities that do not sum to 1, and, in some cases, they might $\notin [0, 1]$. Post-processed corrections that involve averaging and standardization help to resolve these issues (Deutsch and Journel, 1998).

Sequential Indicator Simulation (SIS)

The essence of the sequential indicator simulation approach is to use the conditional probability distribution estimated through indicator kriging to draw a categorical value by using Monte Carlo simulation. The conditional probabilities are estimated using both the conditioning data and values simulated at previous locations. As a consequence, the indicator covariance model is reproduced in the realizations. The workflow proceeds as follows:

1. Generate a random path through the grid of nodes. At each node:

2. Search for conditioning data and previously simulated nodes

3. For each category $k = 1, ..., K$, estimate the conditional probability by simple or ordinary kriging as appropriate

4. Correct the estimated probabilities to satisfy the axioms of probability, that is, reset negative estimates to zero then restandarize the sum to one.

5. Simulate a categorical value from the set of probabilities

6. Go back to step 2 until all nodes of the grid have been visited

Sequential indicator simulation is a robust and straightforward technique for transferring uncertainty into the numerical model. The method, however, considers only two-point structural information which implies that the method is not suitable to modeling complicated morphological geologic features. Furthermore, the geologic models sometimes look patchy, and the domains appear unstructured. Finally, the transitions between simulated domains are difficult to control.

2.3.3 Truncated Gaussian Simulation (TGS)

Matheron et al. (1987) proposed an alternative method for geostatistical modeling of categorical attributes based on the simulation of a standard Gaussian random function and its truncation according to specified thresholds. To simulate the Gaussian function, the method requires the conversion of categorical data to normal scores values. The truncation rule establishes the ordering of the contacts and controls the proportions of the geologic domains.

An extended model was proposed by Galli et al. (1994). The idea is fundamentally the same, but instead of using an individual Gaussian random function, the method proposes the use of multiple functions. This model is referred to as Pluri-Gaussian simulation, and the objective is to characterize more complicated geologic domains interrelations. Normally, a maximum of two Gaussian functions are considered to represent relevant geological features in the models (Rondon, 2009; Carrasco et al., 2007; Fontaine and Beucher, 2006). However, the extension of Gaussian functions leads to the definition of a more sophisticated truncation rule; a large number parameters need to be selected from limited drill hole information. Optimization techniques (Deutsch and Deutsch, 2014) may help to setting up the thresholds and manage the transition probabilities between domains accordingly.

Issues related to the connectivity of geologic domains are often observed in the simulated models. This is due to the property of maximum entropy of the Gaussian distribution function, which establishes a maximum variability of highest and lowest quantiles. Whether the geologic phenomenon is characterized by highly connected domains, object-based methods may be preferred.

2.3.4 Object-based Methods

Object-based modeling has been extensively applied in petroleum reservoir characterization and offers an option to model more complicated morphological geological features than SIS or TGS techniques. The method relies on the randomized allocation of entire geologic objects in the simulated field (Bridge and Leeder, 1979). The advantage of the method is that the final models reproduce the pre-conceived

27

geologic objects such as large-scale curvilinear shapes, connectivity features and spatial heterogeneity structures. Their geometric properties are parameterized through distributions that can result in a difficult problem. Yet, the major disadvantage relates to honouring conditioning data. Perturbation techniques that change the size, shape and location of the object help to correct the conditioning problem (Deutsch and Wang, 1996).

An alternative method, that introduced the concept of event-based process, was proposed by Pyrcz et al. (2005) for modeling subsurface based models. Here, instead of parameterizing the distribution function of the geologic objects, the method simulates the physical process that led to the deposition of the geologic system. A large variety of strataform sediments reservoir may be modeled, improving their characterization since the method accounts for erosional features at multiple scales.

2.4   Multiple Point Statistics Simulation (MPS)

Multiple point statistics simulation was developed to overcome the limitations of conventional variogram-based simulation methods. The novelty of the MPS methods consists of replacing the inference of two-point spatial statistics from the data by multiple point configurations from training images. The conditional probabilities are estimated considering a variety of multiple point configurations allowing the reproduction of geologic features from training images at distinct scales and orientations.

The method was originally proposed by Guardiano and Srivastava (1993). Several methods of MPS and important advances on practical applications have been developed since then. Almost ten years since its advent, (Strebelle and Journel, 2000) established the algorithm known as *snesim*, which popularized the use of MPS by implementing a search tree storage procedure, addressing implementation details and reducing computer requirements. There are other proposed methods based on Gibbs Sampler iteration (Deutsch, 1992; Srivastava, 1992; Lyster and Deutsch, 2008) and neural networks (Strebelle and Journel, 1998; Caers, 2001). Ortiz (2003)

introduced the integration of runs and indicator simulation to reproduce higher order statistics. Dimitrakopoulos et al. (2010) introduced data-driven algorithms based on higher order moments known as cumulants for reproducing complex geologic patterns without appealing to training images. Another approaches implement secondary information for handling the reproduction of trends in simulations (Chugunova and Hu, 2008; Straubhaar et al., 2011). Recently, Mariethoz et al. (2010) proposed a direct sampling technique, extensible to continuous variables, that samples the higher order statistics directly from the training image avoiding computing the conditional probabilities.

The *snesim* method is a flexible and robust MPS simulation technique. The algorithm is important for this thesis. A detailed reviewed of theoretical and practical aspects along with the terminology that will be used later are presented below.

Multiple Point Statistics Terminology

MPS algorithms have been designed for simulating categorical geologic attributes. Consider that the variable S may take K possible states $\{s_k, k = 1, ..., K\}$ for all locations $\mathbf{u} \in D \subset \mathbb{R}^3$. The following terminology will be used hereafter for describing the theory and practice of *snesim* algorithm.

- A data event D_n describes the spatial configuration of n surrounding data with respect to the centered location $\mathbf{u}$. The locations are described by the set of vectors $\{\mathbf{h}_\alpha, \alpha = 1, ..., n\}$, and their attributes $\{s_k(\mathbf{u} + \mathbf{h}_\alpha), \alpha = 1, ..., n\}$

The problem is to estimate the conditional probability of the indicator random variable $S(\mathbf{u})$ given an n-multiple spatial configuration of D_n.

2.4.1 Single Normal Equation Algorithm

The conditional probabilities are estimated based on the frequency of occurrence of the data event D_n in the training image.

The indicator binary random variable A_0 denotes the state s_k of occurrence of

$S(\mathbf{u})$:

$$A_0 = \begin{cases} 1, & \text{if } S(\mathbf{u}) = s_k \\ 0, & \text{otherwise} \end{cases} \tag{2.12}$$

In the same manner, the indicator binary variable D is set out for the joint state of occurrence of conditioning data $S(\mathbf{u}_\alpha) = s_{k_\alpha}, \alpha = 1, ..., n$ in the data event D_n:

$$D = \begin{cases} 1, & \text{if } S(\mathbf{u}_\alpha) = s_{k_\alpha}, \forall \alpha = 1, ..., n \\ 0, & \text{otherwise} \end{cases} \tag{2.13}$$

Alternatively, the indicator variable D may be decomposed into the product of individual indicator variable Eq (2.12) associated to the n data:

$$D = \prod_{\alpha=1}^{n} A_\alpha, \text{ with } A_\alpha = \begin{cases} 1, & \text{if } S(\mathbf{u}_\alpha) = s_{k_\alpha} \\ 0, & \text{otherwise} \end{cases} \tag{2.14}$$

Considering one datum at a time, as classical two-point algorithm, the conditional probability is estimated by the simple indicator kriging equation:

$$Prob(A_0 = 1 | D = 1) = E\{A_0\} + \sum_{\alpha=1}^{n} \lambda_\alpha [1 - E\{A_\alpha\}] \tag{2.15}$$

$A_\alpha = 1$ is the indicator value of datum at location $\mathbf{u}_\alpha$ corresponding to state s_{k_α}, while $E\{A_\alpha\} = Prob(A_\alpha)$ and $E\{A_0\} = Prob(A_0)$ are the marginal probabilities. Guardiano (1993) showed that the contribution of lower order events including the $\{A_\alpha, \alpha = 1, ..., n-1\}$ single-point indicators are redundant with respect to the information provided by the joint indicator variable D. Thus, the conditional probability may be expressed just in terms of the richer indicator of n data D:

$$Prob(A_0 = 1 | D = 1) = E\{A_0\} + \lambda [1 - E\{D\}] \tag{2.16}$$

where $E\{D\} = Prob \prod_{\alpha=1}^{n} A_\alpha$ is the probability of occurrence of D. Then, the weight of the simple kriging is obtained by solving the single extended normal equation:

$$\lambda Var\{D\} = Cov\{A_0, D\} \tag{2.17}$$

Developing the common expressions for the indicator variance and covariance, one has that:

$$Var\{D\} = E\{D\}(1 - E\{D\})$$

$$Cov\{A_0, D\} = E\{DA_0\} - E\{D\}E\{A_0\}$$

Thus, replacing in Eq 2.17 $\lambda = \dfrac{E\{DA_0\} - E\{D\}E\{A_0\}}{E\{D\}(1 - E\{D\})}$, and then, into equation Eq 2.16:

$$Prob(A_0 = 1 | D = 1) = \frac{E\{DA_0\}}{E\{D\}} \tag{2.18}$$

The extended normal equation of the data event D_n establishes that the conditional probability may be estimated by scanning and extracting its frequency of occurrence from the training image.

The steps of the *snesim* implementation are summarized as follows:

1. Scan the training image and store the frequencies of occurrence of each multiple-point configuration in a search-tree

2. Assign each conditioning data to the closest node of the grid

3. Visit each unsampled location following a random path

4. Compute the conditional probability from the search-tree conditioned to the nearby data event D_n

5. Simulate a categorical value from the conditional probability distribution and go back to step 3 until all nodes of the grid have been simulated

The Training Image

The training image conceptualizes the spatial arrangements and interrelations of the geological features that are expected to be found in the deposit. It represents a richer structural tool and provides a direct and more intuitive understanding of the geologic relations than variogram or covariance functions.

The primary concern of MPS simulation methods is to find a representative training image for the phenomenon being modeled. An incorrect training image may lead to severe bias in proportions and spatial features, compromising future decision-making.

2.5 On the Stationary of Training Images

Multiple point statistics simulation methods rely on the assumption of stationarity for allowing the inference of higher order multiple point events from the training images. In early views, it was assumed that only stationary training images are statistically valid to be used for MPS methods. The current subsection discusses the concept of stationary training images for geologic modeling with the focus on mineral deposits.

2.5.1 Stationary Random Function Model

A brief introduction to the concept of stationarity was provided in Section 2.3.1. A detailed description of the assumptions of strong, weak and intrinsic stationarity for random functions are presented in the following subsection. The concepts reviewed here will be related to the stationary training images.

Strong Stationarity

The assumption of strong stationarity considers that the probability distribution function for any spatial configuration of n data of the random function $\{Z(\mathbf{u}), \mathbf{u} \in$

D}; remains the same for all locations:

$$F_{\mathbf{u}}(z(\mathbf{u}_1), ..., z(\mathbf{u}_n)) = F_{\mathbf{u}+\mathbf{h}}(z(\mathbf{u}_1 + \mathbf{h}), ..., z(\mathbf{u}_n + \mathbf{h})),$$

$$\forall \mathbf{h} \qquad (2.19)$$

Here, the distribution function of the stochastic sequence of n random variables is measured as a whole. The assumption of strong stationarity cannot be directly evaluated since it requires access to repetitions of the multiple configuration of n data. A weaker hypothesis that considers a less restrictive configuration of data is needed.

Weak Stationarity

For the case of weak stationarity, the configuration of multiple data is reduced to a two-point relation. Even though this hypothesis restricts the capability of the model for reproducing complicated geologic phenomena, it allows the statistical inference of covariance/variogram functions. Weaker stationarity of the random function $\{Z(\mathbf{u}), \mathbf{u} \in D\}$ is defined by:

$$E[Z(\mathbf{u})] = E[Z(\mathbf{u} + \mathbf{h})] = m, m \in \mathbb{R}$$

$$E[Z(\mathbf{u})^2] = E[Z(\mathbf{u} + \mathbf{h})^2] = \sigma^2, \sigma^2 \in \mathbb{R} \qquad (2.20)$$

$$E[Z(\mathbf{u})Z(\mathbf{u} + \mathbf{h})] = C(\mathbf{h}), \forall \mathbf{h}$$

The mean and variance are constant for all locations $\mathbf{u}$, and the covariance function only depends on the separation vector $\mathbf{h}$. In some cases, data shows a trend behavior making the hypothesis of weak stationarity unreasonable. Strong structural controls of the mineralizing solutions may condition the spatial distribution revealing trends. An even weaker hypothesis would offer an alternative to modeling this type of phenomenon; this is the intrinsic stationarity assumption.

Intrinsic Stationarity

Nonstationary phenomenon are challenging for geostatistical methods. A useful approach consists of discomposing the data into an averaged trend that reproduces the large scale nonstationary behavior, and a stationary residual which may fluctuate around this trend, Eq (2.21). The intrinsic hypothesis intends to extract the stationary residual directly from the data, avoiding the assumption of a specified structure for the trend. The hypothesis does not work directly with the data, but with its increments.

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}) = \sum_{l=1}^{k} a_l f^l(\mathbf{u}) + R(\mathbf{u}) \tag{2.21}$$

The increments for a location $\mathbf{u}$ are defined as a linear combination of data, that is, $Z_\lambda(\mathbf{u}) = \sum_{\alpha=1}^{m} \lambda_\alpha Z(\mathbf{u}_\alpha)$; where $Z(\mathbf{u}_\alpha)$ are the m number of observed data involved in the increment of order k. The intrinsic hypothesis establishes that the increments are second order stationarity, namely, its mean and variance are finite over the domain D:

$$E[Z_\lambda(\mathbf{u}) = \sum_{\alpha=1}^{m} \lambda_\alpha Z(\mathbf{u}_\alpha)] = 0$$

$$Var[Z_\lambda(\mathbf{u}) = (\sum_{\alpha=1}^{m} \lambda_\alpha Z(\mathbf{u}_\alpha))^2] = 2\gamma_k(\mathbf{h}) \tag{2.22}$$

In order to remove the trend $m(\mathbf{u})$ from the data in Eq 2.21, the coefficients of the increments must be configured such that $\sum_{\alpha=1}^{m} \lambda_\alpha [\sum_{l=1}^{k} a_l f^l(\mathbf{u})] = 0$, filtering the trend from the data. Thus, the variogram analysis may be performed directly from the stationary residuals.

The polynomial of degree k reflects the complexity of the trend. The simplest case remains when the trend is constant, that is, a polynomial of order zero. This

is the classic definition of the intrinsic model:

$$E[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\alpha + \mathbf{h})] = 0$$

$$Var[(Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\alpha + \mathbf{h}))^2] = 2\gamma(\mathbf{h}) \qquad (2.23)$$

The hypothesis of the intrinsic model suffices for a wide range of natural variables for mineral deposits.

2.5.2 Stationarity Multiple Point Statistics Simulation

In variogram-based geostatistical methods, the stationarity is an important assumption that makes possible to perform statistical inference. Once the histogram and variogram/covariance functions have been inferred, multiple realizations are drawn from the random function. MPS simulation methods rely on a different paradigm. Instead of characterizing the random function through covariance/variogram functions before simulating, the inference of multiple point statistics is performed directly from training images. Under this view, the stationarity plays a different role which consists of allowing the estimation of local conditional probabilities given multiple configurations of data, and not the inference of parameters of the random function.

Similarly to the hypotheses of strong and weak stationarity described for random functions, one could define related concepts applied to MPS simulation methods. The type of stationarity assumed determines the manner that the local conditional probabilities are estimated.

Strong Stationary for MPS Methods

Under this assumption, one location is enough for estimating the conditional probabilities associated to multiple point configurations. In geosciences, multiple images of the same phenomenon are difficult to find. An artificial set of n ellipsoid-based images are used to illustrate the idea of strong stationarity, see Figure 2.5. The ellipses are placed randomly in the grid using the object-based program *ellipsim* of GSLIB.
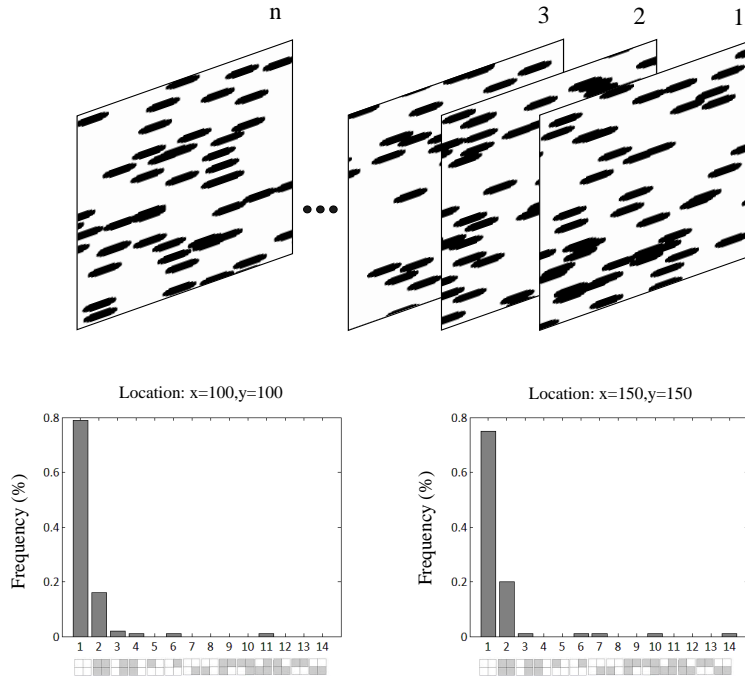
Figure 2.5: Illustration of the extended definition of strong stationarity for training images. The set of ellipsoid-based image represents realizations of the same random process. Independent of the location, the multiple density function remains equal; no assumption of stationarity is needed for the conditional probability inference.

The distribution of the multivariate random function may be characterized through the computation of multiple point density functions (Boisvert et al., 2010). Higher order statistics depend on the configuration of points and the number of categories contained in the training image, the relation is given by k^n where k is the number of categories and n the locations of the multiple point event. The multiple point density functions determine the frequency of appearance of each configuration, and each data event has associated a multiple point density function. Two locations in the grid are considered ($x = 100, y = 100$) and ($x = 150, y = 150$), along with a template of 2 by 2. The multiple point density functions are presented on the right part of the figure. Independent of the location considered, the multiple point density function remains equal when compute it through the set of images. The strong stationarity leads to practical assumptions to facilitate the statistical inference, and the conditional probabilities may be estimated directly from the set of realizations and for any location.

36
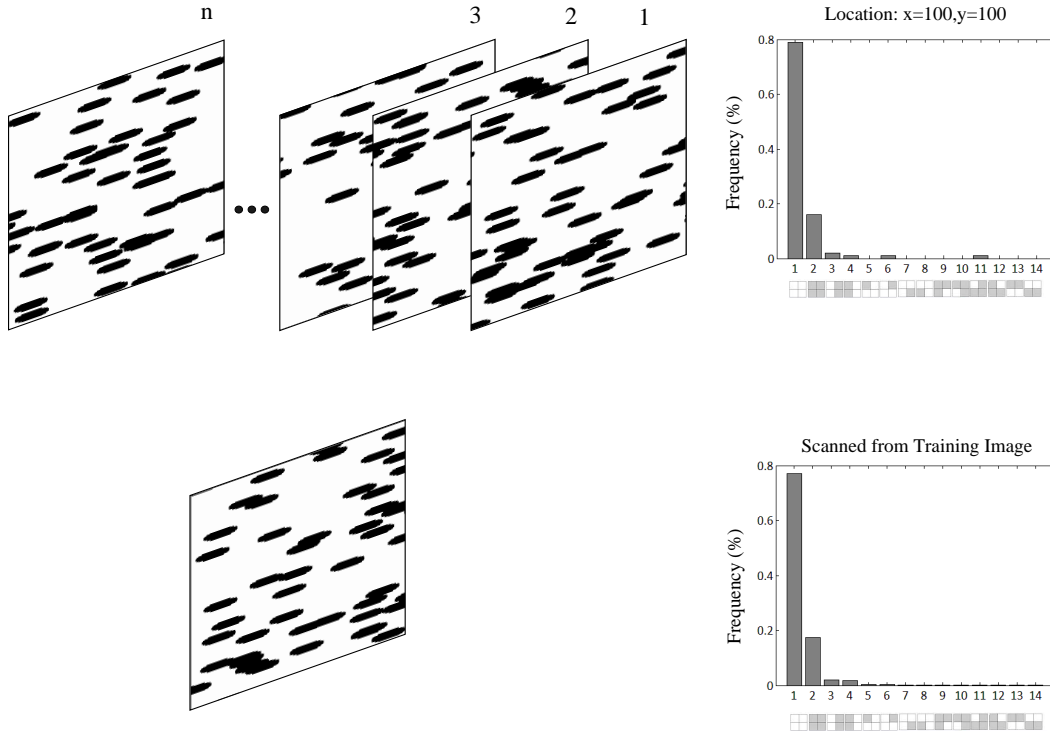
Figure 2.6: Comparison of the multiple density functions when assuming strong training image stationarity, and an individual training image assumed with the hypothesis of weak stationarity.

Weak Stationary for MPS Methods

The weak stationarity of training images requires an assumption that alleviates the lack of a set of realizations. Here, the estimation of local conditional probabilities of multiple data configurations calls for the probabilities to be invariant by translation over the image. Under this definition, the set of training images is replaced for an individual image large enough to estimate the conditional probabilities. Figure 2.6 compares the multiple point density functions at using an individual and the set of training images based on the previous two by two template. Both multiple point density functions remain the same. The assumption of stationarity over an individual training image is considered enough for most MPS methods in order to derive the local conditional probabilities.

37

Transference of Geologic Features from Training Images

An adequate estimation of local conditional probabilities ensures the correct transference and reproduction of geologic features from the training image to the simulation grid. The richness of the information carried by the local conditional probabilities is fundamental to obtain a reliable reproduction of the geologic features, or at least, and adequate transference of the geologic essence of the training image.

However, the estimation of local conditional probabilities might lead to multiple point geologic configurations not contained in the training image. This is due to the MPS methods estimating the local conditional probabilities in a sequential stepwise fashion in order to derive the simulation globally. In order to allow the estimation of conditional probabilities for specific multiple point configurations not found in the training image, remedial actions need to be undertaken. Solutions aim to modify the configuration of the conditioning data event. In this respect, (Strebelle and Journel, 2000) proposes a heuristic alternative that consists of dropping off the furthest data up to encounter the exact configuration that allow estimating the probability distribution function.

For MPS simulation, the objective is the reproduction of the essence features of the training image. These methods rely more on finding the exact multiple point configuration in the training image than count with a larger number of occurrences of it. This is a substantial difference with regard to variogram-based geostatistical approaches where a large number of repetitions ensures the correct inference of the parameters of the model.

Multiple-Scale Local Conditional Probability Estimation

To transfer the essence of the training image, it is necessary to consider the large scale features first and then consider higher resolution small scale features. The MPS simulation needs templates of a reasonable size for convergence. The template must account for the estimation of conditional probabilities at multiple-scales. Templates of dynamic size are considered by implementing the use of nested multigrids. The configuration of the template remains the same, but, the scale changes according

to the multigrid size. The MPS simulation depends on the multigrid and template selected. The correct parameterization of these components is unclear. The size of training images and templates is constrained by memory and CPU speed.
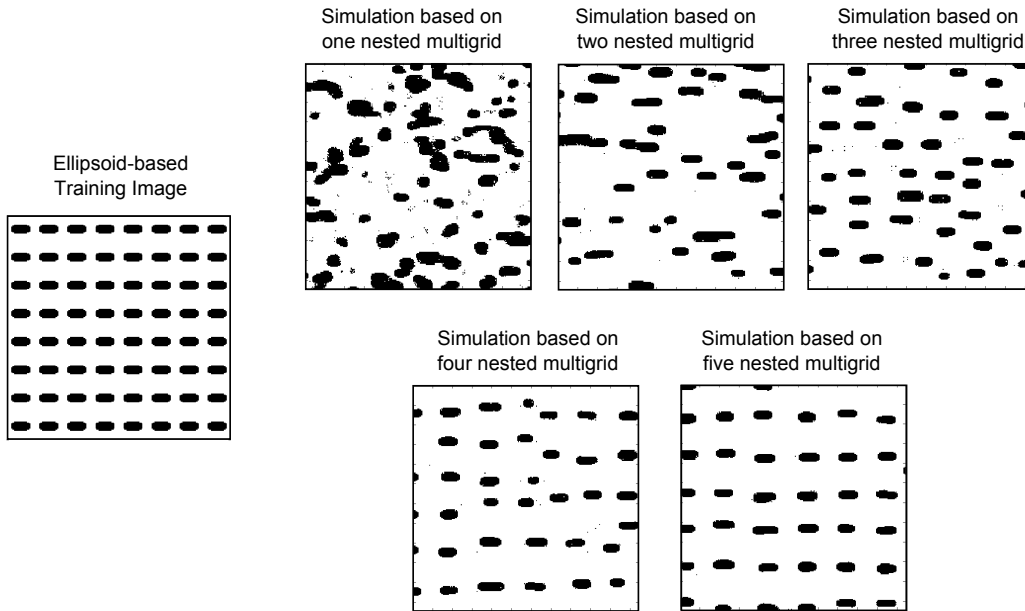
Figure 2.7: Effect of the multigrid parameter upon the reproduction of the ellipsoid-based training images on the left. An implementation of decreasing nested-multigrids ensures the transference of the essence of the training image.

The effect of implementing a different number of nested multigrid is presented in the example of Figure 2.7. For multigrid 1, that is, no rescaling, the simulated features do not look like the ellipsoid-based training image. After increasing the number of multigrids, the correct features start appearing in the realizations. In the last example, by implementing five multigrids the essence of regular-spaced ellipsoids of the training image is transferred to the simulation; the sequential scale reduction of the features being modeled ensures a correct reproduction.

The Stationary Character of Training Images

In a geostatistical context, the concept of stationarity refers to the assumption that statistical variation is invariant over the domain being modeled, that is, the geologic phenomenon remains spatially homogeneous. This definition also applies for training images. An image is said to be stationary if the geologic features are similar when

39

looking at different locations; there is no strong relation between the features and their location in the image. In early days, it was established that only training images considered as stationary are statistically valid for a suitable implementation of MPS methods (de Vries et al., 2009). This view is changing.

When dealing with a stationary training image, the idea of spatial homogeneity is directly related to the scale of observation and the size of the training image. Figure 2.8 illustrates this perception of stationary training images. Based on an image composed of sinusoidal channels with strong stationary essence, three scales of observation are considered. The red squares show three zoomed images from the original sinusoidal image. If each image is deemed individually, our perception of the idea of stationary increases along with the scale of observation due to replicates of channels are found with more frequency. Even though the three images are part of the same channel-based stationary image, not all of them are assumed as stationary. In other words, stationarity is not a property attached to the data, but a hypothesis of the stochastic model adopted; it is the process which is stationary, not the image. The image, in turn, could manifest certain stationary aspects but this cannot be proved.
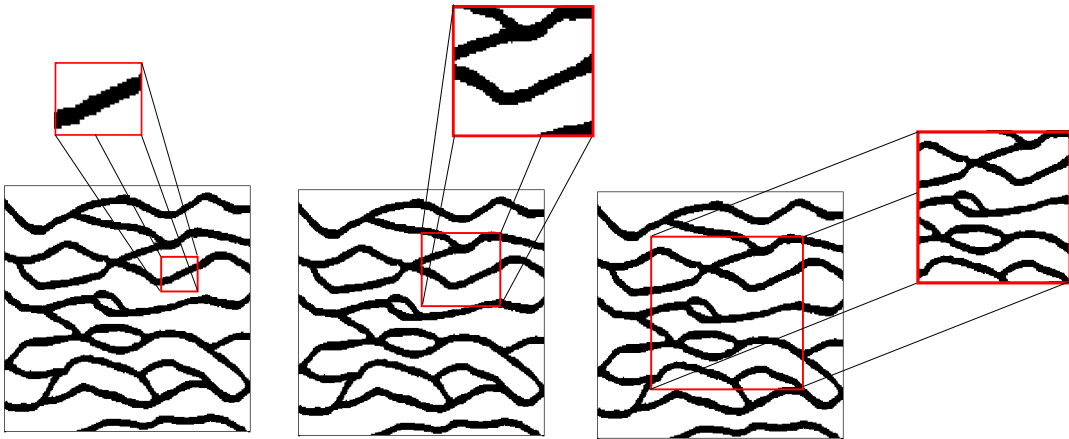
Figure 2.8: Representation of how the scale of observation affects the perception of a training image considered stationary.

A stationary process might generate non-stationary images, or inversely, a non-stationary process could generate stationary images depending on the scale that

the image is observed. Criteria for the invariant location of geologic features based on the scale, orientation and regions of replicates of patterns allows checking if a training image is considered suitable to the application of MPS methods Mirowski et al. (2009). Stationary training images ensure reproducing their patterns as close as possible in the simulations, but not restrict the use of MPS methods for non-stationary images. The MPS algorithm, independent of the class of training images, acts as a conveyor of geological features to the simulations grid. What is not ensured is the suitable transportation, matching and reproduction of the essential features of the image.

As pointed out, the previous view was that MPS techniques were designed to modeling deposits with a strong stationary component. Any training image may be seen as a repository of geologic multiple point configurations and used for MPS simulation. Thus, the definition and field of application of MPS method may be extended since they are no longer restricted to the stationary character of training images.

2.5.3 Data-Driven Training Image: An Accurate Approach for MPS Mineral Deposit Modeling

In the past, MPS methods have not been used in the field of mineral resources modeling. Perhaps, reliance on the need for stationary training image discouraged the use of these techniques. In particular, MPS for mineral deposits must account for the considerable amount of data involved. The large amount of data helps to characterize the interrelations of geologic domains. Thus, the focus is on relatively small scale geologic features.

An Example

An application based on a porphyry deposit is used for illustrating the potential of MPS methods for mineral deposits. The example considers an application performed in 2-D. Figure 2.9 (top left) shows a schematic cross section of intrusive units for a porphyry deposit used as training image. The host rocks are in green and cyan;

two intrusive units in orange and red. The exercise consists of sampling the training image and then using the data for simulating the porphyry. A scatter plot with the sampled conditioning data is shown at the bottom left part. Four resulting MPS simulations are presented on the right. Note that the essence of the training image is transferred to the simulations. Both the intrusive and host rock units maintain their structural geologic interrelations at large scale. The MPS simulation process may be divided into two steps: first, the data delineates the large scale features, in particular, the shape and size; second, the modeling of short scale features, specifically the contacts between units, by matching the geologic features to the those contained in the training image
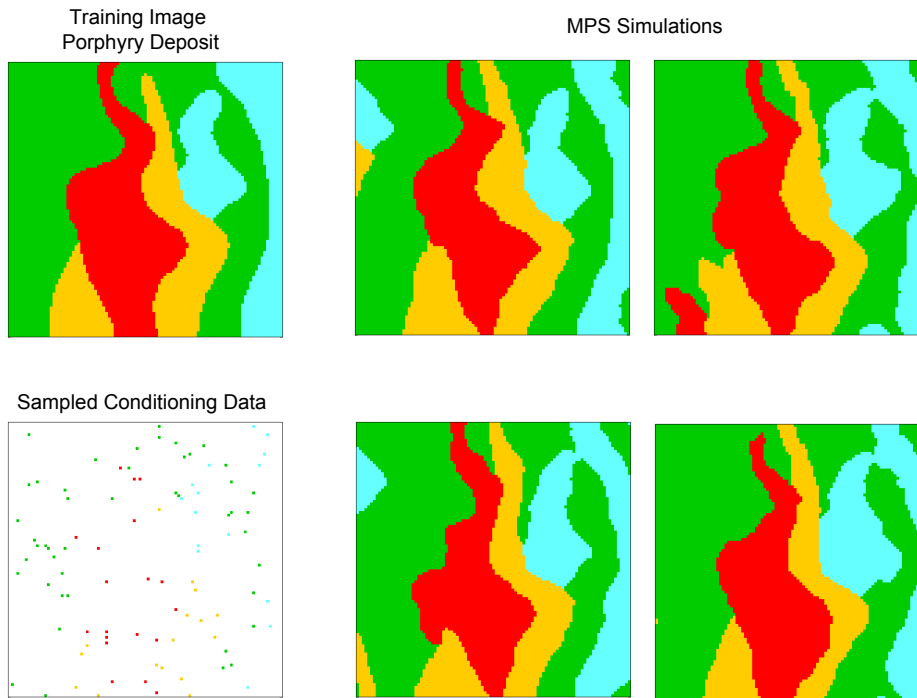
Figure 2.9: MPS simulations using as training image a schematic cross-section of a porphyry deposit. Data is sampled from the training image and used as conditioning data. On the right, four realizations.

The use of data-driven training images is of special interest for mineral deposits modeling. Account for a training image derived from data avoids future discrepancies in the simulations and the presence of artifacts is reduced. Data-driven training images along with the use of large amount of data guarantees the generation of

models that inherit the conceptual design imposed in the training image.

Chapter 3

Multiple Signed Distance Function Implicit Geologic Modeling

Implicit geologic modeling techniques aim for objective, repeatable and flexible models. One limitation of most current implicit methods is that they only work with two domains at a time. The current chapter presents an algorithm to generalize the use of an implicit function in the case of multiple domains. Examples illustrate the advantages of the algorithm and practical implementation considerations describe the robustness and limits of the technique.

3.1 Implicit Functions

The use of implicit modeling was first introduced in the field computer graphics for the creation of objects of distinct geometry and complexities (Bloomenthal and Bajaj, 1997). The core idea relies on the use of an implicit function to demarcate regions in the space of diverse shape and extension.

The implicit functions are an alternative to the classical formalism $f(x) = x$. In the implicit definition, the multiple n arguments and its image are placed on the same side of the equation in the form $\phi(x_1, ..., x_n) = 0$. This allows defining

a contour at 0 in $n + 1$ dimensions in a lower dimensional space n. The contour determines the regions where the sign of the value of the implicit function changes, that is, negative for *inside* and positive *outside*. The classical equation of the circle $\phi(x, y) = x^2 + y^2 - 1$ serves as an illustrative example, see Figure 3.1. On the left, the function $f(x, y) = x^2 + y^2$ is plotted for the region [-2,2]x[-2,2]; on the right the projection of function $\phi(x, y) = x^2 + y^2 - 1$ into the plane xy. This projection is equivalent to the circle of radius 1 intersected by the black plane at $f(x, y) = 1$ as shown on the right of the figure. The circle of radius 1 establishes the limits where the function $\phi(x, y)$ changes sign . Different size circles may be constructed by changing the level where the function $f(x, y)$ is defined.
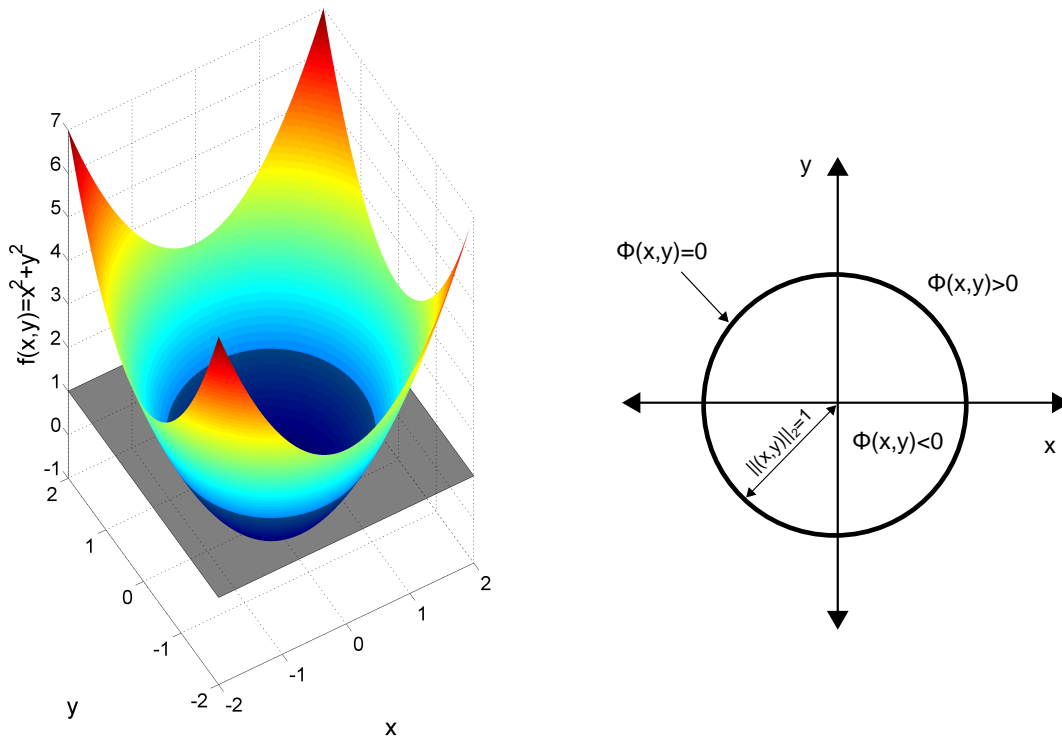
Figure 3.1: Definition of regions in space by an implicit function. On the left, the surface-image of the function $f(x, y) = x^2 + y^2$. The black plane is located at $f(x) = 1$; on the right, the intersection of the plane and the surface that represents a circle of radium 1. The sign of the implicit function $\phi(x, y)$ establishes the boundary of the circle.

This example is a simple case based on a parametric function and, in general, would not be suitable for geological features. The following section describes how

implicit functions are adapted, constructed and applied to operate in a geological perspective.

3.2   The Signed Distance Function Algorithm

The signed distance function methodology (Osher and Fedkiw, 2003) is an implicit modeling technique where the function is constructed by interpolating a distance measure based on the conditioning data. For each sample, an anisotropic distance between itself and the nearest sample belonging to an opposite domain is computed and assigned. Negative values represent the distance to the boundary for samples within the domain, and positive values are used for the distance to the boundary for samples outside the domain. The interface that separates the regions in space is determined by the sign of the estimated signed values. Specifically, the algorithm proceeds as follows.

First, the set of data $\{z(\mathbf{u}_\alpha), \alpha = 1, ..., n\}$ is coded as a binary indicator in order to specify the samples that belong to inside or outside domains, see Eq 3.1. Note that the selection of which domain is considered inside is arbitrary and does not affect the proceeding.

$$i(\mathbf{u}_\alpha) = \begin{cases} 1, & \text{if } z(\mathbf{u}_\alpha) \text{ is inside} \\ 0, & \text{if } z(\mathbf{u}_\alpha) \text{ is outside} \end{cases} \tag{3.1}$$

Then, signed distance values are calculated for each sample according to Eq (3.2). If the sample is inside the domain, the distance is set as negative; otherwise, the distance is positive. The location $\mathbf{u}_\beta$ correspond to the closest sample of the opposite domain to $\mathbf{u}_\alpha$. The Euclidean norm is used to measure the distance.

$$d(\mathbf{u}_\alpha) = \begin{cases} -||\mathbf{u}_\alpha - \mathbf{u}_\beta||, & \text{if } i(\mathbf{u}_\alpha) = 1 \\ +||\mathbf{u}_\alpha - \mathbf{u}_\beta||, & \text{if } i(\mathbf{u}_\alpha) = 0 \end{cases} \tag{3.2}$$

The signed implicit function is then interpolated to all locations of interest. Krig-

ing is preferred for this task due to its ability to account for directions of continuity and the spatial configuration of data. Other techniques such as inverse distance could also be utilized.

$$d^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha^{OK}(\mathbf{u}) d(\mathbf{u}_\alpha) \qquad (3.3)$$

The interpolation technique of ordinary kriging is robust and widely used. The kriging estimator is presented in Eq 3.3. In particular, implementation of global ordinary kriging is recommended. This method does not search for data, using all the available data to estimate distance values at all locations, avoiding the presence of artifacts in the resulting models. The main disadvantage relates to the capacity of computer processors for handling large amount of data which could lead to computational intractability.

$$i^*(\mathbf{u}) = \begin{cases} 1, & \text{if } d^*(\mathbf{u}) \leq 0 \\ 0, & \text{otherwise} \end{cases} \qquad (3.4)$$

At the end, the domains are classified at the unsampled location $\mathbf{u}$ in function of the sign of the signed distance estimates, Eq 3.4. If the estimate is negative, the location is classified as inside; otherwise, the domain is considered as outside, and the boundary remains established.

2-D Signed Distance Modeling Example

Figure 3.2 shows an artificial example of the signed distance method. On the left side, the two domains considered (white and black); each sample is flagged with the distance to its closest opposite sample according to Eq (3.2). Global ordinary kriging is performed at all locations, and colored estimates may be seen in B. The classified domains are presented in C. The sign of the kriged estimate defines the domains, positive for white and negative for black. Note how the smoothness of the estimates ensures the smooth transitions and a well-delineated boundary between the classified domains.

47
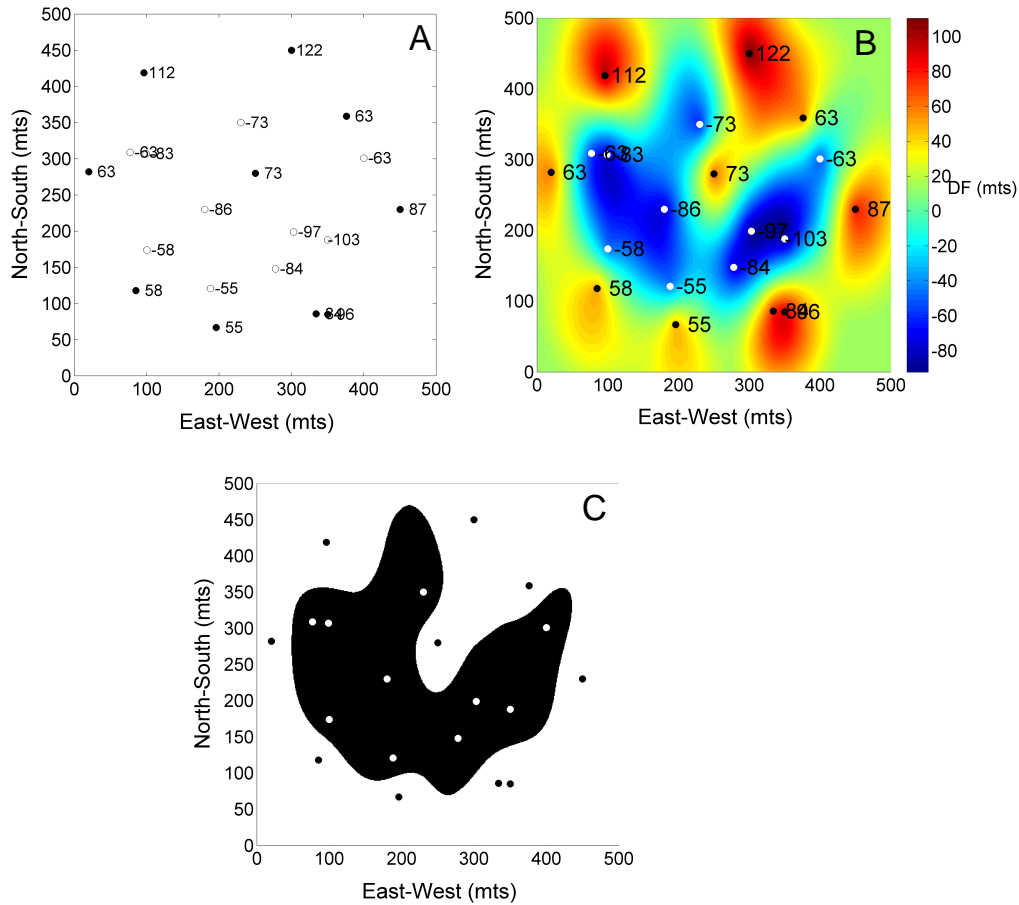
Figure 3.2: Three fundamental steps that summarize the signed implicit geologic modeling. On the left, the samples on the plane *xy* flagged with the signed distance; on the right, the distances estimated by *ordinary kriging*, and at the bottom, the implicit model. In black, the estimates lower than zero, in white the opposite.

3.2.1   Discussion

The signed distance function modeling is a reliable technique that aids the geologic interpretation process. The use of an implicit function ensures a reasonable characterization of geological features into numerical models. Some advantages and disadvantages of the method:

Advantages:

- Geologic models are updated easily when additional data becomes available. The method saves time since manually updating is avoided by the interpolation procedure.

- Using the same parameters, the models are reproduced with exactitude. This makes auditing and checking the models more straightforward.

- A variety of sensitivity analyses may be performed in the models by modifying the parameters involved in the interpolation of the implicit function.

- The process is fast. Geologists may focus their time on understanding the geological processes instead of the tedious construction of cross-sections and plan views.

Among the *Disadvantages:*

- The interpolation of signed distances does not consider that a trained professional can often understand important trends and specific features that may lead to improved results.

- The algorithm is built on kriging so only linear relations between domains are modeled. More complex geometries such as folded structures or curvilinear shapes call for other geostatistical techniques.

- The non-stationarity behavior of the signed distance makes variogram inference arbitrary and questionable.

- There are often edge effects, that is, how far away from the geologic feature are positive data kept? That will change the final model.

- It does not account for uncertainty.

- When dealing with multiple geologic domains, a hierarchical implementation that manages each unit individually provides a satisfactory solution. However, post-processing is needed due to overlapping issues.

Depending on the amount of data involved, the algorithm will provide of interpretations that increase in complexity as data is increasingly collected. When dealing with scarce and widely spaced information, no other arrangements than linear relations may be reproduced. Whether a large amount of data is available,

the algorithm can reproduce more complicated geometries since the drillholes have intersected more details of the deposit.

In what follows, an extension of the distance function algorithm for multiple domains is presented.

3.3 Distance Function Approach for Multiple Geologic Domains

The geologic modeling of mineral deposits involves the interpretation of multiple domains. However, the described signed implicit geologic modeling may only be applied for binary cases. The algorithm presented in this section proposes an extension to multiple domains.

3.3.1 The Algorithm

Suppose there exists K multiple domains in the deposit. For all samples locations $\{z(\mathbf{u}_\alpha), \alpha = 1, ..., n\}$, an indicator vector of K elements is coded according to:

$$i_k(\mathbf{u}_\alpha) = \begin{cases} 1, & \text{if } z(\mathbf{u}_\alpha) = k \\ 0, & \text{otherwise} \end{cases} \qquad k = 1, ..., K \qquad (3.5)$$

Thus, the $k - element$ of the vector is one while the remainder $K - 1$ elements are set to code 0.

In the same manner as Eq 3.2, the signed distance value to the closest opposite domain is computed individually for each k element of the vector. By convenience, the signs of the distance remain equal as the binary case: negative for inside and positive for outside:

$$d_k(\mathbf{u}_\alpha) = \begin{cases} -||\mathbf{u}_\alpha - \mathbf{u}_\beta||, & \text{if } i_k(\mathbf{u}_\alpha) = 1 \\ +||\mathbf{u}_\alpha - \mathbf{u}_\beta||, & \text{if } i_k(\mathbf{u}_\alpha) = 0 \end{cases} \qquad k = 1, ..., K \qquad (3.6)$$

Depending on the information from exploratory data analysis, different anisotropies

50

could be incorporated for each indicator. No correlations among the distance values are considered.

The interpolation is performed individually for each element k. Ordinary kriging is applied multiple times:

$$d_k^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha^{OK}(\mathbf{u}) d_k(\mathbf{u}_\alpha) \quad k = 1, ..., K \tag{3.7}$$

When the multiple variograms appear similar, a single variogram may be considered. This would reduce substantially the processing time since the kriging equations are solved at once for all indicators. Then the final rock type model is determined:

$$i^*(\mathbf{u}) = k' \quad \text{such that} \quad d_{k'}^*(\mathbf{u}) = \min\{d_k^*(\mathbf{u})\}_{k=1}^{K} \tag{3.8}$$

The estimated distance provides a measure of proximity to the closest opposite domain. In this sense, the minimum estimated signed distance value may be seen as the most probable domain to be found at an unsampled location. The classifier presented in Eq 3.8 summarizes this idea. The indicator associated with the smallest estimate determines the domain present at that location.

3.3.2 A Heuristic Measure of Uncertainty for Signed Distance Estimates

The proposed geologic distance-based algorithm does not characterize uncertainty. Practical notions of uncertainty may be coupled with the implicit modeling formalism. This heuristic representation of uncertainty is not the consequence of multiple realizations drawn from a random function; it is derived from a post-processing transformation method.

There exist applications where uncertainty have been incorporated in the implicit framework. McLennan and Deutsch (2006) proposed a methodology for quantifying the uncertainty in the volume function of geologic objects. In this approach, the uncertainty in the mean of the volume function is characterized by the bootstrap sampling technique, and it is established prior to interpolation of the distance

function. The interpolated function honors the expected values of the mean from the bootstrap distribution; the geometry and volume will depend on the sampled quantile.

Munroe (2012) proposed to calibrate a bandwidth of uncertainty along the boundary of domains with two parameters C and β. C controls the thickness of the bandwidth and β the position and tonnage distribution within the bandwidth.

A similar idea was implemented by Wilde and Deutsch (2011) where uncertainty is considered directly in the kriging formalism by adding a constant C to the estimates. The value of the constant represent an uncertainty bandwidth. Then, a set of uniform values are simulated within the bandwidth using a *p-field* simulation approach; the classification of domains involves the estimated distance and simulated uniform values. Cross validation allows calibrating the value of the constant C.

When implementing the proposed multiple signed distance function algorithm, additional information might be retrieved from the estimated distances. As pointed out, the set of estimated distances provides a predicted measure of proximity to the closest boundary. This Euclidean distance measure may be carried out to a more amenable transformed variable within 0 and 1 scale and used as a practical measure of uncertainty.

Transformed Estimated Distance Values

A widely used approach for multiple class classification methods is the *softmax activation function* or also called *multiple logistic regression function* (McCullagh and Nelder, 1989). The idea is to transform the estimated signed distance into outputs that could be interpreted as posterior probabilities. The transformed values lie between 0 and 1 and to sum to 1. The function is commonly used to provide of probability meaning classification methods of multiple outputs as linear discriminant analysis and multinomial logistic regression.

$$P(i(\mathbf{u}) = k) = \frac{e^{\frac{-d_k^*(\mathbf{u})}{\gamma}}}{\sum_{k'=1}^{K} e^{\frac{-d_{k'}^*(\mathbf{u})}{\gamma}}} \quad , \{k = 1, ..., K\} \tag{3.9}$$

52
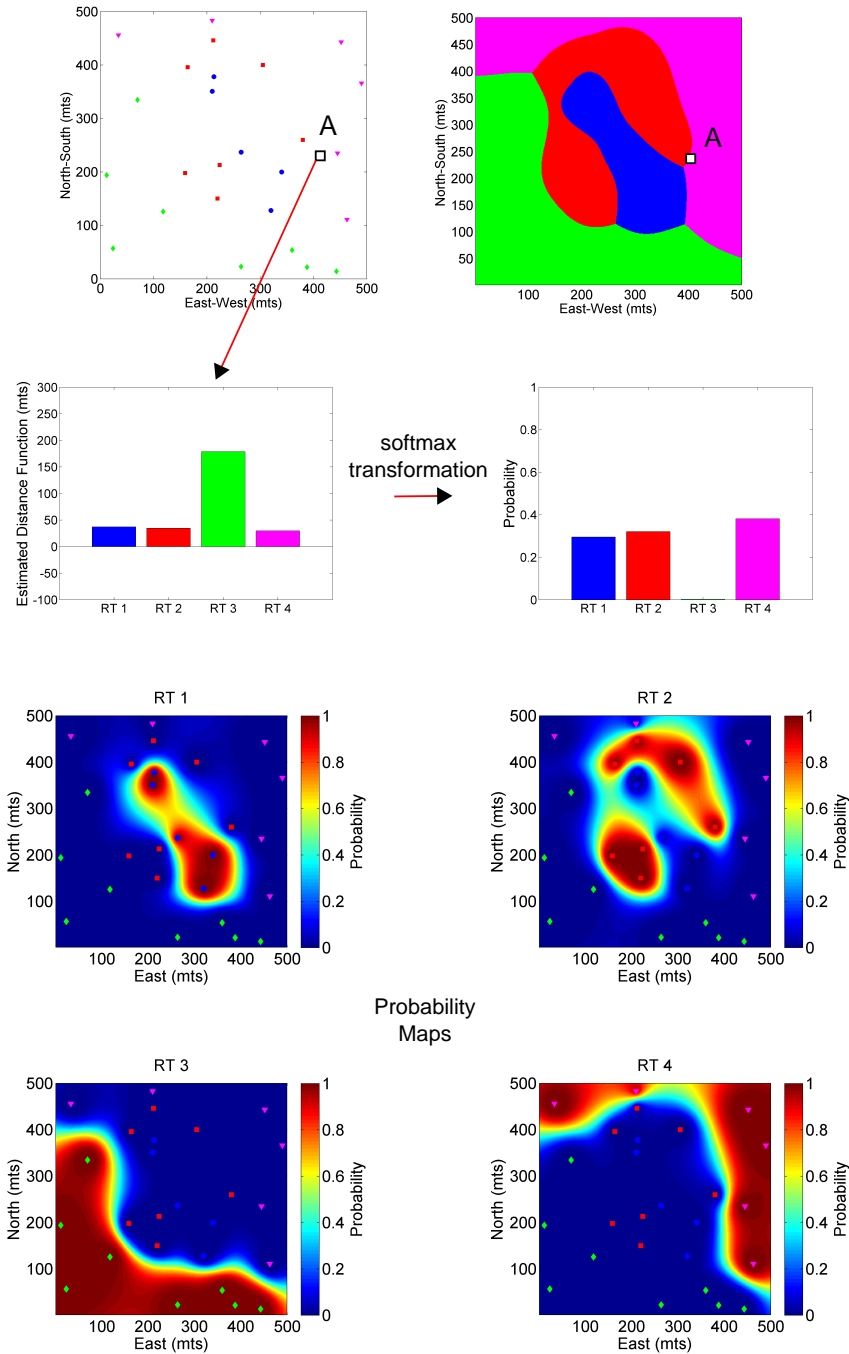
Figure 3.3: Posterior probability estimation based on distance values. At the top on the left, the samples configuration of multiple domains; at the top on the right, the domains classified according to Eq 3.8. In the middle, an example of the softmax transformation function. At the bottom, illustrations of the posterior probability for each domain after applying the softmax function.

The *softmax activation function* is presented in Equation 3.9. $P(i(\mathbf{u}) = k)$ represents the probability at location $\mathbf{u}$ to belong to category k, $-d_k^*(\mathbf{u})$ the estimated signed distance for category k and γ is a parameter that regulates the interrelation among the probabilities of the K categories, a large parameter leads to greater differences in the probabilities.

To illustrate the application of the softmax function, consider the two-dimensional example in Figure 3.3. At the top on the left, samples of four domains colored in blue, red, green and magenta; on its right the geologic model given the minimum estimated distance value according to Eq 3.8. The location A exemplifies the softmax transformation. The red arrow on the left shows the estimated distance for each category in a bar graph, and on the right, the transformed estimates on values scaled 0 to 1. The four maps at the bottom present the probabilities associated to each domain after transformation is applied.

3.3.3 Use of Transformed Distances for Global Domain Proportion Correction

In geostatistical modeling, one of the main concern is to avoid the introduction of bias in the geologic models. The bias may be introduced through interactions between the algorithm and parameters, geologic variability, trends models and local particular geologic configurations.

Typically, drilling campaigns focus on areas of high grades as future profits are expected to be obtained during the first years of mining operation and more information is desired in these areas. This is known as *preferential sampling*, and may lead to biased global proportions of geologic domains in the deposit if not properly treated. There are several declustering techniques that account for the preferential sampling problem such as polygonal, cell declustering, nearest neighbor and kriging weights.

This section proposes an extension of the algorithm to generating geologic models that matches these representative statistics of domains. The algorithm is based on the probabilities obtained after *softmax* activation function is applied, and follows

the principles of the servosystem correction tool proposed by Strebelle and Journel (2000).

Servo System Correction

The servosystem is a mechanism that allow gaining control over the reproduction of pre-established set of target marginal distribution of domains. The core of the algorithm consists of updating the probabilities of the domains based on the difference between the target and current marginal distributions. The amount of the correction is proportional to the magnitude of the difference. Eq 3.10 shows the servosystem equation.

$$P(i(\mathbf{u}) = k)^{updated} = P(i(\mathbf{u}) = k) + \mu(p(k) - p_k^c(\mathbf{u})) \tag{3.10}$$

The current probability $p_k^c(\mathbf{u})$ denotes the proportions of previous estimates nodes classified as domain k when visiting the location $\mathbf{u}$; $p(k)$ corresponds to the target marginal distribution and μ is the servosystem correction parameter defined as $\mu = \dfrac{\lambda}{1-\lambda}$, there is more of a correction when λ is closer to one.

The Algorithm for Domain Proportion Correction

The servosystem correction is embedded in the multiple distance function algorithm. The difference relies in the use of the transformed estimate probabilities after applying the softmax function in Eq 3.10. After the probabilities have been updated, the classification of domains is no longer determined by the minimum estimated signed distance, but by the maximum posterior probability. Thus, the classifier presented in Eq 3.8 becomes:

$$i^*(\mathbf{u}) = \operatorname*{argmax}_{k} P(i(\mathbf{u}) = k)^{updated} \tag{3.11}$$

Algorithm 1 summarizes the methodology. First, the nodes are visited randomly, this does not alter the final estimates, but avoids the introduction of artifacts when updating the probabilities. At each node location $\mathbf{u}_i$, the kriged signed distance is

determined and transformed into posterior probabilities using Eq 3.9. The next step considers the computation of the marginal distributions for each domain including all previous estimates up to location $\mathbf{u}_i$. Then, the transformed probabilities are updated according to Eq 3.10. Finally, the domains are classified based on the maximum updated probability. The process ends when all the nodes have been classified.

Algorithm 1 Multiple Signed Distance Function Implicit Algorithm Equipped with Servo System Correction

---

Input: Samples database, variography of domains, estimation parameters and targeted proportions of domains
Output: Models of geologic domains
1: Code the samples for each domain k as $i_k(\mathbf{u}_\alpha)$
2: For all samples, compute the signed distance measure $d_k(\mathbf{u}_\alpha)$
3: Generate random path throughout the grid node
4: **for** each node $\mathbf{u}_i$ $\{i = 1, ..., (n_x n_x n_x)\}$ in the grid **do**
5:    **for** each domain k **do**
6:       Estimate the distance using ordinary kriging: $d_k^*(\mathbf{u}_i)$
7:       Update posterior probabilities, $P(i(\mathbf{u}_i) = k)^{updated}$, using Eq 3.10
8:    **end for**
9:    **if** $\operatorname{argmax}_k P(i(\mathbf{u}_i) = k)^{updated} = k$ **then**
10:       Node $\mathbf{u}_i$ is classified as domain $i^*(\mathbf{u}_i) = k$
11:    **end if**
12: **end for**

A Fortran-based program in GSLIB-like format called *dfmod* was written to implement the algorithm; details of the code may be found on Appendix B.1. The use of servosystem in the flowchart modeling is optional and its activation is determined by the user.

3.3.4 Discussion

The proposed algorithm is an extension of the binary signed implicit distance version. A classifier based on the minimum estimated distance allows establishing boundaries between geologic domains. The improvements of the proposed extension:

- Multiple signed distance function algorithm is straightforward, fast and avail-

able for multiple rock type modeling in an implicit framework. It is not longer necessary to contour a complete set of sections and plan views of multiple domains.

- The extension from binary to multiple domains is computationally efficient. In the case that a single variogram is used, the ordinary kriging system of equations is solved once for all domains.

- The distance function approach avoids overlapping issues among domains. No priority criterion is needed.

- A post-processing transformation of the estimated distance provides of a heuristic measure of uncertainty.

- The softmax transformation of estimated distances offers a platform for handling representative global proportions in the resulting implicit geologic domains.

Smooth boundaries are produced in the models due to the implementation of ordinary kriging. Although there are no complex short scale geologic features, the large geologic features are adequately addressed.

3.4   Application: Multiple Domain Porphyry Deposit

A multi-domain porphyry deposit serves as a case study for demonstrating the algorithm. In order to maintain the confidentiality of the project, there is no mention of its geographical location, grades values, and geologic domain mapping. The example is used only for illustrative purposes.

The deposit is composed of five domains, three intrusive bodies and two oxide-sulphide hosted domains. The spatial distribution of samples colored by domains is shown in Figure 3.4. In yellow, magenta and cyan, the intrusives 1, 2 and 3 respectively; in green and red, the oxide and sulphide domains.
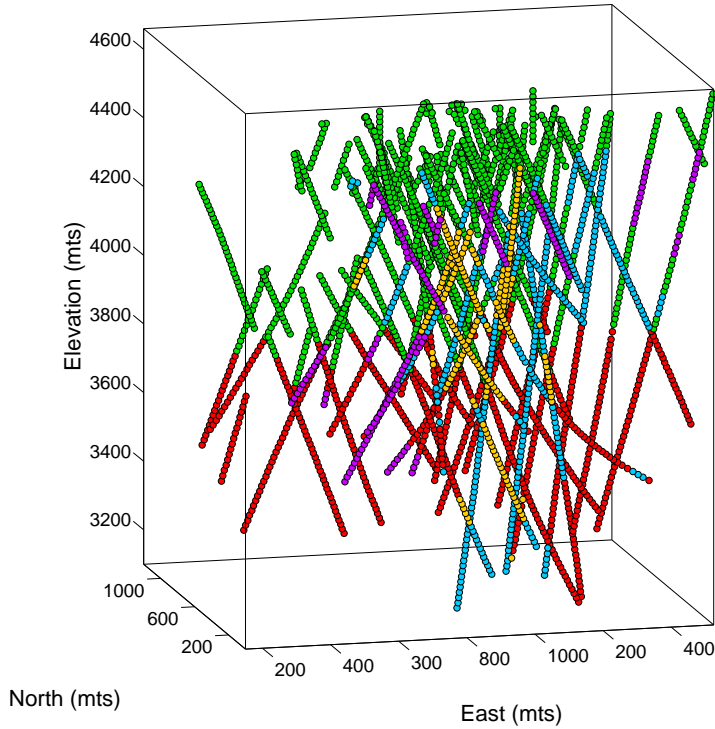
Figure 3.4: Samples scattered plot. In cyan, magenta and yellow the intrusive mineralized domains; in green and red, the oxide and sulphide host rock.

3.4.1 Parameter Selection

Geologic modeling for mineral deposits is often performed on a gridded model. The grid for this case is presented in Table 3.1.

Direction	Number of Blocks	Size of Blocks (fts)
Northing (fts)	52	25
Easting (fts)	48	25
Elevation (fts)	90	15

Table 3.1: Parameters of block model for distance function estimation and classification of domains.

Figure 3.5 shows five plots of the data colored by signed distance values computed according to Eq 3.6. Note the trends of the signed distance and the spatial distribution of high and low values. Special care is needed for this situation, the non-stationary may lead to extrapolation problems at the edge of the block model.

58
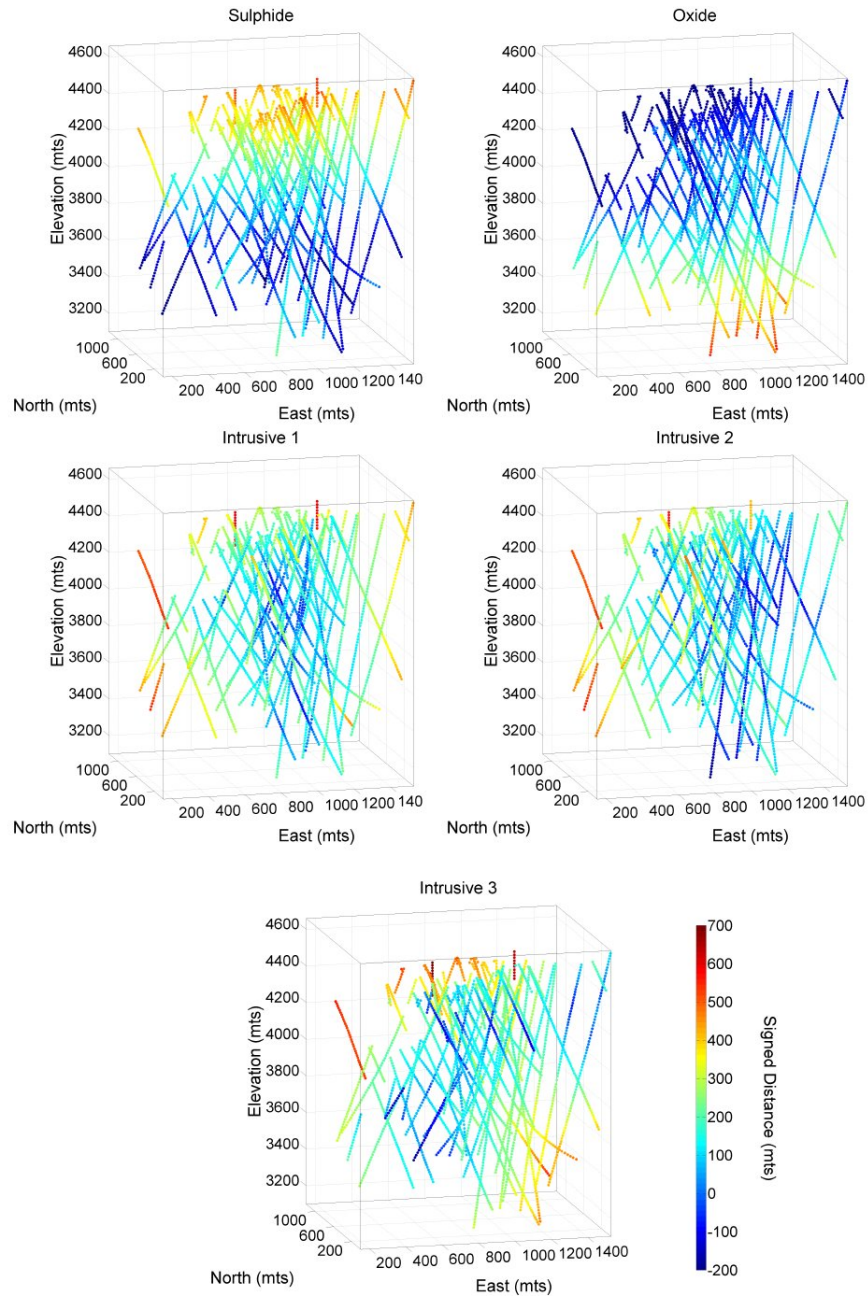
Figure 3.5: Scattered plot of the samples colored by signed distance. The distance is negative when the sample belongs to the current domain and positive in the contrary case.

Figure 3.6 shows the experimental and modeled variograms for each domain. Omnihorizontal and vertical directions are considered, no preferential directions were found. Due to non-stationarity of the distance values, the variograms do not have stationary sills. Also, the linearity of the distance make that the origin behave
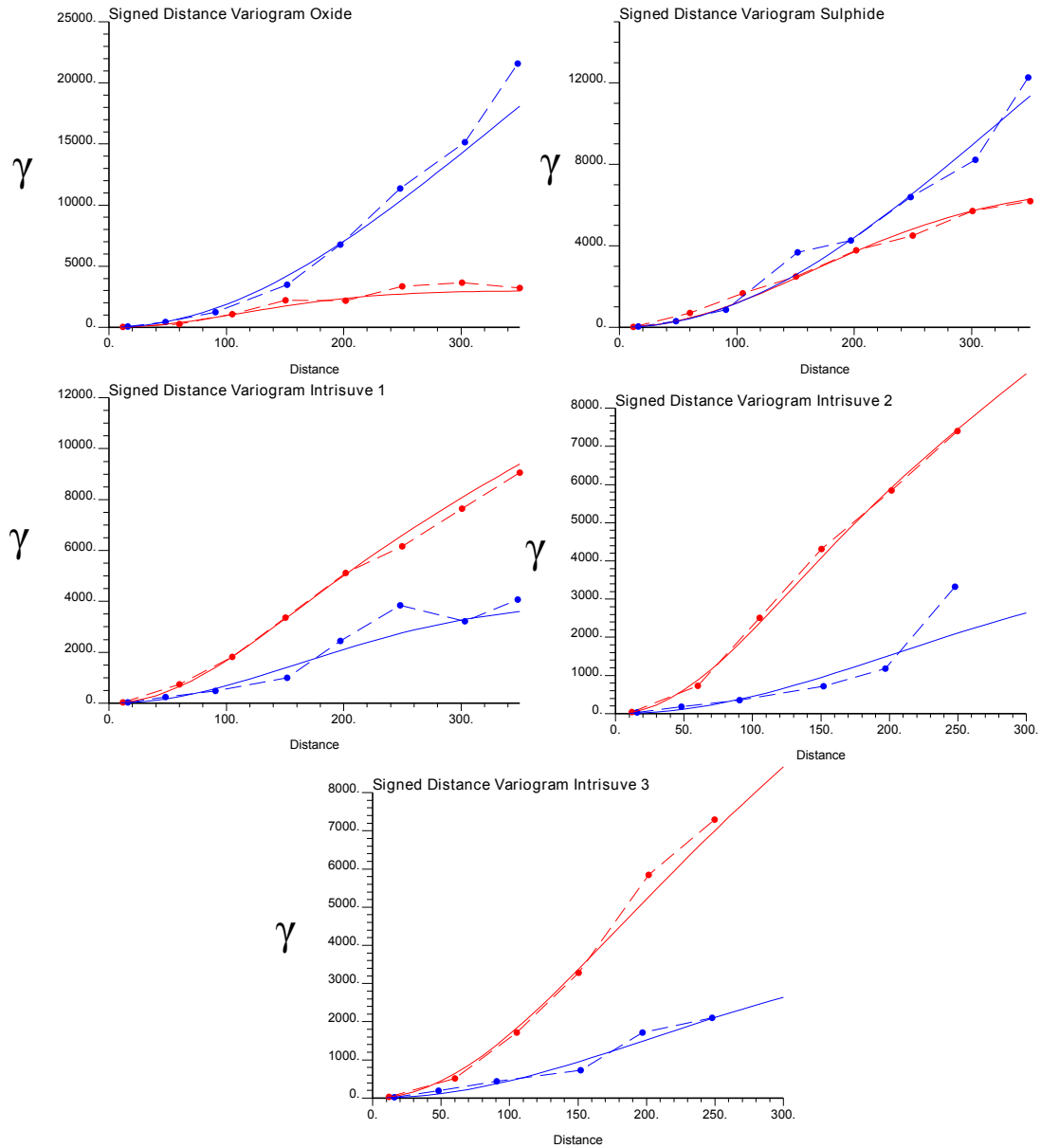
Figure 3.6: Experimental and modeled variogram for each domain. The linear nature of the signed distance provoke that the variograms present a square growth at the origin; Gaussian models were used. In red, the omnihorizontal variogram; in blue, the vertical direction.

close to a quadratic form. The gaussian model is a well-suited structure for this type of variograms. Table 3.2 shows the nugget effect, the range, and the sill of the variogram models. Two structures are used to model all directions.

Mode	Nugget Effect	First Structure			Second Structure		
		Range$_{xy}$	Range$_z$	Sill	Range$_{xy}$	Range$_z$	Sill
Intrusive 1	0.0	320	500	4,000	820	∞	15,000
Intrusive 2	0.0	400	800	7,000	∞	800	20,000
Intrusive 3	0.0	300	400	4,000	900	∞	15,000
Oxide	0.0	250	500	4,000	820	∞	15,000
Sulphide	0.0	320	500	4,000	820	∞	15,000

Table 3.2: Parameters of variogram models for each geologic domain.

3.4.2 Implementation

Before the implementation, a practical aspect involved in the use of global ordinary kriging need to be clarified. Computational requirements restrict the implementation of global kriging for projects that involve a large amount of samples. Typically, kriging related programs may deal up to a maximum of 10,000 samples at once in order to maintain mathematical stability of computations. Instead, a large neighborhood is considered.

A total of 224,640 blocks were estimated. Table 3.3 summarizes the parameters used for kriging estimation. The dimension of the search neighborhood is approximately half of the area under study. All domains use the same parameters, this provides consistency to the models; even though, different parameters could be used if desired.

Estimation Parameters	$Radius_x$	$Radius_x$	$Radius_x$	Number of Samples
All Domains	700	1,000	1,000	128

Table 3.3: Ordinary kriging estimation parameters.

Isometric views of the modeled intrusive and hosted rock domains are presented in Figure 3.7. The models in A, B and C display the intrusives individually; image D summarizes the entire model with the oxide and sulphide hosted units at the background. The intrusives appear rather continuous, and no artifacts are observed in these simple views. The transition between domains does not show abrupt changes or artificial geologic arrangements. The continuity of domains shown by the samples is maintained in the models, and isolated objects are minimized. The oxide-sulphide
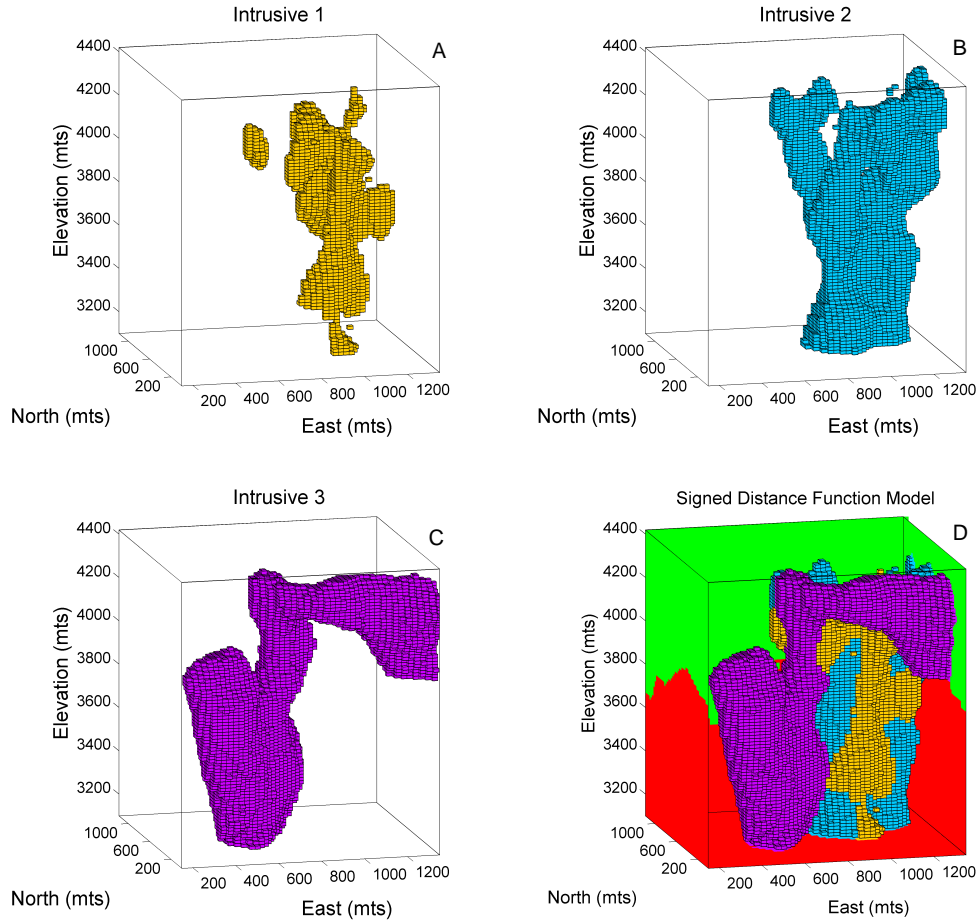
Figure 3.7: Implicit geologic model for the porphyry deposit constructed by the proposed multiple distance function algorithm. Each intrusive domain is represented individually, and the entire model portrayed on the bottom-right corner.

boundary presents a natural transition at distinct depth.

Figure 3.8 complements the analysis providing of six cross-sections oriented along east-west and north-south directions. Intrusives are well modeled, maintaining a representative spatial organization for this type of deposits. The samples are also displayed for comparison purposes; the data are reproduced. There is no bias in proportions and the shape, extension and geologic variability of the domains do not show artifacts or anomalous features.

A second illustration in Figure 3.9 shows maps of the estimated transformed probability values based on the softmax function (Eq 3.9). Transitions between
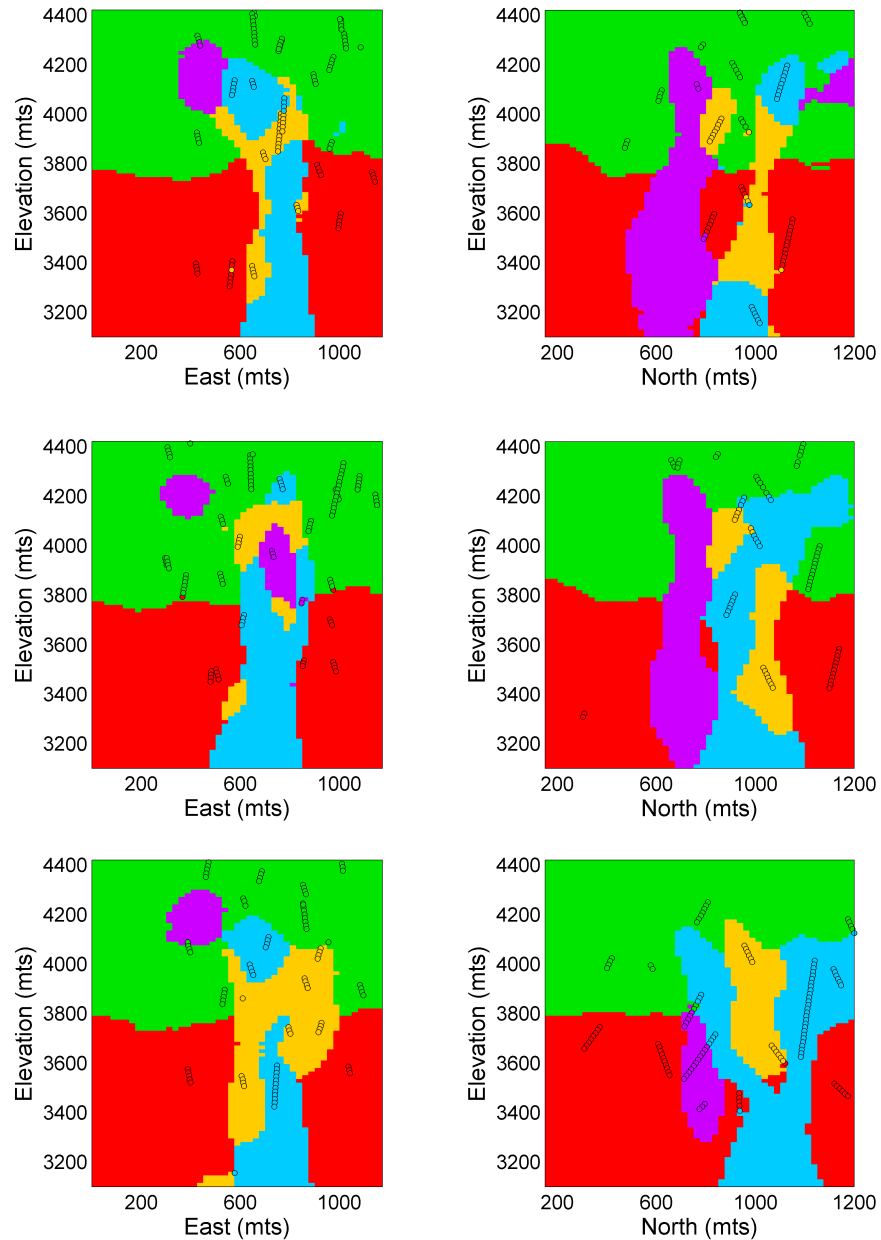
Figure 3.8: Six North and East cross sections of the model in Figure 3.7. In cyan, magenta and yellow the intrusive domain; the green and red correspond to the oxide sulphide contact. The samples are also deployed. No visual evidence related to the introduction of bias in the model is observed.

domains are well described by the probability maps. To evaluate the effectiveness of the posterior probability, the samples are also displayed; in black, the data that belong to the actual domain, in white the samples of remainder domains. The ap-
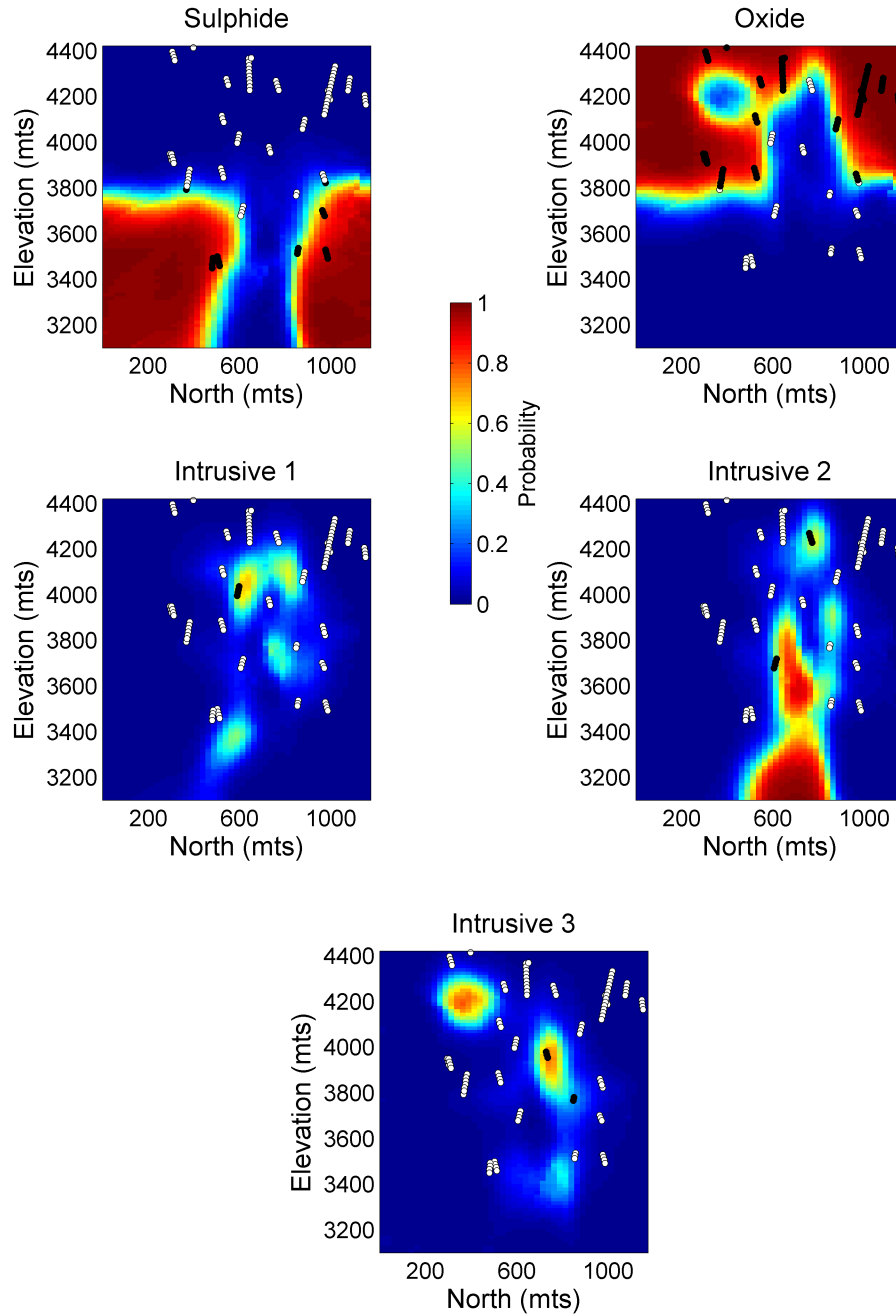
Figure 3.9: Representation of the estimated posterior probability. The cross section in the middle on the left of Figure 3.8 was used. The black dots represent the samples that belong to the current domain and the white ones the samples of any opposite domain.

proach serves as a heuristic complementary analysis and provides a general reference to aspects of uncertainty.

3.4.3 Example of Global Domain Proportion Correction

In this example, the performance of the servosystem method is evaluated. The target proportions for the corrections of the five domains are set up to different values and the servosystem applied. Three cases are assessed; the same variogram models and estimation parameters are used for all the cases. Table 3.4 shows the resulting proportions in the models compared to their targets for each domain. Case 1 and 3 match perfectly the target, subtle differences are observed in Case 2 for the oxide and sulphide domains.

Global Proportion Servo System Correction Results						
Domains (%)		Oxide	Sulphide	Intrusive 1	Intrusive 2	Intrusive 3
Case 1	Target	40	38	4	8	10
	Reached	40	38	4	8	10
Case 2	Target	35	33	7	11	14
	Reached	36	34	6	10	14
Case 3	Target	48	45	1	2	4
	Reached	48	45	1	2	4

Table 3.4: Resulting global proportion of geologic domains after servo system correction.

Figure 3.10 presents the geologic models for the three cases. Each column represent one target setting of domain proportions. The models are compared to the original model considered in Section 3.4. Case 1 shows similar results since the target proportions were set to the same proportions as the original model. Different results are seen for cases 2 and 3. Case 2 considers models where the proportions of intrusive units are increased. From the images observed (second column), all intrusive units shown a substantial increase in their size. The target proportions of the intrusives are decreased in Case 3. The size of the intrusive is constrained to small areas around the conditioning data. Issues related to conditioning of the samples may appear if the models are corrected significantly. The algorithm relies on coupling the servosystem correction with representative statistics determined by well-established declustering methods.
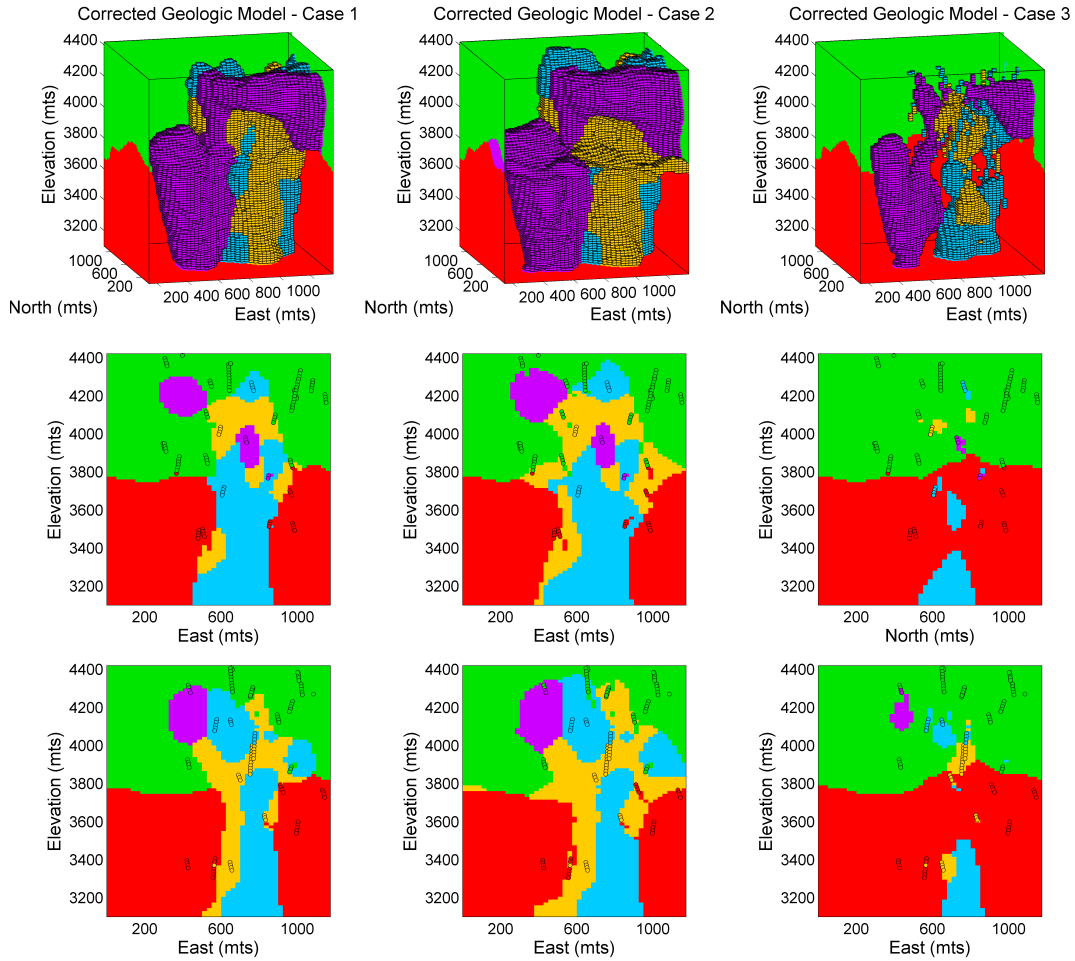
Figure 3.10: Geologic models for three different configurations of global domain proportions; an isometric view and two cross sections for each case are presented. The first column corresponds to the global proportions: 40%, 38%, 4%, 8%, 10%; the second column set to 35%, 33%, 7%, 11%, 14%, and the third column the model set to 48%, 45%, 1%, 2%, 4% for the oxide, sulphide, intrusive 1, 2 and 3 respectevely.

3.4.4 Accounting for Boundary Extrapolation of Domains

A concern of the multiple distance function approach relates to the extrapolation of the estimated domains influence of samples. This requires a sound knowledge of the characteristics of the deposit. There exist situations where the domains need to be constrained to control excessive extrapolation and to avoid the introduction of bias in the models.

In particular, the spatial configuration of samples plays a fundamental role in the
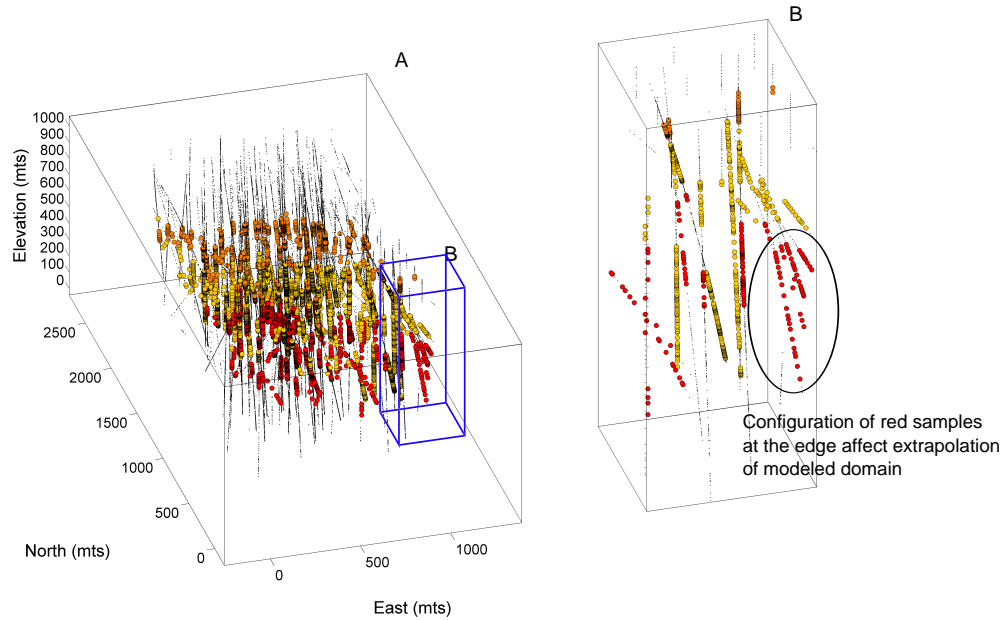
Figure 3.11: Example based on a gold deposit. On the left, isometric view of the scattered samples; four domains are included red, orange and yellow represent the folded sedimentary units, the smaller black dots the host rock. On the right, a zoomed view of the blue rectangle of figure A. The samples within the black ellipse represent a risk of extrapolation.

extrapolation of distance estimates. Depending on this, there may exist situations where the extrapolation may not be controlled satisfactorily. A real case example based on a folded-sedimentary gold deposit is introduced in order to analyze the causes and to propose alternatives for handling the extrapolation problem.

Figure 3.11 on the left presents an isometric view of the samples. The red, yellow and orange samples portray three folded-sedimentary domains, the small black dots represent the host rock. The blue rectangle B on the left is zoomed in on the right to enhance the visualization of the configuration of samples at the edge of the model. The set of red samples inside the circle is an illustrative example of a potential area of uncontrolled extrapolation. The drillholes within the circle do not have samples of host rock nearby. This situation may occur repeatable times in the deposit and typically will produce extrapolation problems beyond the limits of the grid node.

To observe the consequences of uncontrolled extrapolation, the multiple signed distance function algorithm was applied, see Figure 3.12 on the left for the illus-

67
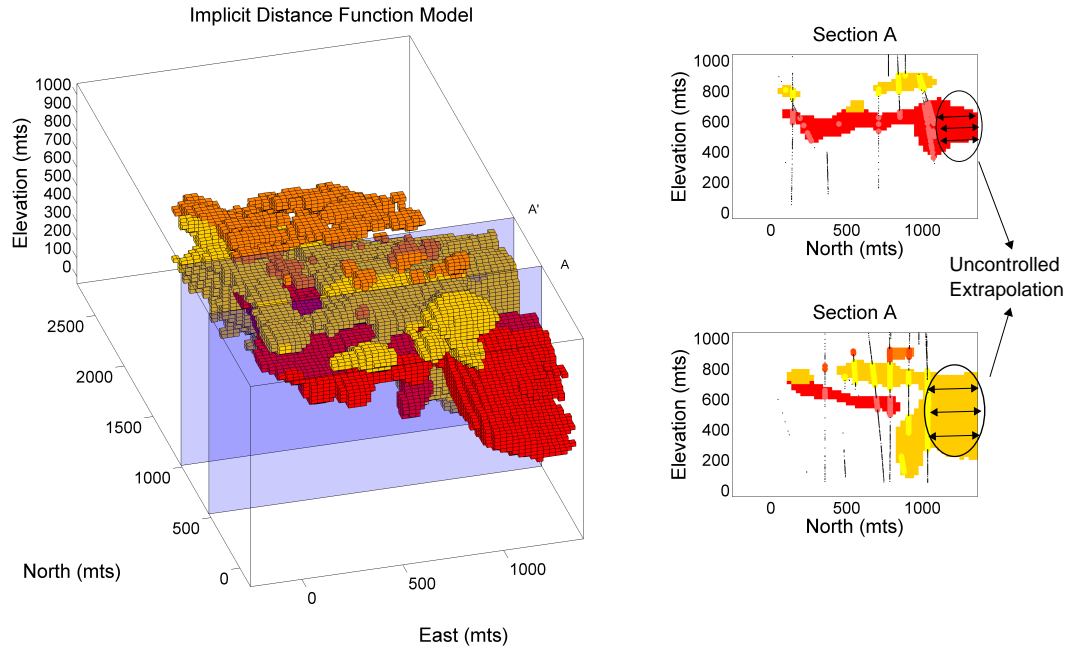
Figure 3.12: On the left, the entire model in three-dimension, the blue cross sections A and A' are plotted on the right of the figure. Extrapolation of domains red and yellow is highlighted.

tration. The image also shows an exaggerated representation of the red domain on the north-east corner. Two blue slices that cut the model are shown as sections A and A' on the right. The final model and the scattered samples are also plotted for visual inspection. Both sections show how the red and yellow domains depart excessively from the limits of the model.

Extrapolation depends on the kriging weights applied to the local conditioning samples, see Eq 3.7. Thus, when locations far away from samples are estimated, the kriged signed distance estimates become relatively constant. This situation causes the issues described above.

An alternative treatment consists of implementing a restriction upon the estimates beyond the range of the variogram, and applying the proposed servosystem correction. The approach proceeds as follows: all the nodes beyond the range of the variogram are automatically assigned to the host rock limiting the extrapolation. Local problems as the tendency to generate isolated structures and abrupt transi-
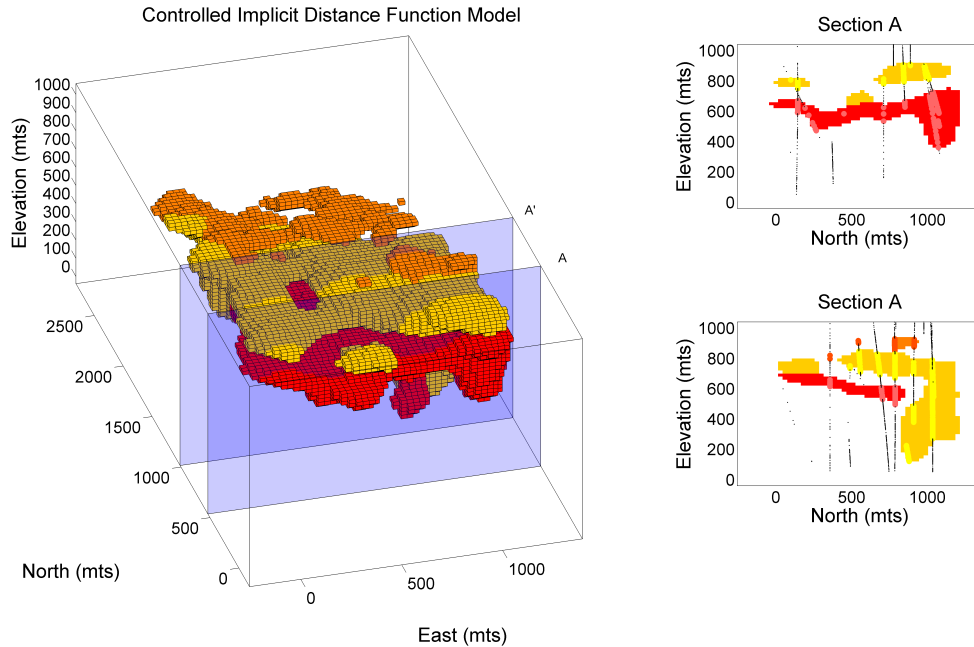
Figure 3.13: Same model as Figure 3.12. The nodes beyond the limit of the variogram are not estimated and directly assigned to the host rock domain.

tions are solved by a servosystem. Figure 3.13 shows an example of the proposed corrector. Sections A and A' show how the extrapolation is controlled up to 150 meters (the range of the variogram), and the application of servosystem helps to maintain the coherence in the model.

Other alternatives could also be considered such as the implementation of artificial controlling samples at the edge of the models. The major disadvantage of this is that, without proper geologic knowledge, artifacts could be introduced making the models even worse. Another alternative consists of applying the distance function algorithm in its primal binary version hierarchically, Section 3.1, change to simple kriging and manipulate artificially the mean to manage the extrapolation. However in both cases, the additional manipulation would harm the simplicity of the algorithm.

69

3.4.5 Discussion

The proposed multiple signed distance implicit algorithm is able to reproduce and transfer large scale geological features into numerical models. The method is not restricted by the number of domains considered. The geological information is incorporated through the calculation of signed distance, but also, by the directions of continuity in the variograms models. The use of large searching parameters for the estimation process approximates the results as if global kriging were applied. Some additional practical considerations are addressed below.

3.5 Assessment and Practical Considerations of the Algorithm

Geologic modeling algorithms must satisfy certain principles to be considered as reliable. A detailed investigation developed by McLennan (2007) established six principles related to concepts including *simplicity*, *speed*, *objectivity*, *data integration*, *access to uncertainty* and *geologic realism*. The multiple signed distance function is evaluated in the light of these principles, pointing out its strengths and weakness.

Simplicity: The implementation of the algorithm is straightforward. No complicated mathematical functions are needed. The idea of a signed distance to a different domain is easy to understand.

Speed: The core of the algorithm relies on the implementation of kriging, and kriging methods are quite fast. Some limitation comes from the use of a large number of samples.

Objectivity: The approach was devised to overcome the subjectivity involved in explicit geologic modeling methods. Reproducible models are expected when estimation parameters remain constant. The models are not subject to arbitrariness of one modeler.

Data Integration: The algorithm is flexible to incorporate additional information of the same type; however, it has not been designed to account for secondary information.

Access to Uncertainty: The algorithm does not account for uncertainty. An alternative based on the transformation of the distance values allow incorporating an heuristic measure of uncertainty into the models.

Geologically Realistic: Models are considered to be realistic when they provide an acceptable level of agreement to geologic interpretation and the evidence gathered by the modeler. The agreement comes from the fact that the spatial continuity of geologic domain is properly reproduced, that is, reproduction of large scale structures and their placement in the space. However, the short scale resolution is questionable. The smoothness property of kriging leads to geologic models with smooth transitions that may not be realistic.

The algorithms are dependent on the input parameters and may show instability. In this sense, a seventh principle is added to the assessment: the *robustness* of the algorithm.

3.5.1 Robustness of Algorithm

The robustness of the algorithm is checked through sensitivity analysis on the parameters. In particular, the concept of robustness refers to the generation of slightly less efficient models, but that still provide acceptable results when large deviations are applied to the parameters.

Variograms models and estimation parameters are considered for sensitivity analysis. Global proportions of domains and visual inspection are used to evaluate the robustness of the algorithm. Comparisons are made with the porphyry case study shown in Section 3.4. Three scenarios are proposed: 1) change the sill and range for all domains simultaneously, 2) modify the variogram model for an individual domain and 3) change the number of samples considered for estimation.

Variation of Variograms Model for All Domains: The nugget effect of the variogram model is modified in 10%, 30% and 50%, whereas the range changed in 50% and 200%. Table 3.5 presents a summary of the global proportions for each domain after the modifications are applied; the global proportion of the base case is also presented in the first column. Figures 3.14 and 3.15 provides isometric views of
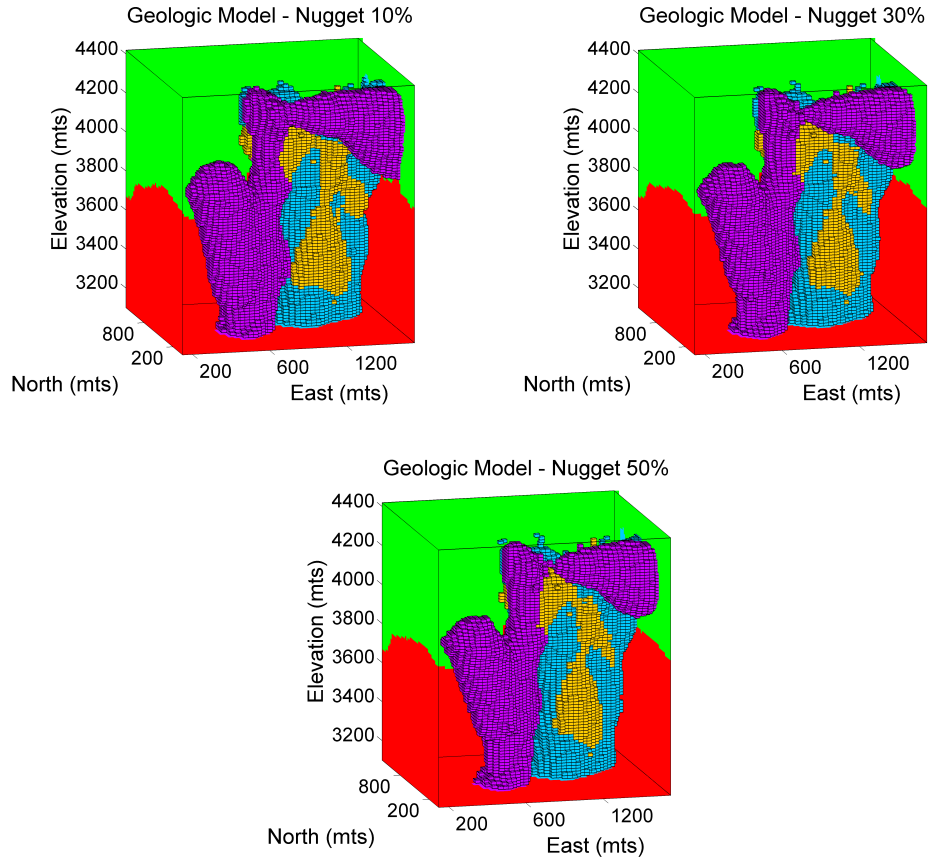
Figure 3.14: Models for three different nugget effect in the variogram model. The values used were 10%, 30% and 50% of the original nugget effect.

the models. In general, the spatial organization among domains remains similar to the base case (Figure 3.7). The large scale general shape and orientation remain the same. However, some changes in the models may be linked to the nugget and variogram range parameters. The increment to the nugget effect disconnects the intrusive domains. This is illustrated by the purple domain in Figure 3.14 that tends to disconnect from the rising trunk as the nugget effect increases. Thus, less continuous variogram models lead to disconnected bodies. Table 3.5 quantifies the previous effect, the global proportion of all intrusive domains decreases. The same effect is noted when augmenting the range of the variogram models. Table 3.5 and Figure 3.15 shows similar results, an increment in the proportions of the intrusive.

Variation of Variogram Model for Intrusive 3: Here, the variogram model is

Percentage Domain (%)	Base Case	Nugget Effect			Range	
		10%	30%	50%	50%	200%
Intrusive 1	2.1	1.8	1.6	1.4	1.9	1.7
Intrusive 2	5.7	5.8	5.8	5.8	5.6	6.1
Intrusive 3	8.4	7.6	6.7	6.0	6.5	9.9
Oxide	42.8	43.4	44.0	44.4	44.3	41.3
Sulphide	40.9	41.3	41.8	42.2	41.6	40.9

Table 3.5: Resulting global proportion of geologic domains when modifying variogram models.
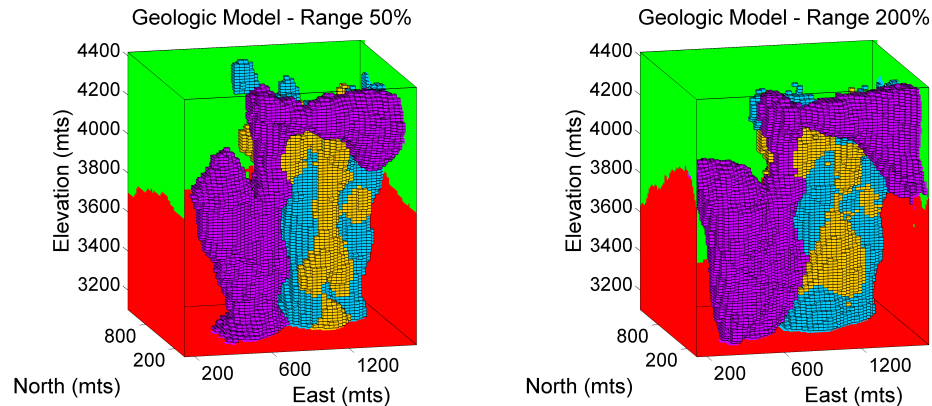
Figure 3.15: Geologic models when the range of the variogram for all the domains is changed to 50% and 200% respect to the original variogram.

modified individually for intrusive 3. Results are shown in Figure 3.16 and Table 3.6. The intrusives 1 and 2 remains equal in terms of global proportion if compared to the base case. The major change is related to the relation between the intrusive 3 and the host rock. The proportion of intrusive 3 decreases when the range is modified to 50% of the original range, the contrary is observed when using 200 % of the range. The changes to intrusive 3 may be seen on the right side of the Figure 3.16 where the absence of conditioning samples allows spreading the intrusive.

Variogram models influence the shape and spatial extension of the geologic domains. High nugget effect and low range in the variograms lead to generating isolated geologic objects. On the contrary, continuous variogram models generate larger objects.

Number of Samples: A third element to be considered is the number of sam-

Percentage Domain (%)	Base Case	Range	
		50%	200%
Intrusive 1	2.1	2.1	2.1
Intrusive 2	5.7	5.8	5.8
Intrusive 3	8.4	7.9	8.6
Oxide	42.8	41.1	40.1
Sulphide	40.9	43.1	42.8

Table 3.6: Resulting global proportion when modifying variogram model of domain intrusive 3.
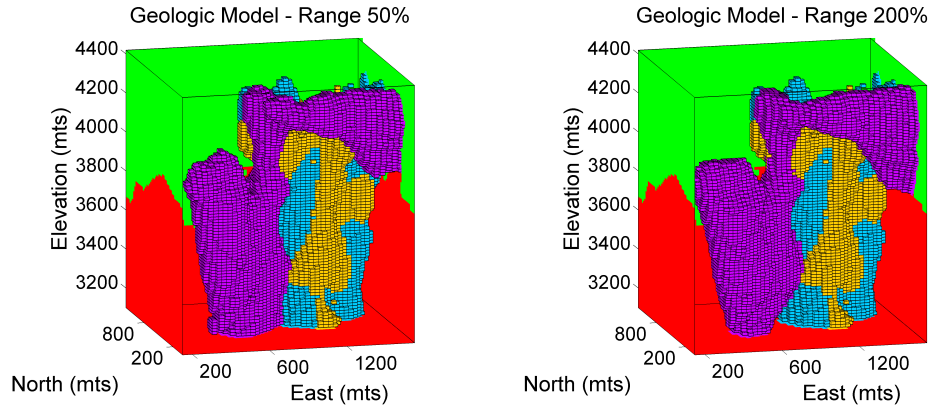
Figure 3.16: Same case as Figure 3.15, but the range of the variogram was only changed for the intrusive purple domain. The objective is to analyse how sensitive is the model when one individual domain variogram model is altered.

Percentage Domain (%)	Base Case (128 Samples)	Number of Samples		
		32	64	96
Intrusive 1	2.1	2.1	2.1	2.1
Intrusive 2	5.7	5.8	5.7	5.8
Intrusive 3	8.4	8.8	8.2	8.3
Oxide	42.8	43.2	43.1	42.8
Sulphide	40.9	39.9	40.6	40.9

Table 3.7: Resulting global proportion when modifying the number of conditioning samples for signed distance estimation.

ples used for signed distances estimation. The effect of the number samples upon the models may be observed in Figure 3.17 and Table 3.7. A maximum number of samples of 32, 64 and 96 was applied. Visually, unrealistic edges and transitions of objects are observed when using 32 and 64 samples; the domains deviate substantially from the base case Figure in 3.7. In the case of 96 samples, the global
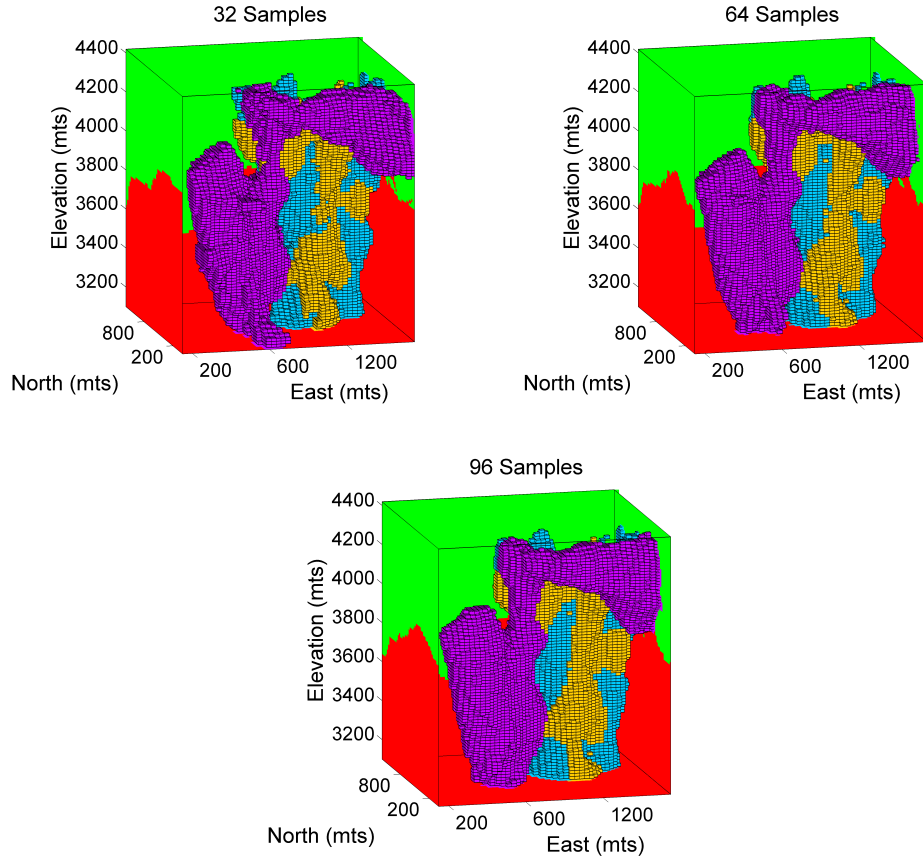
Figure 3.17: Geologic models using 32, 64 and 96 samples for ordinary kriging estimation.

proportions of domains remain similar to the base case.

Variograms models and estimation parameters may affect the final proportions of the geologic domains. This is important since there is some risk of introducing bias in the models. Bias in the global proportions of domains might affect the quantification of tonnage and grade of minerals attributes.

3.6 Discussion

An alternative geologic modeling approach based on signed distance function estimates has been presented. The method accounts for objectivity, reduces the time-consuming interpretative work, facilitates auditable revisions and offers an easy manner to update geologic modeling with new information. It is straightforward

to be implemented and does not require heavy parameterization. A post-processing correction for global proportions of the geologic domains is proposed. It adds modeling flexibility and allows maintaining coherence in the models by matching previous established representative global geologic domain proportions.

The proposed implicit geologic modeling algorithm shows promise results for mineral resource evaluation. Future avenues of research would include the implementation of a grid-free scheme that facilitates the construction of wireframes at any resolution. The introduction of local trends maps for non-stationary modeling. Developing hybrid techniques considering explicit, implicit and stochastic approaches also remains an interesting path of investigation.

Chapter 4

MPS Simulation with Multiple Data-Driven Training Images

This chapter discusses the use of multiple data-driven training images for spatial modeling of geologic domains through a scheme of data integration of conditional probabilities known as a linear opinion pool. The proposed methodology reproduces features from the training images with the correct amount of geologic continuity and variability.

Section 4.1 presents an introduction to the concept of data-driven training images and discusses the advantage of their integration in the context of geologic variability reproduction.

Section 4.2 provides a general description of the proposed methodology. Details are presented and discussed in the subsequent three sections. In particular, **Section 4.3** describes the process of constructing data-driven training images. **Section 4.4** explains the fundamentals of the integration of data-driven training images based on the linear opinion pool scheme. **Section 4.5** describes the calibration process for determining the correct degree of mixing between the training images.

The last two sections are devoted to the analysis of the realizations generated by the methodology. **Section 4.6** reviews the reproduction of expected geologic features when considering the linear integration of training images and **Section 4.7**

evaluates the performance of the geostatistical model.

Finally, a brief discussion of the results and final remarks of the method are presented in **Section 4.8**.

4.1 Introduction

One concern with multiple point statistics methods is find a valid and suitable training image that reflects the level of prior geologic knowledge of the reservoir/deposit. Reproduction of geologic structures contained in the training image is not guaranteed if the conditioning data are not consistent with the image (Journel, 2004). If the link between the data and training image is not strongly supported by geologic knowledge, the models may not correctly represent the spatial variability. Journel and Zhang (2006) demonstrated large differences in flow performance of models with the same histogram and covariance function, but differing multiple point geological structures. The prior geologic model defined by a training image requires a minimum level of agreement with the data to ensure a consistent reproduction of data and prior features. Thus, generating a training image based on data would guarantee consistency and local relevance, providing strong site-specific geological support.

To use training images conditioned by data is particularly interesting in mining projects since there are many more data than most petroleum reservoirs. The application of MPS methods is not particularly popular in mining partially due to the difficulty of obtaining a representative training image where the depositional and alteration processes are quite complex. Some methods have been proposed. (Boisvert, 2007) developed a genetic geologic modeling process for vein type and weathered lateritic type deposits. The use of exhaustive amount of data (blastholes) to model nearby zones that share common geologic structures has been also applied (Perez, 2011; Boisvert, 2007). However, these approaches do not explicitly account for the risk of a mismatch between the data and the training image. In this respect, the following chapter develops the concept of a data-driven training image to the task

of establishing geologically acceptable realizations.

This chapter presents an innovative method that incorporates valuable information from two valid, but individually unrealistic, deterministic and stochastic geologic interpretations (data-driven training images). The advantage of utilizing more than one training image is the ability to capture distinct heterogeneities, variability and multiple point configurations from each image and combine them. The proposed methodology reproduces the correct geologic spatial variability and continuity of domains supported by the conditioning data. By combining the conditional probabilities estimated from both images, the approach is able to model high-order relations that range from the original deterministic to stochastic training images.

4.2   Description of the Method

A variant of Multiple Point Statistics simulation is developed to use multiple (two in practice, but the methodology is flexible) training images. The general outline of the proposed methodology is summarized in Figure 4.1. It consists of four stages: (1) the construction of data-driven training images, (2) the generation of realizations based on the integration of training images using the linear opinion pool (LOP) (Stone, 1961), (3) the calibration process, and (4) the simulation based on calibrated weights. A detailed description of each stage involved is provided in separate sections.

(1) Construction of Training Images:

Deterministic and stochastic data-driven training images are constructed based on the same conditioning data.

(2) Generate a Set of Combined Simulations:

The idea of the proposed method is the integration of the conditional probabilities inferred from both training images using the linear opinion pool combination scheme in the sequential MPS simulation framework. Algorithm 2 in Section 4.4.3 provides
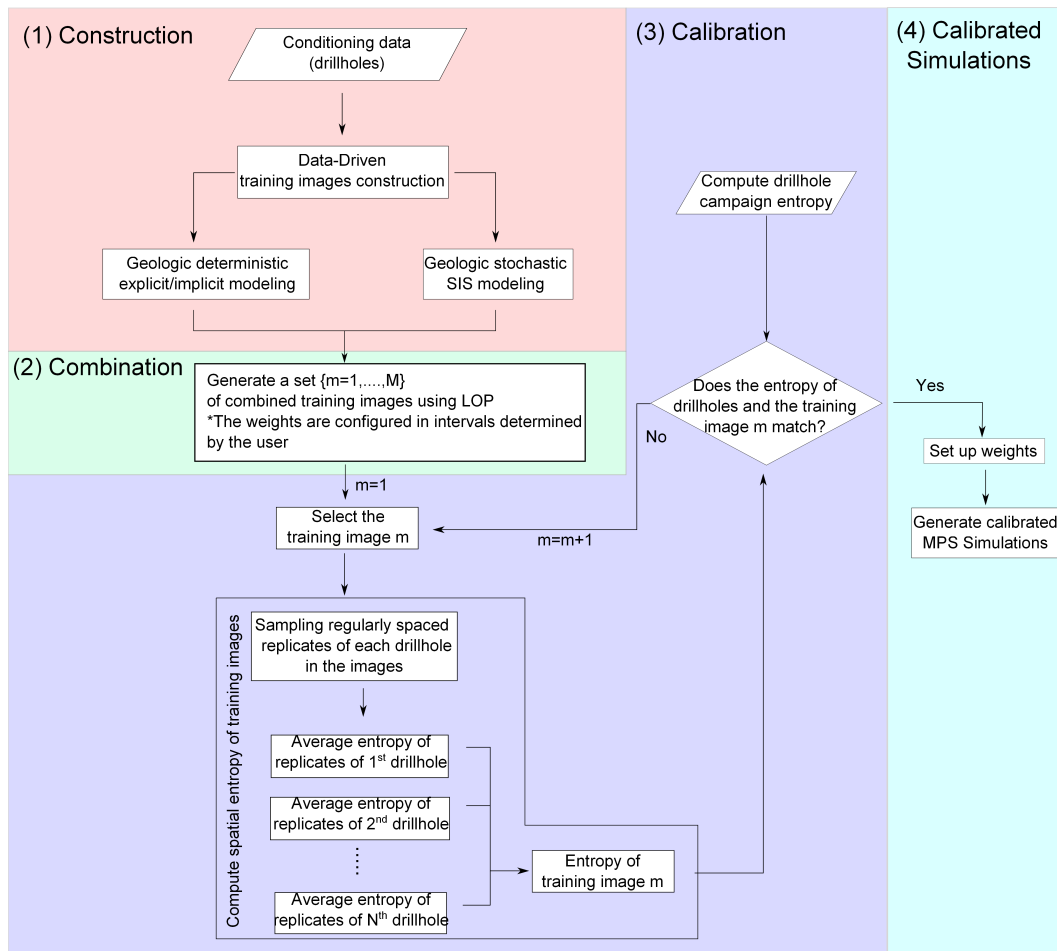
Figure 4.1: Schematic flowchart of the proposed algorithm. Four major areas are involved: in red, the process of construction of training images; the green shows the integration of multiple images for MPS simulation; in light blue, the calibration of the geologic variability between realizations and the drillholes; and in cyan, the generation of calibrated MPS realizations.

a detailed description of this process.

This second stage involves the generation of multiple realizations with a different combination of the training images. Each realization is the result of a particular linear combination based on weights set up in intervals determined by the user; these weights are specified within the $[0, 1]$ range.

(3) Calibration:

This stage will be reviewed in detail in Section 4.5. The concept of entropy (Journel and Deutsch, 1993) is introduced as a measure of the spatial variability of the geology. The set of realizations generated in the previous step are calibrated with respect to the spatial variability of the conditioning data.

(4) Generating Calibrated MPS Simulations:

Once the weights have been calibrated, the method generates realizations that reproduce the same amount of geologic variability as shown by the conditioning data. No other technique in current usage guarantees the correct modeling in the context of geologic variability reproduction.

4.3   Construction of Data-Driven Training Images

Both deterministic and stochastic training images represent interpretations based on the available data; the difference lies in the principles assumed for their construction. The LOP integration scheme unifies those principles in a flexible manner.

4.3.1   Geologic Deterministic Training Image (DTI)

A deterministic training image is generated by a smooth interpretative process. Both the explicit and implicit approaches described in Section 2 could be used. The first uses manual cross-sectional and plan view interpretations; the second generates three-dimensional objects by defining implicit functions in space. These approaches efficiently model large scale geological structures. The relation among distinct domains is well understood and the spatial arrangements are correctly set out, constituting reasonable large scale morphological arrangements.

The implicit geologic modeling technique based on signed distance functions developed and presented in Section 3.3 will be used. The method is incorporated into the flowchart of the algorithm to provide quick and geologically plausible deterministic models.

4.3.2 Geologic Stochastic Training Image (STI)

The construction of the stochastic training image is based on the use of SIS, but other geostatistical methods could be used. Even though the method can appear somewhat patchy and unstructured (Deutsch, 2006), it gives a good approach for simulating models with somewhat random geological transitions. A completely different training image in terms of continuity is produced by SIS. Unlike the distance function approach, the SIS methodology and applications are well documented in the geostatistics literature (Chilès and Delfiner, 1999; Deutsch and Journel, 1998; Goovaerts, 1997; Journel, 1983; Journel and Isaaks, 1984); however, the details are not reproduced here.

4.4 Integrating Training Images for Reproduction of Geologic Variability

This section addresses the problem of integrating multiple training images. The objective is to develop a MPS algorithm that allows combining the geologic features from deterministic and stochastic training images. Both linear and multiplicative data integration methods are explored. Examples provide insights into the linear integration of training images and the proposed method.

4.4.1 Problem Statement

Based on the terminology developed in Section 2.4, consider a data event $D_{m,n}$ centered at location $\mathbf{u}$ surrounded by n conditioning data that may take K values from the set of categories $\{s_k, k = 1, ..., K\}$. Now, consider that the data event is related to the training image m from the set of images $\{m = 1, ..., M\}$, that is, the inference of the conditional probability must be computed using the training image m. By definition the conditional probability of the indicator variable $S(\mathbf{u})$

that takes the state s_k conditioned to the data event $D_{m,n}$ is given by:

$$P(S(\mathbf{u}) = s_k | D_{m,n}) = \frac{P(S(\mathbf{u}) = s_k, D_{m,n})}{P(D_{m,n})} \qquad (4.1)$$

As reviewed in Section 2.4.1, this conditional probability is the ratio of the frequency of appearance of the joint event $\{S(\mathbf{u}) = s_k, D_{m,n}\}$ and the data event $D_{m,n}$ in the training image m. In the same manner, the conditional probability could be extended to the case of considering M multiple training images. Now, the problem consists of how to integrate the information from these multiple training images, that is, provide a solution for the posterior conditional probability:

$$P(S(\mathbf{u}) = s_k | D_{1,n}, ..., D_{m,n}). \qquad (4.2)$$

The correct solution would require knowledge of the joint multivariate distributions of $\{D_{1,n}, ..., D_{m,n}\}$ and $\{S(\mathbf{u}) = s_k, D_{1,n}, ..., D_{m,n}\}$. This information is not available because we do not have a repetition of multiple training images.

Integration of multiple data sources as in Eq 4.2 has been widely studied in the field of geostatistics (Allard et al., 2012; Hong, 2010; Journel, 2002; Polyakova and Journel, 2007). The limited and sometimes poor understanding among the sources of information calls for assumptions that considers each individual source in a suitable theoretical and practical framework. Different approaches have been proposed with different assumptions for different objectives. Two classes of probability combination models to solve Eq 4.2 may be found in the literature: those based on multiplicative and additive aggregation operators (Allard et al., 2012). In the context of combining training images, special attention has been placed on the additive linear method, specifically a branch of management science known as consensus theory (Winkler, 1968), where a weighted linear opinion pool offers a scheme with desirable properties. In particular, consensus theory considers the importance of weighting expert opinions to combine the conditional probabilities from each source of information.

4.4.2   Additive Aggregation: Linear Opinion Pool

Instead of estimating the unaccessible joint multivariate distribution in Eq 4.2, one could consider the integration of the conditional probabilities inferred from each training image $m = \{1, ..., M\}$ by a function $f(D_{1,n}, .., D_{m,n})$. A reasonable option for this function would consists of using a weighted linear integration; this scheme is known as linear opinion pool (Stone, 1961):

$$P(S(\mathbf{u}) = s_k | D_{1,n}, .., D_{m,n}) = f(D_{1,n}, .., D_{m,n}) = \sum_{i=1}^{m} w_i P(S(\mathbf{u}) = s_k | D_{i,n}), \quad (4.3)$$

where $w_i \geq 0$ and $\sum_{i=1}^{m} w_i = 1$. The result must be a consistent probability value between 0 to 1 and satisfy the closure property, that is, sum to unity. The linear opinion pool assumes that each training images is conditioned to the opinion of an expert. The problem is to quantify the weights w_i for each training image. These weights can be based on the expertise of the decision maker, determined by some previous mathematical model or, as presented here, calibrated based on the geologic variability of the conditioning data.

The linear opinion pool (Eq 4.3) is convex, which is a good property in the context of geologic variability reproduction. The combined probability falls inside the range of the individual probabilities computed from the training images. The convexity of the combined probabilities corresponds to a mixture model. The individual conditional probabilities have access to the same information but are considered differently. This redundancy justifies using the linear opinion pool for mixing geologic features instead of commonly used multiplicative aggregative methods in earth sciences, which sometimes lead to a non-convex solution due to the distinct information within each source of information.

4.4.3   Linear Opinion Pool MPS Combination Scheme Algorithm

Algorithm 2 summarizes the steps involved for MPS simulation based on the integration of training images using the LOP scheme. The grid node is defined (Line 1). All the nodes of the grid are randomly visited; this is represented by the loop

between Line 2 and Line 11. To simulate a node, the set of n hard data surrounding the current location is found, that is, the data event $D_{m,n}$ associated to location $\mathbf{u}$ (Line 3). At this point, the original MPS algorithm is modified; the data event is used to compute the conditional probabilities derived from all training images $m = \{1, ..., M\}$; a second loop that visits the M training images is defined between Lines 4 and 7. Given the unique evidence provided by the data event, the conditional probabilities express distinct information from each training image (Line 5). These probabilities are computed for each training image (Line 6). Then, the conditional probabilities are combined linearly according to weights w_i allocated to each image (Line 8), Eq 4.3. The posterior probabilities are updated and a categorical value drawn from the set of probabilities (Line 9), the value is added to the location and the process continues (Line 10). The simulation ends when all nodes have been simulated (Line 12).

Algorithm 2 MPS Simulation with Multiple Training Images

Input: Samples database, Data-Driven Training Images
Output: Simulations that combine geological features from both training images
1: Set up a regular grid of n_x, n_y, n_z and the size of nodes $xsiz, ysiz, zsiz$ over the area to be modeled
2: **for** each node along the random path $\{i = 1, ..., (n_x n_y n_z)\}$ **do**
3:    Retrieve the data event $D_n(\mathbf{u})$ associated to the spatial configuration of n data
4:    **for** each training image $m = 1, ..., M$ **do**
5:        Associate the data event to the training image m, $D_{m,n}(\mathbf{u}) = D_n(\mathbf{u})$
6:        Compute the conditional probability $P(S(\mathbf{u}) = s_k | D_{i,n})$
7:    **end for**
8:    Integrate the conditional probabilities by the linear combination scheme $P(S(\mathbf{u}) = s_k | D_{1,n}, ..., D_{m,n}) = \sum_{i=1}^{n} w_i P(S(\mathbf{u}) = s_k | D_{i,n})$
9:    Draw the categorical variable from the posterior probability $P(S(\mathbf{u}) = s_k | D_{1,n}, ..., D_{m,n})$
10:    Add the simulated value s_k at location $\mathbf{u}$ and move to the next node
11: **end for**
12: **Return** or repeat the process for anew simulation

The previous algorithm was implemented by the program *snesimmti* written in ANSI standard Fortran 90. The *snesim* algorithm developed by Strebelle (2002) is used as the platform for the linear integration of the conditional probabilities.

Many parameters related to *snesim* apply to *snesimmti* including global and vertical target proportions for the servosystem correction, training image parameterization, template set up, the minimum number of data events, definition of multigrids and conditioning data searching parameters.

4.4.4   Application: A 2-D Example

Three examples based on training images of different nature are used to illustrate the proposed algorithm.
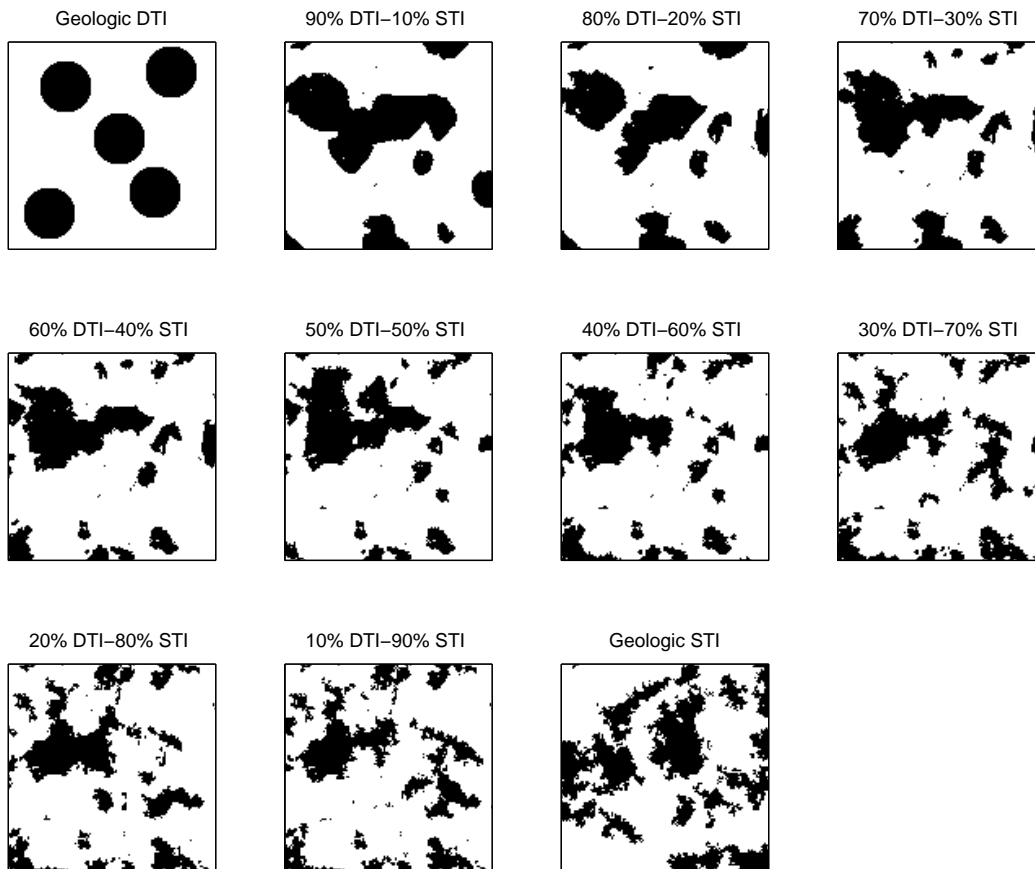
Figure 4.2: Realizations generated by the integration of multiple training images using the LOP scheme. The example consists of an artificial exercise using spheres. The first and last images correspond to the training images. The nine simulations in between correspond to configurations of weigths in intervals of 10%.

The size and configuration of the data event $D_{m,n}$ must be considered in the examples. A large data event probably will not find enough replicates to estimate

the conditional probability. Conversely, a too specific and smaller data event will not capture larger scale features between the data points. The size and configuration of the data event should be chosen to capture the main features, but at the same time, to permit stable calculations.
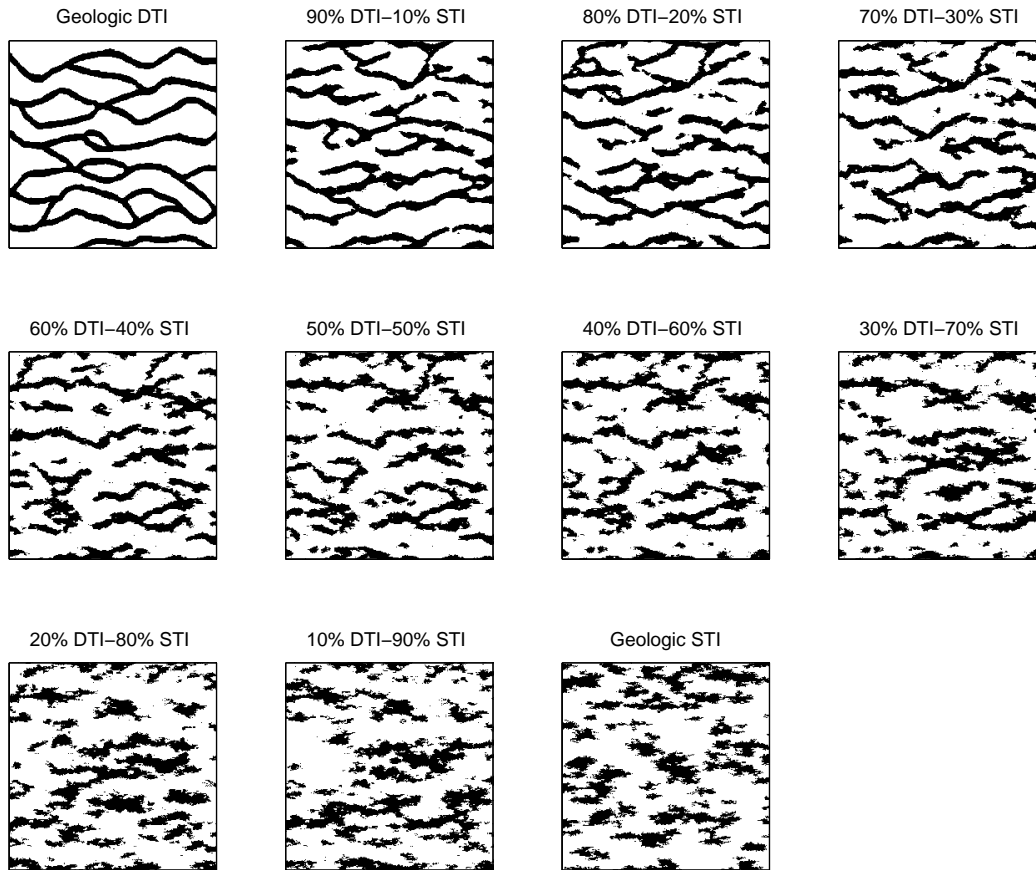
Figure 4.3: Realizations generated by the integration of multiple training images using the LOP scheme. The example is based on training images of a fluvial reservoir. The first and last images correspond to the training images. The nine simulations in between correspond to configurations of weigths in intervals of 10%.

Examples are shown in Figures 4.2, 4.3 and 4.4. Random samples were extracted from training images and used as conditioning data. Nine scenarios in intervals of 10% were generated. To make the comparison among simulation easier, the same random path was used.

The convex nature of the linear data integration scheme is reflected in the resulting simulations. For the three cases, the models transit from one well-structured
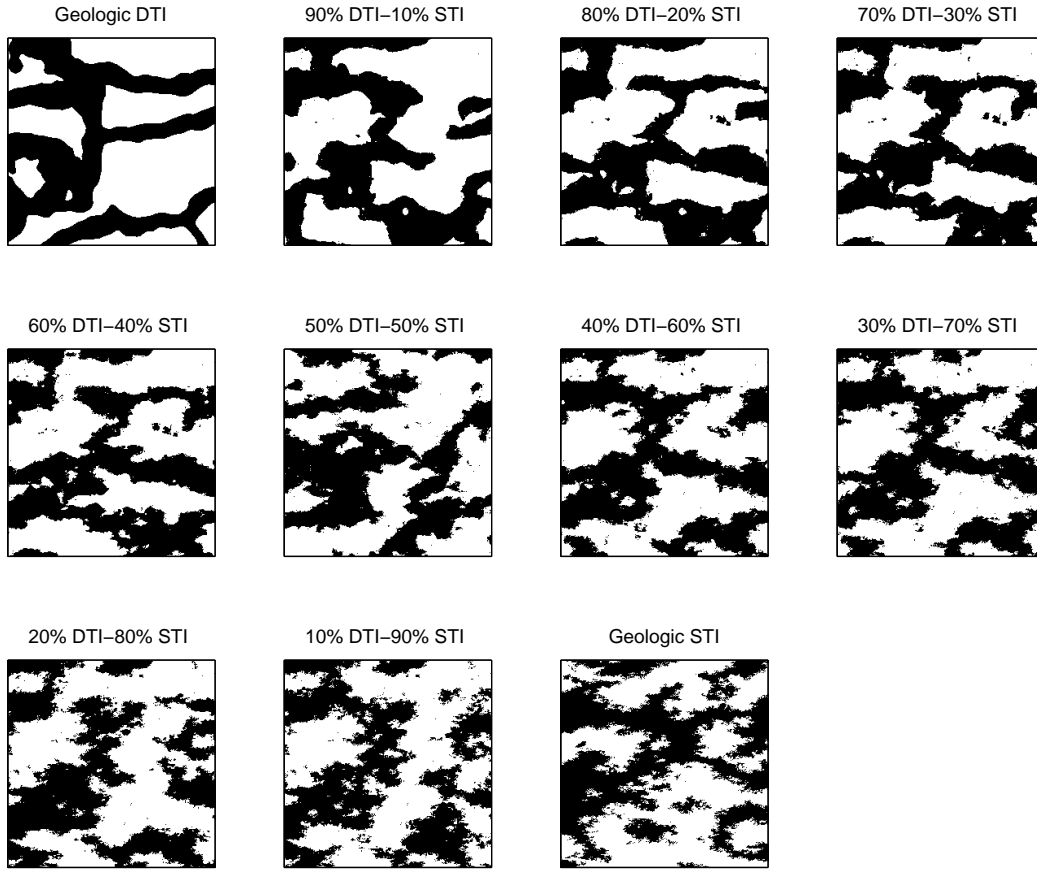
Figure 4.4: Realizations generated by the integration of multiple training images using the LOP scheme. The example is based on a reservoir characterized by turbidite lobes. The first and last images correspond to the training images. The nine simulations in between correspond to configurations of weigths in intervals of 10%.

and continuous realization to another where the patterns are increasingly randomly arranged. The resultant models respond to essential features from both training images according the level of mixing. In all the cases, the models are in between the two training images.

A reasonable amount of continuity and randomness are reproduced in the models. The realizations preserve the geologic structures from both training images by maintaining large scale structures of the deterministic, but incorporating the randomness of SIS at different rates as required. The generation of geologic features in the realization is not exclusively reserved to patterns contained in the training images. By combining the conditional probabilities, the linear opinion pool updates the

information modifying the underlying random function, generating novel geological features not contained in the images. These unique features result in realizations with a controllable degree of randomness in the spatial arrangements of geologic domains.

4.4.5   Issues with Multiplicative Aggregation Methods

The second branch of aggregation probability methods are based on the multiplicative integration of individual conditional probabilities. Among these methods, there exist approaches to weight the sources of information. Journel (2002) proposed the permanence of ratio model based on the conditional independence of the secondary information. The data sources are not entirely conditionally independent, thus, the Tau model Journel (2002) and other weighted combination models (Hong, 2010; Polyakova and Journel, 2007) account for this redundancy and improve the predictive capabilities of the models.

The weighted combination models provide convex and non-convex combined probabilities depending on the specific probabilities and weights. It offers a robust framework for dealing with probabilities from unique data sources and varying levels of reliability and information. As discussed earlier, the multiple training images correspond to two valid interpretations of the same geological phenomenon subject to the same level of information. Even though the Tau model was proposed to deal with redundant sources of information, the redundancy of the training image may be treated better by the linear opinion pool. Here, the convex property of the aggregation method becomes an important requirement when combining probabilities from different training images. The proposed methodology aims at generating realizations in between both training images.

The non-convex property reduces the uncertainty of the primary attributes by increasing its certainty of occurrence. In other words, the combined probabilities tend to be closer to 0 or 1. This property is not desirable when the sources of information are particularly redundant as the case of the proposed data-driven training images. To illustrate this point, let us suppose that no weighting scheme is adopted,
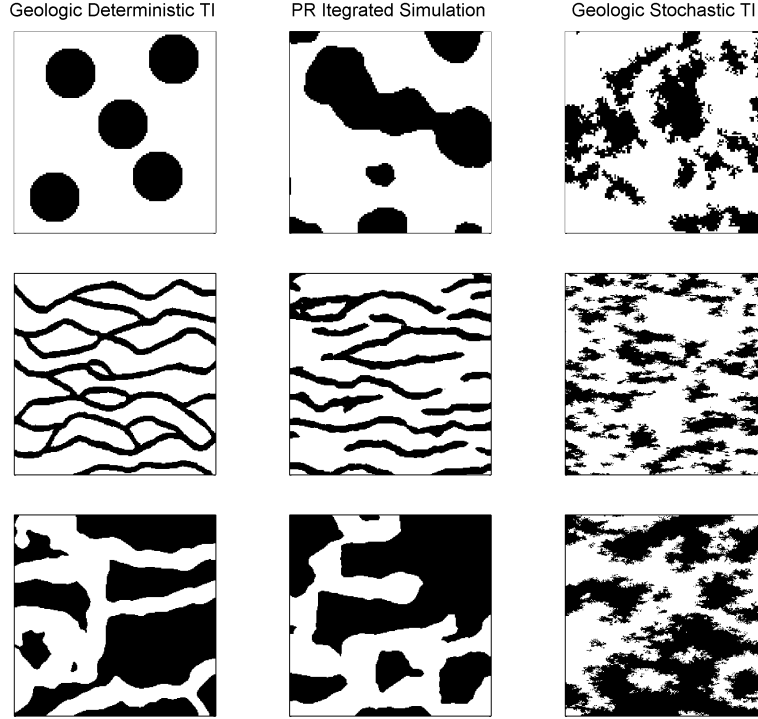
Figure 4.5: Three realizations resulting from the integration of two training images based on the multiplicative approach of pernanence of ratio. Same examples of Figure 4.2, 4.3 and 4.4 are used. The first and last column represent the training images, in the middle the realizations.

and the weights of the Tau model are unity (equally weighted sources). This configuration assumes no redundancy between the training images, turning into the well-known permanence of ratio model. Given a data event D_n, both training images provide unbiased predictions of occurrence of the primary attribute. As both images are highly correlated, it is likely that the combined probabilities over or under estimate the probability of the attribute. Besides, the nature of the features of the training images comes into play an important role in the behavior of the combined probabilities. The training image derived from the SIS will turn into a less certain probability, given the high entropy nature of the image, than the smooth one. Therefore, one would expect that correct (true) probability would be far from the probability estimated from the SIS, but closer to the probability estimated from the smooth image. Thereby, the combined probability should be placed outside the in-

terval of conditional secondary probabilities, but closer to the smooth one. Thus, the non-convex character accentuates even more the tendency to predict the posterior probabilities smaller and greater than the probability computed from the smooth image, pushing the probabilities even closer to the extremes. Figure 4.5 shows an example of applying the permanence of ratios approach when combining training images. Three distinct training image are used, both represented by the first and last column. The column in the middle shows a realization with the same smooth character than the training images on the left, which is not the expected result in the current context. The multiplicative integration of the conditional probabilities accentuated the probability computed from the smooth training image.

The setup of the weights for the tau model is unclear and the assumptions underlying the method are not satisfied. The linear opinion pool is the straightforward and correct technique to combine probabilities in our context of balance the geologic variability in the models, but certainly not in all circumstances.

4.5   Calibration of the Training Images

A data-based approach to calibrate the weights for each training image is proposed. As mentioned before, the assignment of weights could be imposed directly by the expertise of the geomodeler and the resulting realizations validated according to the geological variability expected by the expert. However, one would wish to select the correct mixture between the training images using an objective and practical approach based on available conditioning data. The goal is to correctly reproduce the geological variability, that is, the spatial disorder of the geologic attribute. Thus, a measure of entropy appears as a natural tool for calibrating the mixture model.

4.5.1   Entropy as a Measure of Spatial Disorder

Entropy was introduced in geostatistics by Christakos (1990) as a measure of uncertainty of probability density functions. In the case of discrete variable with n

different probabilities, it is defined as:

$$H = -\sum_{k=1}^{n} p_k \ln(p_k),$$ (4.4)

with p_k being the probability of occurrence of each of the n possible outcomes of the random function and $\sum_{k=1}^{n} p_k = 1$. In the case of continues variables, the p_k is replaced by the probability density function and the summation by the integral. The entropy is a summary statistic similar to the variance, but aimed at categorical variables that have no intrinsic numerical value.
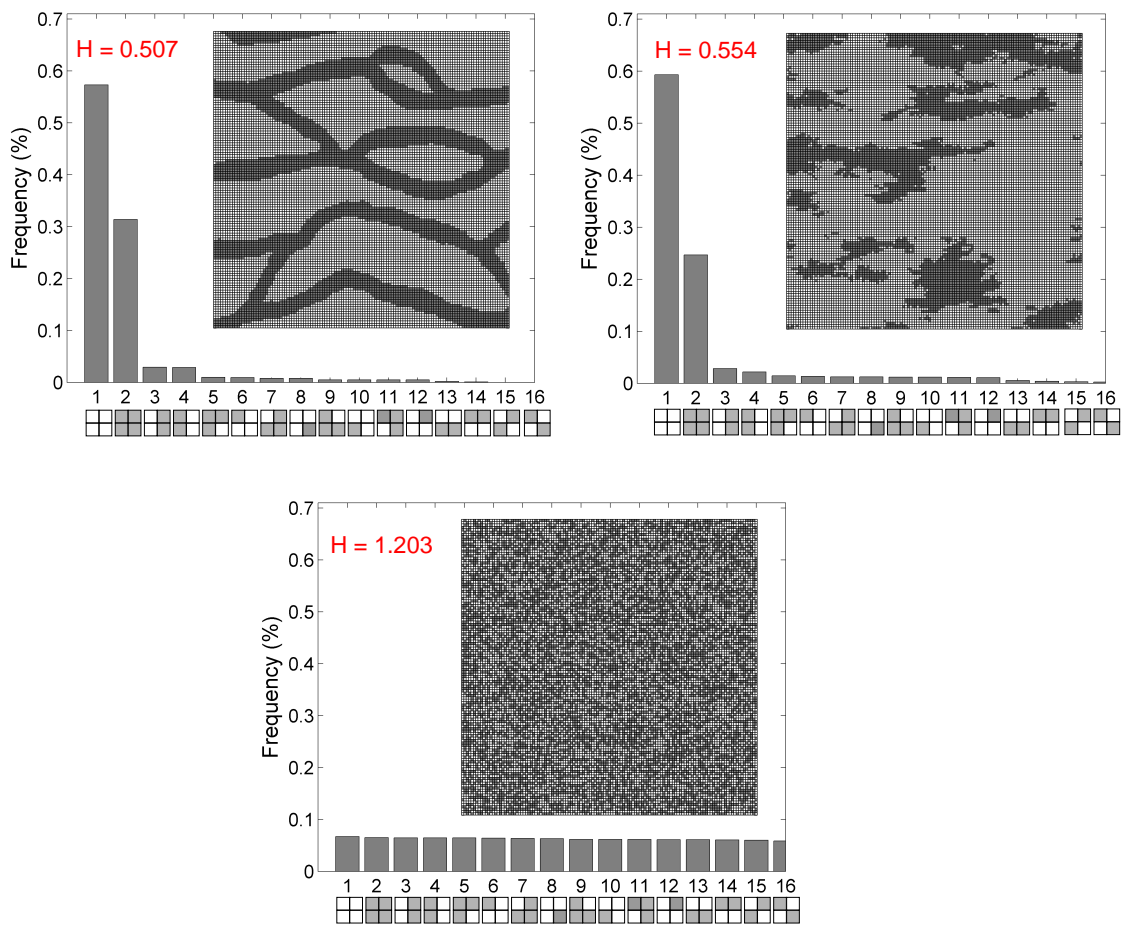
Figure 4.6: Multiple point density functions for three training images. In red, the computed entropy associated with each image. At the top on the left, a well-shaped fluvial image; on its right, a stochastic image generate by SIS. The case of maximum entropy is given by the image at the bottom.

The concept of entropy in a spatial disorder framework was first introduced by Journel and Deutsch (1993) for assessing the resulting response distributions assuming different random functions given the same level of known spatial statistics. In this work, the entropy is used for assessing the spatial disorder of simulated realizations. The spatial entropy allows evaluating the degree of disorder of a training image by using templates of distinct size and configurations.

To illustrate the entropy as a measure of spatial disorder, the example in Figure 4.6 is presented. At the top, two training images with significant differences in terms of continuity. The image on the left shows well-defined limits of channel-like structures; the case of the image on the right evidences a more random behavior in the contacts. The entropy depends on the size of the template and the number of categories in the image; the number of combinations that one may generate from a template of size K and N categories is given by K^N. Thus, for a 2 by 2 template $2^4 = 16$ configurations could be generated. For the case of the continuous channels image, only 14 out of 16 configurations were found; the patterns labeled as 15 and 16 do not appear in the image, and the entropy reaches 0.507. As may be expected, the second training image shows an increased level of spatial disorder reflected in a greater entropy of 0.554. For comparison, the extreme case of a spatially non-correlated image is presented at the bottom of the figure. The image corresponds to the random case where the histogram of the multiple point configurations follows a uniform distribution. This case corresponds to the case of maximum entropy or complete randomness.

4.5.2 Linking Drillholes Entropy to two/three-dimensional Training Image Spatial Entropy

The spatial entropy has been devised as a structural tool for calibrating realizations; however, a major issue arises when dealing with two/three-dimensional training images, and the limited one-dimensional information of the drilling campaign. The spatial entropy comparison and calibration is not straightforward. The assessment of spatial entropy requires the use of multipoint templates that must be computed
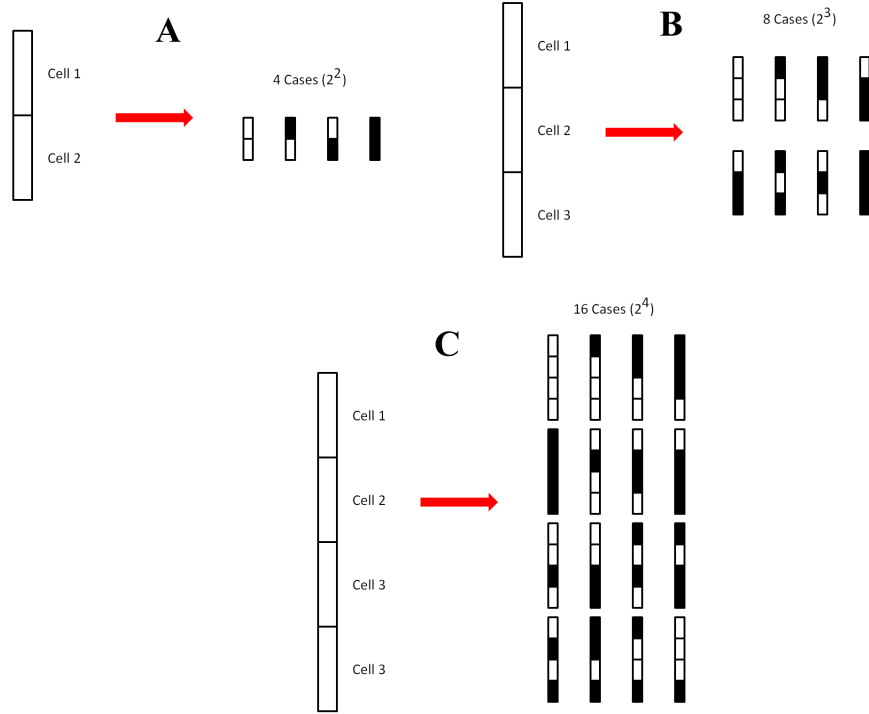
Figure 4.7: All possible multiple point configurations based on three examples of one-dimensional templates when considering two indicator variable for geologic categories. The examples are based on templates of two, three and four locations.

from both the drillholes and the training images; therefore, one-dimensional templates must be considered. Figure 4.7 shows a template for two, three and four locations given two categories. To determine the entropy of the set of drillholes, each drillhole is treated separately based on one-dimensional template. Figure 4.8 shows an example of all templates configuration found along a set of 10 drillholes; then, the entropies of each drillhole are averaged to obtain the expected entropy from the set of drillholes according to:

$$H_{drillholes} = -\frac{1}{n_d} \sum_{i=1}^{n_d} \sum_{j=1}^{N^l} p_{i,j} \ln(p_{i,j}), \qquad (4.5)$$

where n_d is the number of drillholes and N^l the number of possible template configurations. A comparison of $H_{drillholes}$ with the entropy of the training images

is needed. To obtain an adequate entropy comparison from both training images and drillholes the same configuration of template is required. Given the spatial restrictions this comparison must be performed in one-dimensional template dimensional basis.

Number of Replications of Each Template per Drillhole

$H_1 = 0.74$ $H_2 = 1.01$ $H_3 = 0.87$ $H_4 = 0.83$ $H_5 = 1.02$ $H_6 = 0.91$ $H_7 = 0.84$ $H_8 = 0.84$ $H_9 = 0.78$ $H_{10} = 0.93$

$H_{global} = 0.88$
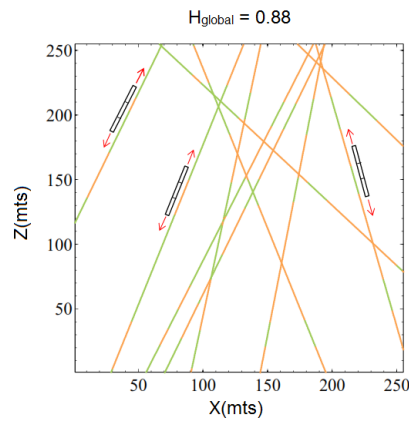
Figure 4.8: At the bottom, a cross section with a spatial configuration for 10 drillholes. At the top, the number of replication for each configuration found along each drillhole. The global entropy is equal to 0.88.

95

Even though the one-dimensional spatial entropy does not completely define the spatial disorder compared to a two-dimensional of three-dimensional template, it represents, at least partially, the spatial order of geologic structures.
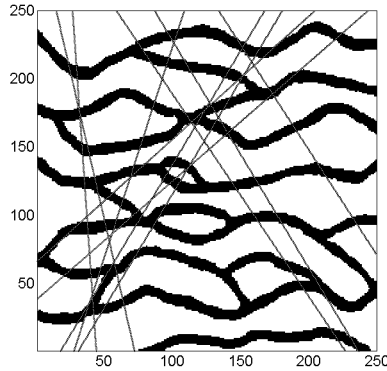
Figure 4.9: Fluvial training image and an artificial configuration of a drilling campaign.

The direction that the entropy is computed is relevant. Figure 4.9 presents a well-known fluvial training image to be used to illustrate this last point. A total of 10 drillholes were sampled from the training image in different directions, and their one-dimensional entropy computed and plotted in Figure 4.10. Changes in the entropy value responds to variations in the direction of the drillholes intersecting specific patterns of the training image. To verify the dependency of entropy according to the dip, the experiment was repeated, but this time sampling 100 drillholes. Figure 4.11 compares the entropy of each drillhole with respect to its dip, as close to vertical provides larger entropy values. It is a reflection of the structure of channels in the training image. If a drillhole is sampled along the vertical, it has to go through more channels that the case when the dip is close to 45 or 135 degrees where several channels split into two channels for a short distance before merging anew into one channel. This demonstrates that the one-dimensional entropy is able to describe structural-oriented patterns from two/three-dimensional training image, and its accuracy as an element to calibrate training images.

The orientation of the drillholes must be accounted for when computing the spatial entropy. The problem arises how to compute the spatial entropy from two/three-dimensional training images subject to a set of drillholes. To overcome this issue

96
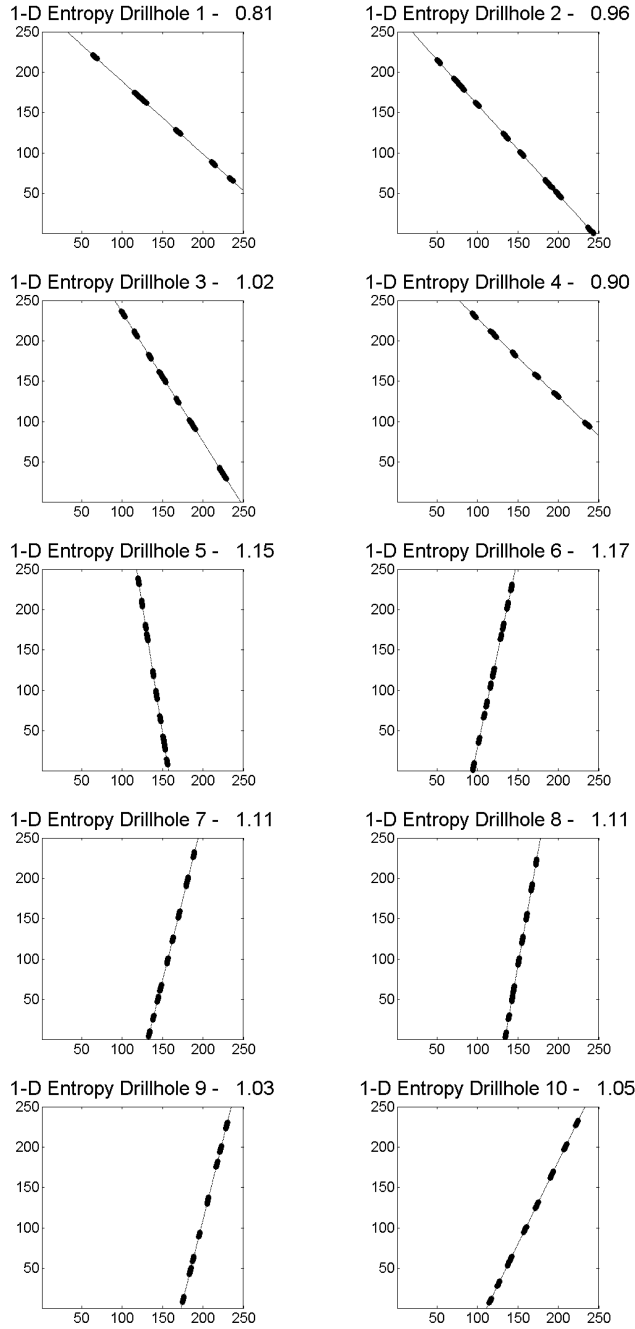
Figure 4.10: One-dimensional entropy of each drillhole sampled from training image in Figure 4.9.

a one-dimensional sampling procedure is proposed upon the training image. The orientation of each drillhole sampling is applied at several locations on the training

image. The average of the results provides a measure of entropy from the training image that is comparable to the entropy of the drillholes.
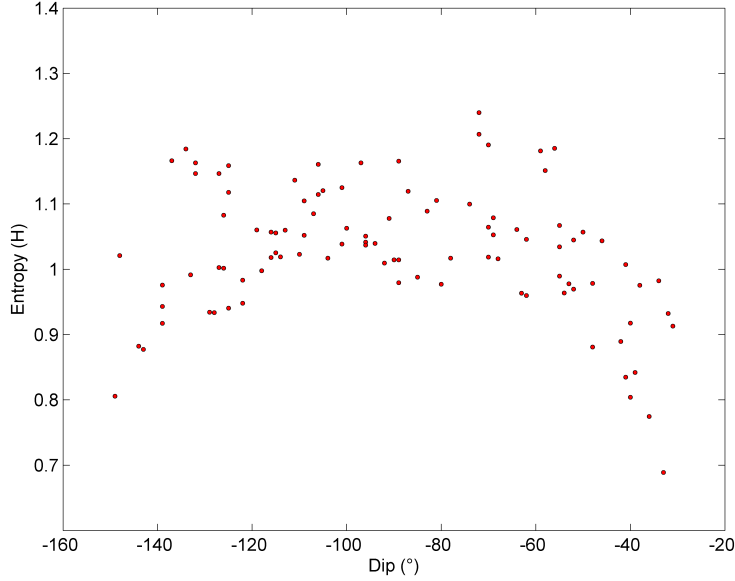
Figure 4.11: One-dimensional entropy subjected to dip of 100 sampled drillholes from training image in Figure 4.9.

The strategy looks for a regular sampling over the training image based on the orientation of each drillhole. For instance, consider a set of 10 drillholes as in Figure 4.12 at the top. In the middle, the figure shows the regular sampling procedure with different orientations of drillholes 1, 2 and 4 on a two-dimensional training image. At the bottom, the resulting sampled replicates with their assigned categories. Then, the entropy of the 1-D replicates sampled from the training image are shown at the bottom. The entropy of the training image corresponding to the entropy's average of each single-orientated drillholes:

$$H_{TraininImage} = -\frac{1}{n_d} \sum_{\alpha=1}^{n_d} (\frac{1}{n_{sd_\alpha}} \sum_{i=1}^{n_{sd_\alpha}} \sum_{j=1}^{n_{Nl}} p_{i,j} \ln(p_{i,j})), \qquad (4.6)$$

Again, n_d is the number of drillholes but n_{sd_α} the number of orientated replicates within the training image for each n_d.

Consider an example of calibrating the spatial disorder from the drillholes to the
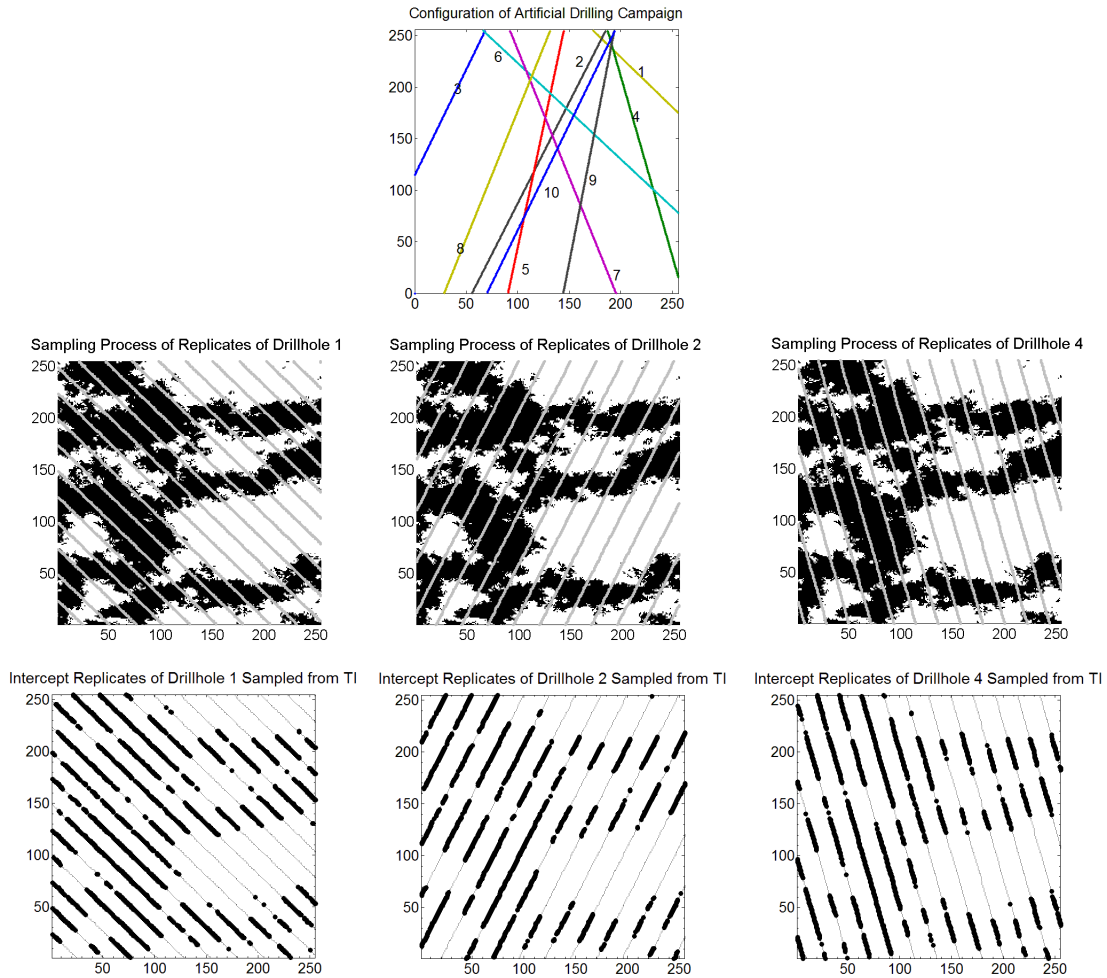
Figure 4.12: Sampling process of drillholes 1,2 and 4 over the training image. The entropy of the training image is calculated over the replicates shown at the bottom.

training image weights. Using the training images from the example of Figure 4.4, a total of 101 realizations were simulated. Each realization corresponds to a change in the weights by 1%. Thus, it starts with weights of 100% and 0% up to the opposite scenario of 0% and 100%. The same two-dimensional configuration of 10 drillholes is used to sampling each training image and utilizing them as a true reference. Later, the entropy for each drillhole is computed and plotted to compare with the reference entropy, see Figure 4.13 . The entropy of the drillholes is accurately corresponded by sampled entropy of the training images. Thus, the entropy of the drillholes can be used to select suitable weights for the training images, matching the same entropy

99

as the drillholes.
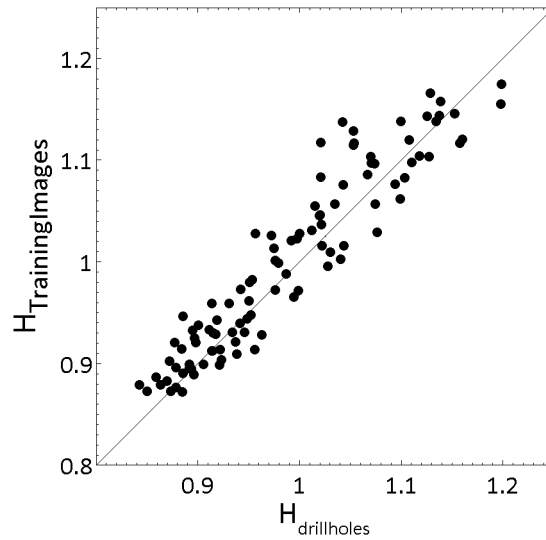
Figure 4.13: Scatter plot between the entropy of 100 training images, and the entropy computed from the replicates of drillholes based on the proposed sampling process.

A program called *caliti* written in ANSI standard Fortran 90 consistent with GSLIB was developed to perform the calibration process. The program reads the set of combined realizations, samples the drillholes, computes the entropy, and then calibrates them.

4.6 Analysis of Reproduction of Geological Features in the Models

When using MPS simulation methods, the primary concern relates to the reliable reproduction of input statistics and geological features of the training image in the realizations. However, when dealing with the combination of multiple training images the challenge is no longer to evaluate whether these models reproduce the input statistics, but understanding the mixing process. In what follows, a series of quantitative methods are used to evaluate which features are preserved along with how the multiple point statistics change as the training images are being combined. These techniques assess the reproduction of univariate statistics and the distribution

of multiple point statistics through the use of multiple density functions.

4.6.1 Quantitative Measures to Assess Geologic Features Reproduction

Checking the reproduction of high order statistics from training images is quite difficult due to the high dimensionality of the configurations that may be generated. Boisvert et al. (2010) suggested a set of metrics to assess the reproduction of high order statistics including elements as the first two order statistics, trends, multiscale histogram, multiple point density functions, and missing bins in the multiple point density function. The goal was to establish a minimum criteria of approval with respect to the reproduction of patterns retrieved from the input training image. In the same manner, three statistics of this set of measures are proposed for evaluating the mixing of geologic features from multiple training images, these are: the use of multiple point density function, a measure of similarity and the entropy.

Multiple Point Density Function

Originally proposed by Boisvert (2007), the multiple point density functions correspond to the distribution of the set of multiple points configurations encountered in an image. The distribution depends on the size of the template and the number of categories in the image; recall that the number of combinations is given by K^N, where K is the number of categories, and N is the locations used in the template. An illustrative example is showed in Figure 4.14, the training image represents a binary (shale and sand) fluvial reservoir. A template of 4 locations (2x2) is considered. The template is scanned over the training image counting for the appearance of each configuration; then, a histogram is constructed.

Various distributions may be constructed using different templates. However, the dimensionality of the template may be a serious obstacle for summarizing the statistics and its visualization. Typically, when dealing with complex geological environments, the size of the template has to be large enough to characterize the complexity of the categorical arrangements and the amount of multiple point con-
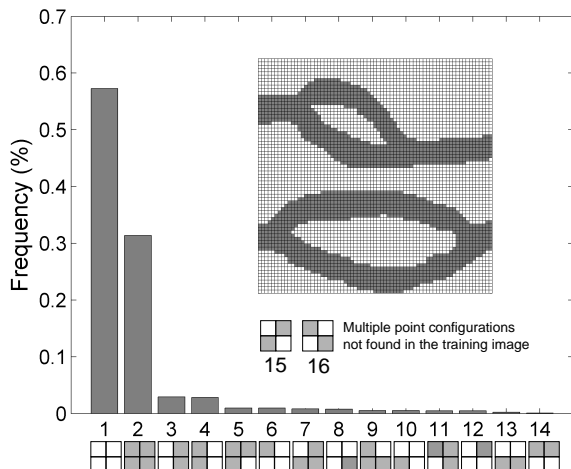
Figure 4.14: Multiple point density function of the two-category image above. The histogram correspond to the frequency of appearance of 14 multiple point configurations using a four-location template.

figurations may be huge; a histogram that involves millions of classes could not be understood. The advantage of multiple point density functions lies in the capacity for checking if the patterns in the realizations have a similar frequency to those in the input training image.

Similarity

The second measure relates to the concept of similarity. This statistics compares if multiple point configurations are found in both the training image and the realizations without considering its frequency of appearance. A similarity measure was proposed by Manchuk and Deutsch (2012), and is presented in Eq 4.7. $N_{sim \in TI}$ corresponds to the number of multiple point configurations in the simulation that are also present in the training image (TI), and N_{TI} is the total amount of multiple point configurations in the training image.

$$S_{sim,TI} = \frac{N_{sim \in TI}}{N_{TI}} \tag{4.7}$$

This measure lies between 0 and 1, as closer to 1 means that more multiple point configurations are shared between the realization and the training image. The
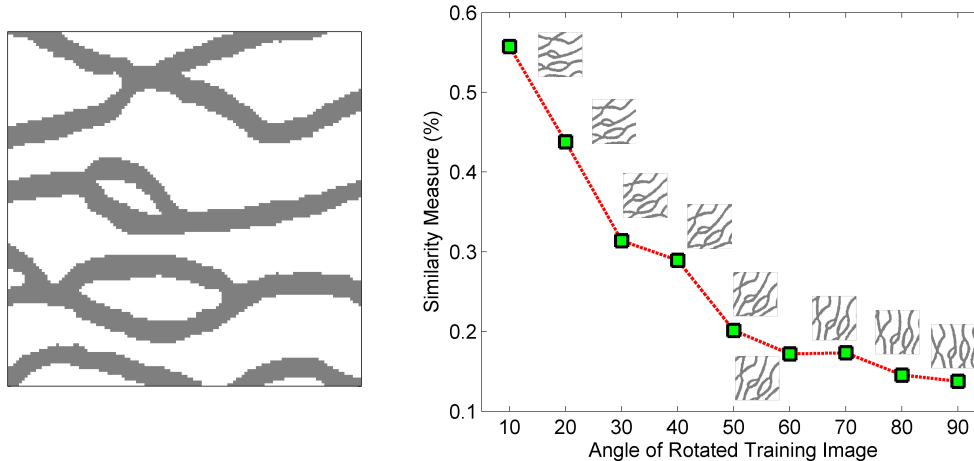
Figure 4.15: The red curve shows the similarity measure between the image on the left and rotated representations of itself; increasing angles from 10 to 90 degrees on the right were used. As large the rotation less similar the images.

similarity also depends on the template used. Special care must be taken when the comparison is realized over images that share a high proportion of certain category; in this cases, the similarity tends to increase despite that the patterns related to categories of low proportions do not fit.

An example of the similarity measure is shown in Figure 4.15. The exercise consists of comparing the training image on the left with rotated images from itself; nine angles ranging from $10°$ to $90°$ were used. On the right, the similarity measure curve shows how the images loss similarity as the angle of rotation increases. The number of shared patterns decreases up to reach around 15% of the total number of multiple point configurations in the training image.

4.6.2 Expected Geological Features in the Combined Models

A convex data integration scheme is a desired quality when combining training images of the same geological phenomenon for calibrating geologic variability. As depicted by Figures 4.2, 4.3 and 4.4 the spatial variability of the geologic arrangements may be balanced. To understand the generation of geological features and the changes in the multivariate distribution, the previous measures are used to compare the resultant geological features in the realizations to those originally contained in

103

the training images.
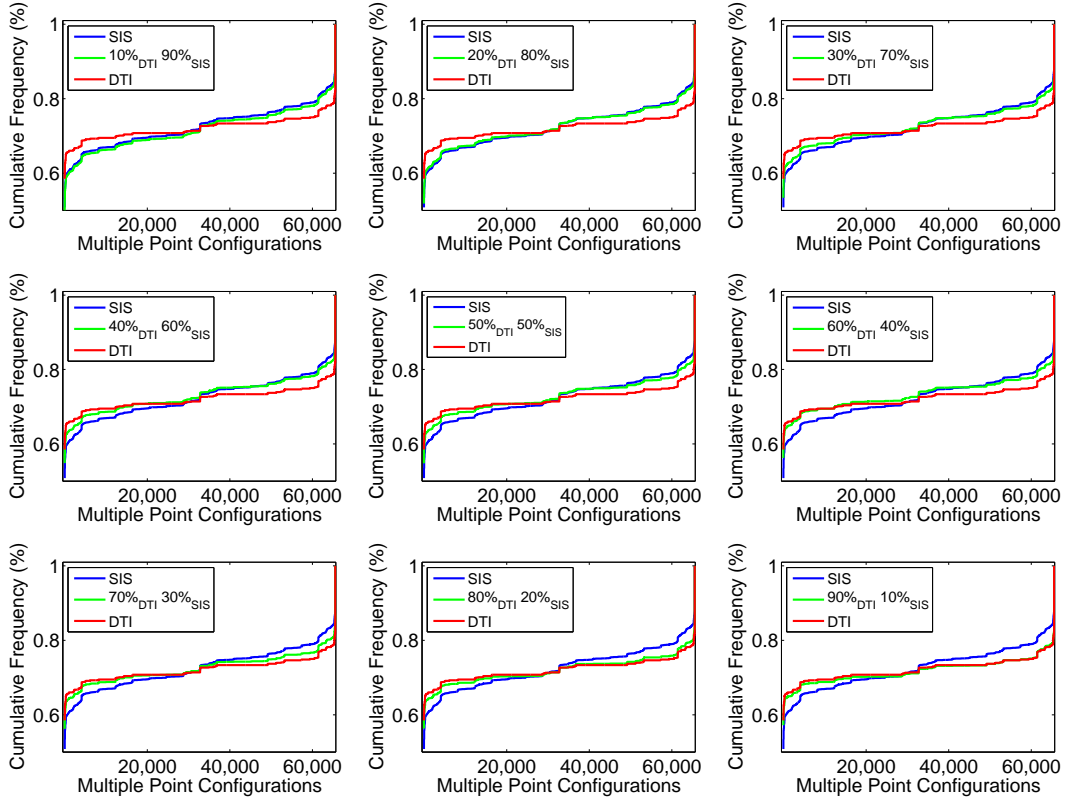
Figure 4.16: Comparison of the cumulative multiple point density function curves between the deterministic, stochastic and combined realizations. Nine examples are shown using weights in intervals of 10%.

The fluvial example in Figure 4.3 is reconsidered for the analysis. The same nine scenarios when combining the images in intervals of 10% are utilized.

Multiple Point Density Function: Figure 4.16 compares the cumulative distribution of multiple point configurations of the deterministic (DTI) and stochastic training images (STI). A template of 4x4 was used; this implies a set of $2^{16} = 65,536$ possible configurations considering two categories (sand and shale). In the abscissa axis the multiple point configurations numbered from 1 to 65,536; the labeling does not consider any special ordering. On the ordinate axis, the cumulative frequency curves. The blue curve represents the distribution of the multiple point configurations contained in the deterministic training image, in red, the distribution of the stochastic one, and in green the distribution of the combined simulations. The nine
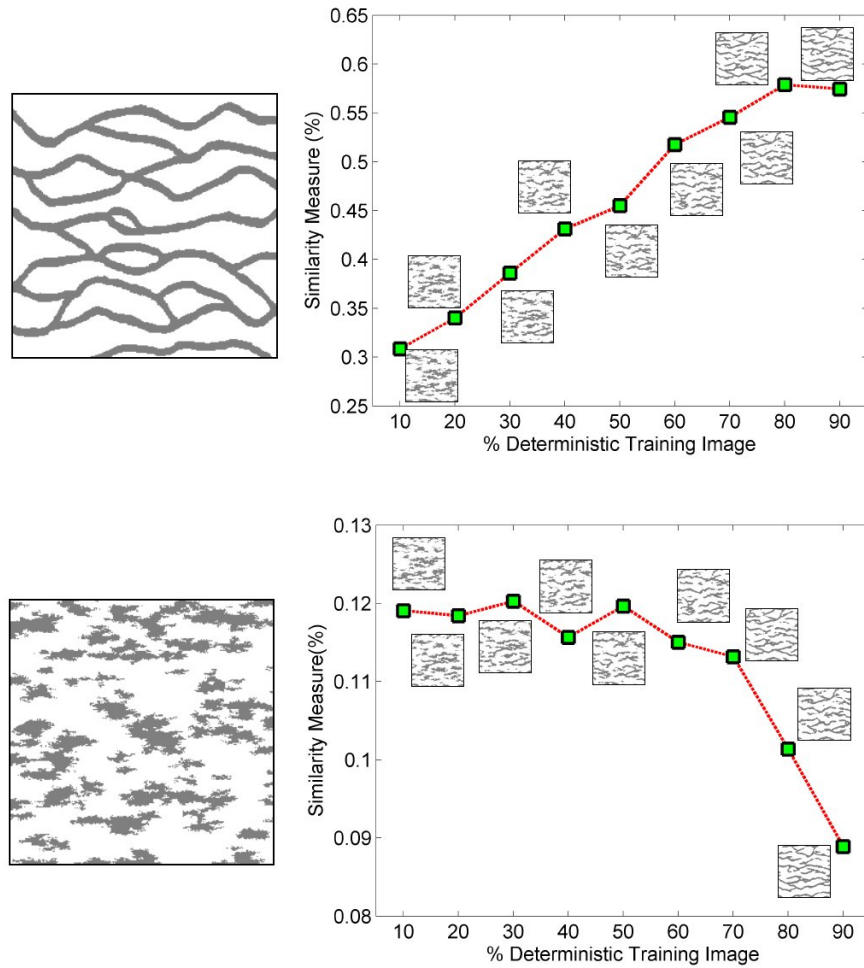
Figure 4.17: At the top, the similarity measure based on the deterministic training image on the left. At the bottom, a second similarity measure, but with respect to the stochastic training image on the left. The nine small images above the curves correspond to realizations based on weights in intervals of 10%. The weight of the deterministic image increases from left to right.

figures correspond to configurations of weights in intervals of 10%. Closer curves imply greater similarity in the combined realizations and the training images. Note how the green cumulative frequency curve transits from the STI to the DTI curves as the images are combined. The convex property of the linear combination makes that the resulting cumulative multiple point density functions of the realizations fall in between those corresponding to the training images. The realizations are balanced within the range of variability established by the training images.

105

Similarity: Figure 4.17 illustrates how the realizations become more similar to the training images when changing the configuration of their weights. At the top on the right, the similarity measure based on the training image on its left. The similarity is computed for each configuration where the weight of the deterministic image increases from 10% to 90%. The small images represent the resulting combined realizations. At the bottom, the same similarity curve, but this time, based on the stochastic training image on its left. One interesting aspect of the results is the confirmation that the linear combination of multiple point statistics produces geological features in the realizations that change linearly in terms of the similarity measure. The former may be expected given the nature of LOP training image integration; it confirms what is visually observed in the realizations.
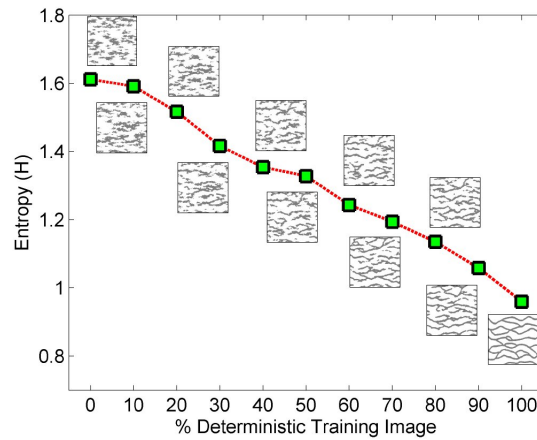
Figure 4.18: The measure of entropy. The first and last values correspond to the entropy of the deterministic and stochastic training images. In the middle, the nine entropy values for the nine realizations based on configurations of weights in intervals of 10%.The weight of the deterministic image increases from left to right.

Entropy: like the similarity measure, the spatial entropy of the simulations (Eq 4.4) follows a linear behavior, see Figure 4.18. That is, the geologic variability of the fluvial simulated reservoir changes linearly as the multiple point statistics are being combined. In terms of spatial disorder, the entire spectrum of variability may be visited just by setting up the correct set of weights. An essential part of the proposed methodology consists of generating training images that represent extreme scenarios of geologic variability; then, the linear combination guarantees generating

106

realizations with the correct amount of geologic variability.
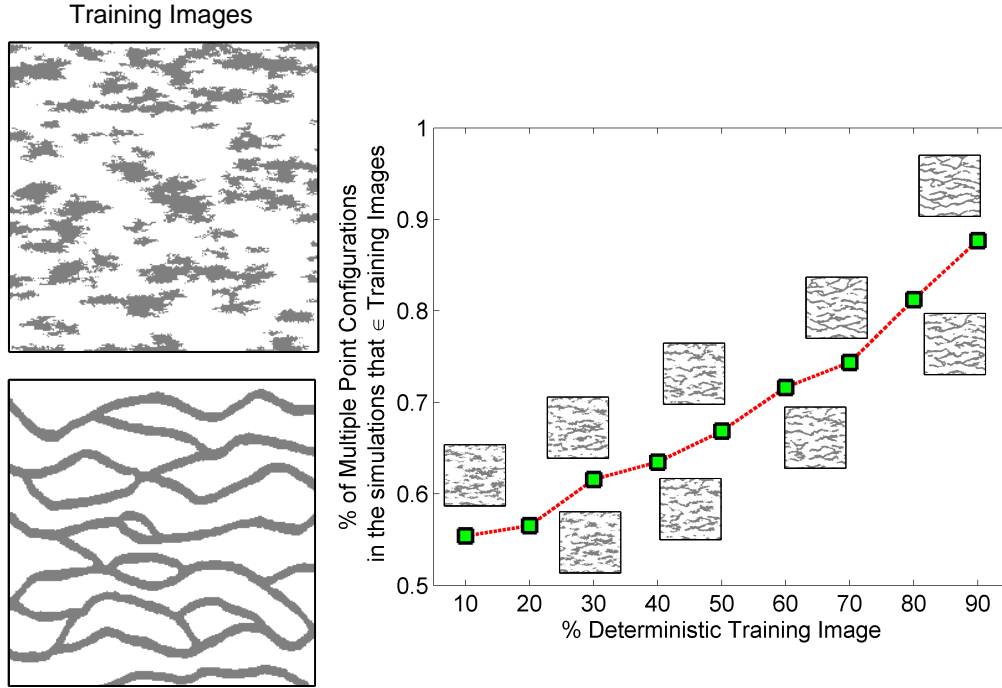
Figure 4.19: On the left the deterministic and stochastic training images. On the right, the curve that shows the percentage of multiple point configurations found in these training images that are also present in the nine realizations. This percentage increases with the weight associated to the deterministic training images.

The previous analysis relates to general aspects of the expected geological features in the realizations. However to completely understand the process of combining multiple point statistics, it is also necessary quantify the nature of the resulting patterns, that is, how many are inherited from the training image and how many correspond to novel features generated by the merging process. For doing this, the same previous fluvial example is utilized. The analysis consists of comparing each simulated multiple point configuration to the configurations contained in the training images, and quantify the proportion of them that correspond to novel features. This measure differs from the similarity in that the number of multiple point configurations is computed using both training images. Figure 4.19 on the right shows the proportion of patterns in the realizations that are found either in the deterministic or stochastic training images. It may be observed that as the weight of the

deterministic training image increases, the proportion of inherited multiple point configurations also increases. In other words, when just a 10% of the deterministic image is used almost the half of the multiple point configurations in the realizations correspond to novel configurations not contained in the training images; the opposite occurs when using 90%, in this case, the 10% of the configurations are new.

The generation of novel multiple point configurations that are not contained in the training images is increased when using training images with a considerable level spatial disorder. A larger set of multiple point configurations leads to novel multiple point configurations in the simulated realizations; it is always feasible to find configurations that match, even though, its frequency is limited to a few number of appearances. For the deterministic training image, there are fewer new patterns due to the restricted set of multiple point configurations available when simulating.

A lack of similarity in the input training images also leads to more novel patterns. Consider the example of Figure 4.20. Here, the fluvial training image is rotated by angles of 15°, 30°, 45°, 60°, 75° and 90°. The exercise consists of combining the original and rotated training images, represented by the first and last column in the figure. Intuitively, the lower the rotation, the more similar the training images. The three columns in the middle represent configurations of weights at 75% - 25%, 50% - 50% and 25% - 75%. Visually, it is possible to observe that increasing the rotation of the image affects the reproduction of features in the realizations making them less similar to the training images. The extreme case corresponds to the last row where the lack of continuity of the channels is evident, especially for the 50% - 50% case. In fact, the channels start losing continuity when the second training image presents a rotation of 60°. The cases for the images rotated at 15°, 30° and 45° are interesting since the simulated channels maintain the continuity with subtle alterations. For all the scenarios, the angle of the simulated channels are mildly elevated, for instance, in the case of 75% - 25%, the horizontal direction of the channels increases when being combined with a more rotated training image. Note that due to the LOP integration, the angle of the simulated channels is always in between the angle shown by the training images.

Simulations combined by LOP

$75\%_{FLTI} - 25\%_{RFLTI}$ $50\%_{FLTI} - 50\%_{RFLTI}$ $25\%_{FLTI} - 75\%_{RFLTI}$
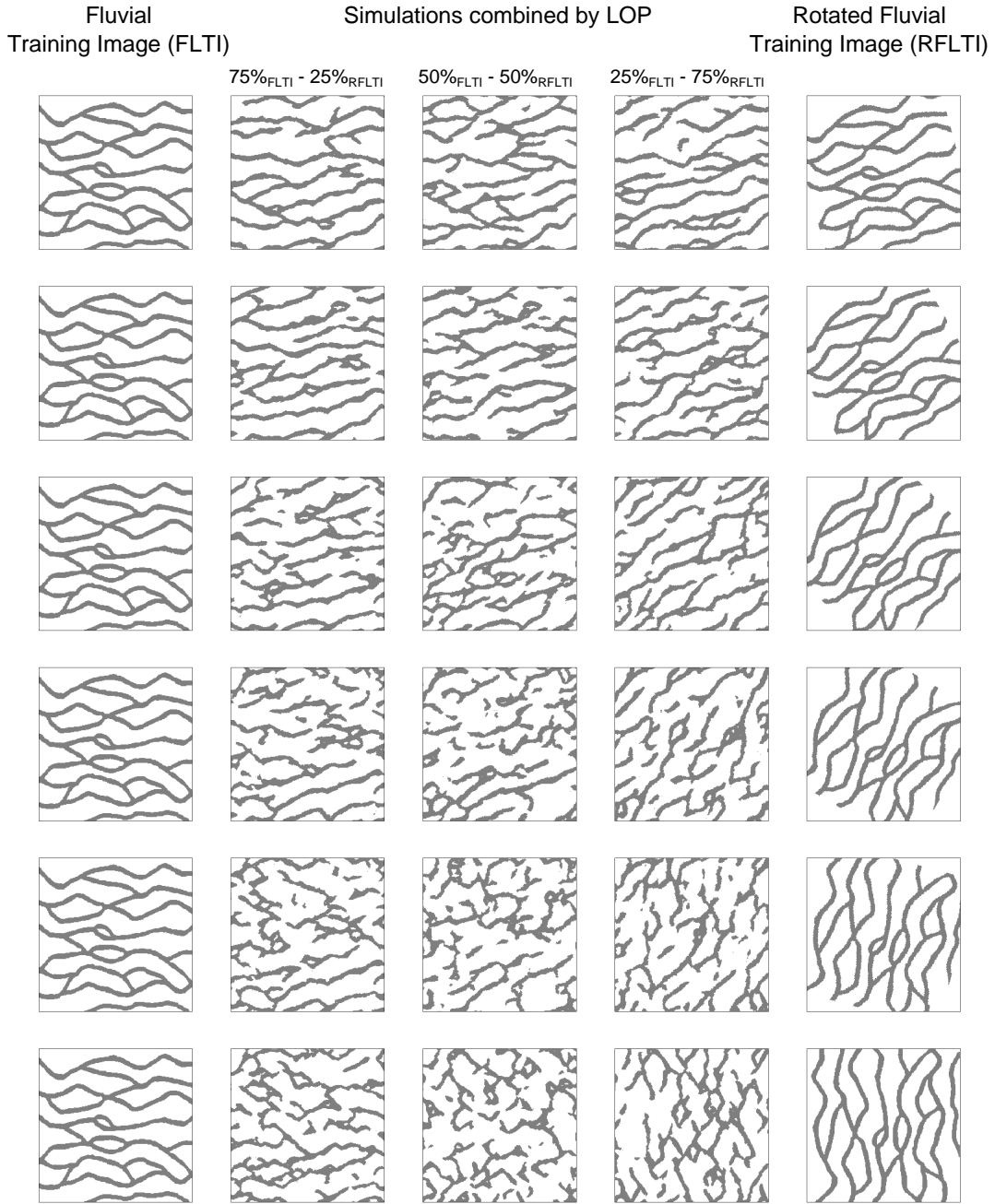
Figure 4.20: An illustration that shows how the level of similarity between training images implies realizations with similar characteristics to the training images features. On the first column, the fluvial training image; on the last column, a rotated version. On each row, the training image is rotated in $15°$. The three column in the middle correspond to simulations based on the configurations of weights of 75% FLTI - 25% RFLTI, 50% FLTI - 50% RFLTI and 25% FLTI - 75% RFLTI.
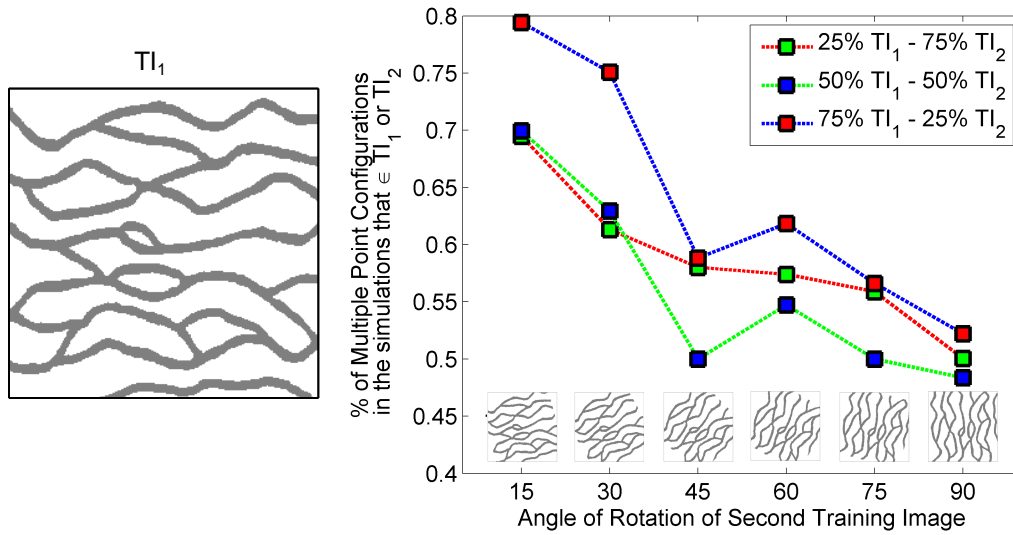
Figure 4.21: The curves on the right represent the percentage of patterns shared for the realizations in Figure 4.20 and the training images presented on the abscissa axis. The red, the green, and the blue curves correspond to the configuration of weights of 75% TI_1 - 25% $TI_{Rotated}$, 50% TI_1 - 50% $TI_{Rotated}$ and 25% TI_1 - 75% $TI_{Rotated}$ respectively.

Figure 4.21 shows the measure of similarity for scenarios of 75% - 25%, 50% - 50% and 25% - 75%. As expected, the lower the rotation, the more similar the images and more multiple point configurations are shared. The similarity measure between training images helps to understand the expected patterns obtained in the realizations and plays an essential role in the success of features reproduction.

4.6.3 Accounting for Multiple Stochastic Training Images

A key assumption of the proposed methodology relies on the randomness of this training image to attain the correct amount of geologic variability. However, the image is a realization of a stochastic process, and the question is whether the use of this unique representation is enough to find the correct geologic variability. The current subsection is devoted to analyse the reproduction of multiple point configurations accounting for multiple stochastic training images, all of them are realizations of the same geostatistical method.

The analysis consists of applying the proposed algorithm in the same manner as described in Algorithm 2, but assuming that for each realization (100 in total)
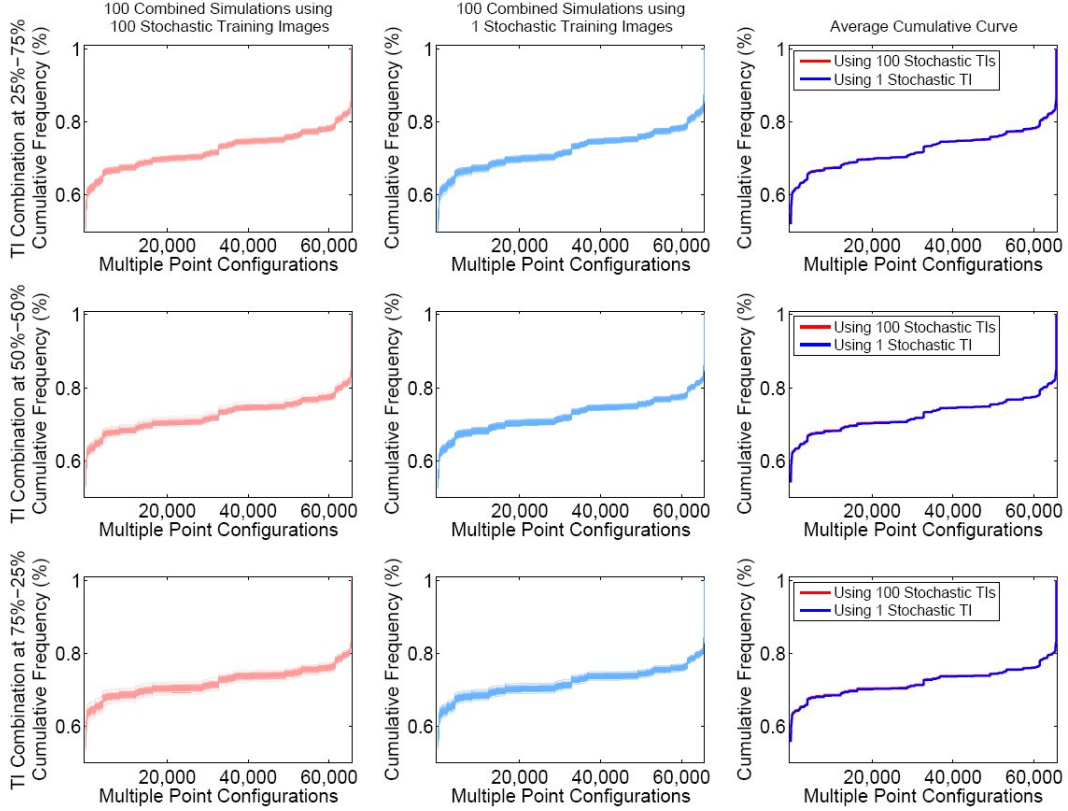
Figure 4.22: Cumulative multiple point density functions curves for 100 simulations. The first column represents the simulations based on one deterministic and 100 stochastic training images generated by SIS. The second column also corresponds to 100 simulations, but this time, using one deterministic and one stochastic training image. The last column shows the averaged cumulative frequency curves for the 100 simulations. Each row represents the settings 25% DTI - 75% STI, 50% DTI - 50% STI and 75% DTI - 25% STI respectively.

an independent stochastic training image generated by SIS is considered. Then, the exercise is repeated in its original version, that is, considering an unique stochastic training image. The first column of Figure 4.22 shows the cumulative frequency curves for each combined simulation when using an independent stochastic training image. The second column also shows the multiple point density function cumulative curves but for those simulations based on a unique stochastic training image. The third column summarizes the previous two columns by plotting the average of the cumulative curves. The rows represent the cases when the training image are being combined using configurations of weights of 25%-75%, 50%-50% and 75%-25%, respectively. In expected value, the distribution of multiple point configurations for

111

all realizations are quite similar for both cases and configurations of weights, see the third column. The variability of the cumulative multiple point density functions is also similar. The introduction of multiple stochastic training image does not necessarily imply an increase in the variability of the simulated realizations.
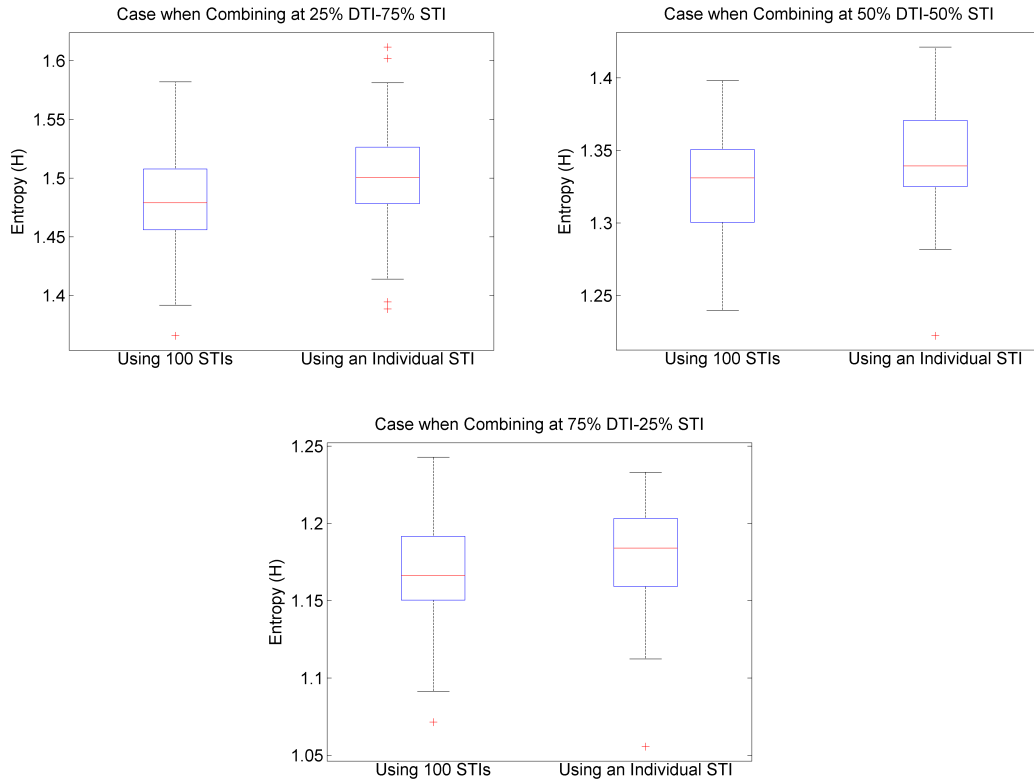
Figure 4.23: Entropy boxplot computed from the 100 simulations. On the left of each figure, the distribution of the entropy when using 100 STI; on the right, the case of an individual STI. The entropy was calculated based on the multiple point density functions using a template of 4 by 4. Three settings at 25% DTI - 75% STI, 50% DTI - 50% STI and 75% DTI - 25% STI are examined.

For a more detailed comparison, the variability of the realizations was evaluated by computing their entropy. Boxplots of the distribution of the entropy for each simulation, and for each weighting scenario are presented in Figure 4.23. Again, the entropy distributions look similar independent of the number of stochastic training images used. A subtle larger variance may be observed in favor of the case when using 100 training images. This is because more multiple point configurations are

available and incorporated to the simulations; the consequence is a subtle increase in the entropy. Despite the huge number of multiple point configurations available when using multiple stochastic training images, minor differences in terms of the reproduced patterns are observed. The use of only one stochastic training image is acceptable to produce the target level of geologic variability in the realizations.

4.7   Model Checking

In order to accept the proposed method as a reliable tool for geologic modeling, it should be validated; however, this is not a straightforward task. The idea of validation calls for the true spatial distribution of the geologic phenomenon, but in earth sciences this distribution remains unknown (Oreskes et al., 1994). Thus, direct comparison between the phenomenon and the model is not possible. Despite this limitation, cross-validation permits some checking of the proposed method.

4.7.1   Measures of Checking for Geostatistical Uncertainty Distributions

In what follows, a series of measures are presented. Issues related to their use in the context of the LOP scheme are shown through an artificial example. Then, an alternative measure for assessing the geostatistical uncertainty distributions is proposed.

Accuracy and Precision

The goodness of a stochastic rock type modeling method must be assessed in terms of its accuracy and precision (Deutsch and Wang, 1996). An adequate categorical geostatistical technique should correctly predict the proportion of samples assigned to a category k according to the predicted probabilities specified for that given proportion Deutsch (2002). Accuracy plots Deutsch and Deutsch (2012) are a good tool for displaying and evaluating the goodness of the resulting uncertainty models.

In practice, accuracy refers to the fraction of locations in the model that were

113

effectively predicted by the estimated probabilities, e.g. if considering all locations with a 50% chance to be indicator k, at least 50% of those locations must belong to that indicator. Precision is measured by the closeness to the estimated probability, that is, when the experimental proportion of the respective quantile k is equal to predicted, then the model is precise.

Closeness of Conditional Probability to the True

The second measure quantifies the closeness of the estimated conditional probabilities to the true. Eq 4.8 corresponds to the average of the estimated probabilities of the true category k; ideally, it would be close to one. It gives a measure of the certainty of the probabilistic model.

$$C = \frac{1}{N} \sum_{i=1}^{N} \{p_k(\mathbf{u}_i)|k = true\}, \tag{4.8}$$

Where N is the number of locations considered, and $p_k(\mathbf{u}_i)$ the probability to belong to the category k at location $\mathbf{u}_i$.

The Informative Level of the Model

The third measure relates to how informative the probabilities are for prediction. This measure is the difference between the averaged probabilities when the indicator is *true* and *false*:

$$B = \frac{1}{N} \sum_{i=1}^{N} \{p_k(\mathbf{u}_i)|k = true\} - \frac{1}{N} \sum_{i=1}^{N} \{p_k(\mathbf{u}_i)|k = false\}, \tag{4.9}$$

The larger the B value, the better the model in predicting where rock types are and where they are not.

4.7.2 An Example

The objective of the example is to evaluate the uncertainty in the models when applying the proposed MPS method. As access to the true distribution is unattain-

able, an individual realization is assumed as the reference model; then, the resulting uncertainty distributions are compared to it.

The exercise considers the case of the turbidite reservoir of Figure 4.4. The same nine configurations of weights in intervals of 10% are considered. One hundred realizations are performed for each configuration. Figure 4.24 shows the probability maps computed from the 100 realizations. Note the increasing certainty for the turbidite lobes when augmenting the percentage allocated to the deterministic training image. This can be noted by the systematic reduction of probabilities in the intermediate quantiles.

Computation of Accuracy and Precision: The reference model corresponds to a realization using 50% of the deterministic and 50% of the stochastic training image. Figure 4.25 summarizes the accuracy plots when evaluating the conditional probabilities respect to this model. The fraction of locations for each quantile on the abscissa must lie above the diagonal to be accurate. The model is precise when the dots are close to the diagonal. The resulting geostatistical models are reasonably accurate and precise when using a large percentage of the stochastic training image; in particular, minimum differences are appreciated for the cases when using 90%, 80% and 70% of STI. On the contrary, the accuracy curve starts deviating from the diagonal when increasing the proportion of DTI in the models. Specifically, a loss of accuracy is observed from the case 40% DTI up to 90% DTI. This is due to the concentration of the conditional probabilities towards the extreme quantiles, which can be seen in the probability maps of Figure 4.24. As the percentage of the DTI increases, the probabilities tend to zero or one which leads to a loss of accuracy and precision.
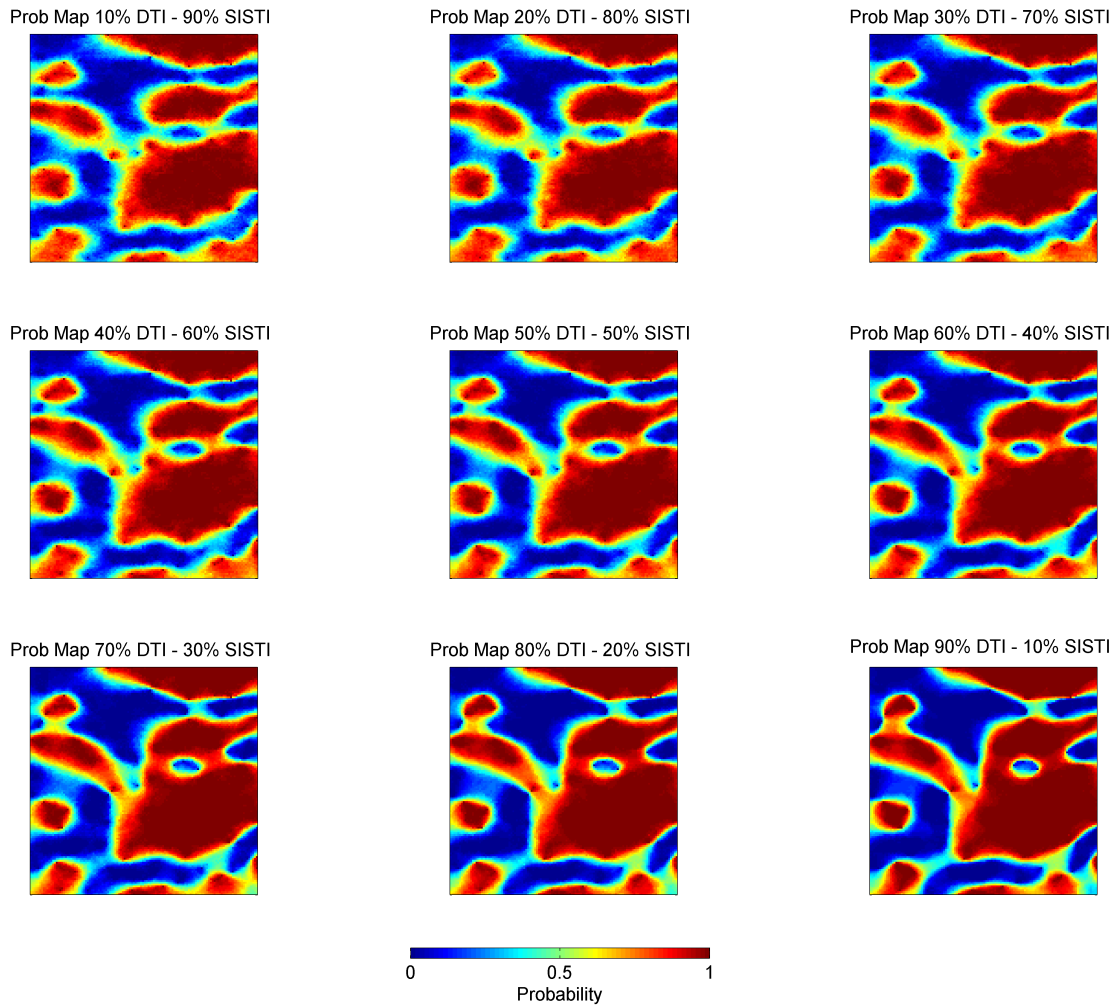
Figure 4.24: Probability maps computed over 100 simulations for the indicator category 1. Each image represents configurations of weights in intervals of 10%.
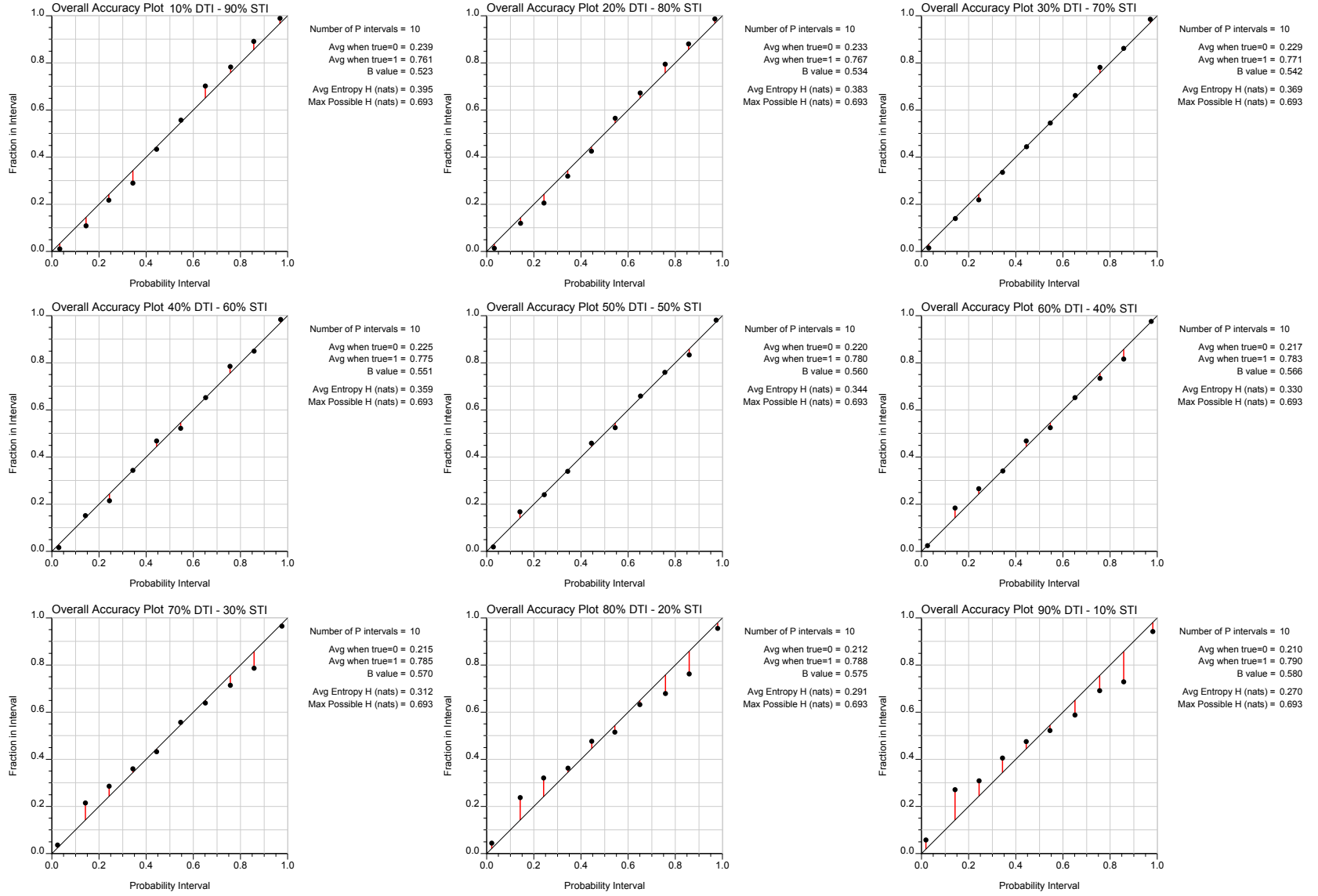
Figure 4.25: Accuracy plots considering the reference model set up at 50% DTI-50% STI.

Determination of Predictive Capability (B Measure): The second assessment consists of the computation of the B measure, Eq 4.9. Figure 4.26 illustrates its computation when considering the extremes cases of 10% DTI and 90% DTI. For the case of 10% DTI, the conditional probabilities tend to be more uniformly distributed; the opposite happens when using 90% DTI. In this case, the probabilities tend to be a highly concentrated close to 0 and 1, and the B measure increases from 0.52 to 0.58.
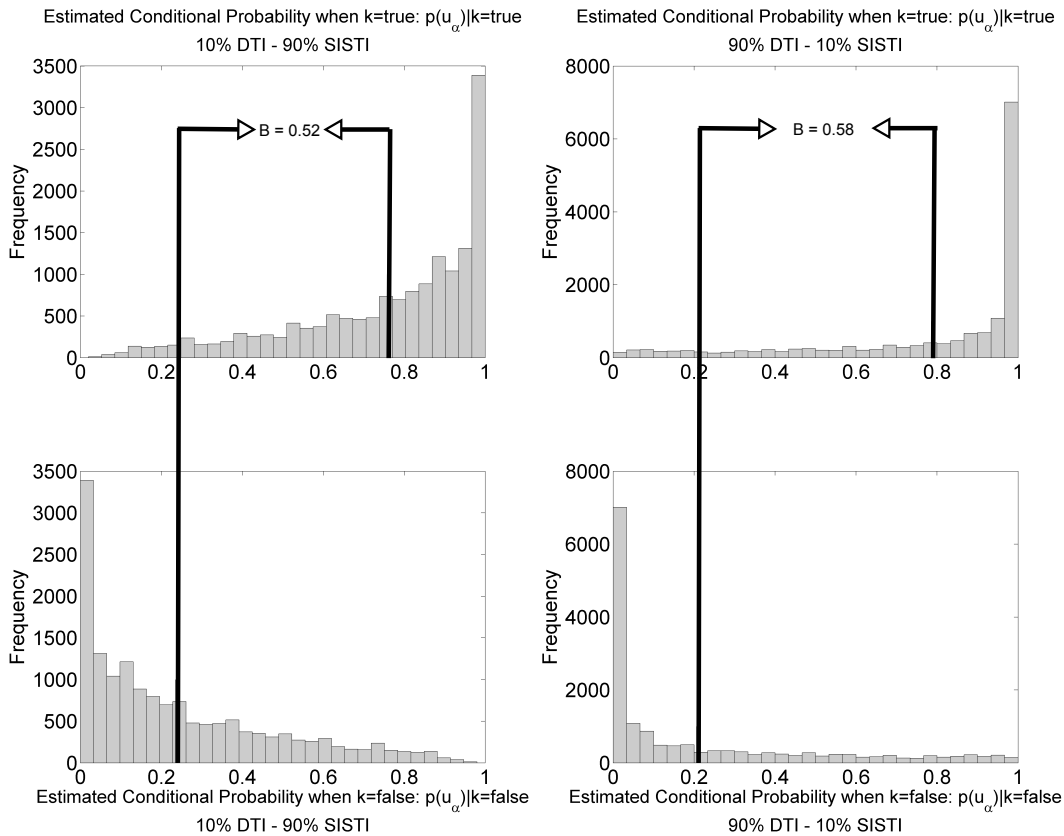
Figure 4.26: Distribution of estimated conditional probabilities when the actual reference model is true and false. Extremes scenarios of 10% DTI-90% STI and 90% DTI-10% STI are considered.

The computation of the B measure is extended to the nine configurations as plotted in Figure 4.27. The B measure increases with the proportion of the DTI. The individual use of the deterministic training image provides a better estimation of the probabilities when the indicator is true. However, the probabilistic model

118

seems to certain, that is, probabilities close to 0 and 1 appear too common given the level of knowledge provided by the conditioning data.
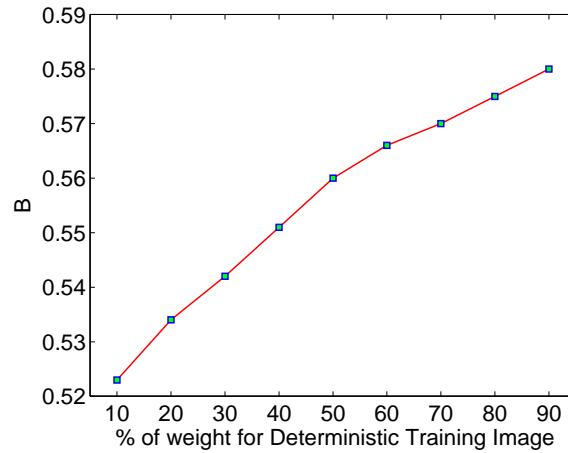
Figure 4.27: Computation of B measure for the nine models considered in Figure 4.24.

For the reasons above, the assessment of the geologic models generated by the integration of training images is not straightforward. The accuracy plots and the computation of the B measure do not constitute of reliable tools for evaluating the uncertainty in the models in the current context. The fact of having probabilities nearby the extremes does not necessarily imply a better model. Both criteria must be combined to express the adequacy of the models .

4.7.3 A Combined Measure of Accuracy, Precision and Predictive Capability

A performance measure is needed that captures an acceptable level of accuracy, precision, and predictive capability. The proposed measure relates the results of accuracy and precision from accuracy plots with the B measure. To summarize the accuracy and precision in one component, the slope of the linear regression between the predicted probability and the true probability is proposed; this corresponds to the regression between fractional interval and probability interval in Figure 4.25. Typically the slope of a linear function $f(y) = mx+n$ can be defined by $m = \frac{Cov[y,x]}{Var[x]}$.

119

Thus, consider:

$$B_{Slope} = \frac{Cov[p, p']}{Var[p']} * B \qquad (4.10)$$

Where p corresponds to the predicted probability (y axis) and the p' the true (x axis), and B is the measure of closeness (Eq 4.9). An interesting property of this measure is that belongs to the interval [0,1]; a large implies greater accuracy, precision and predictive capability.
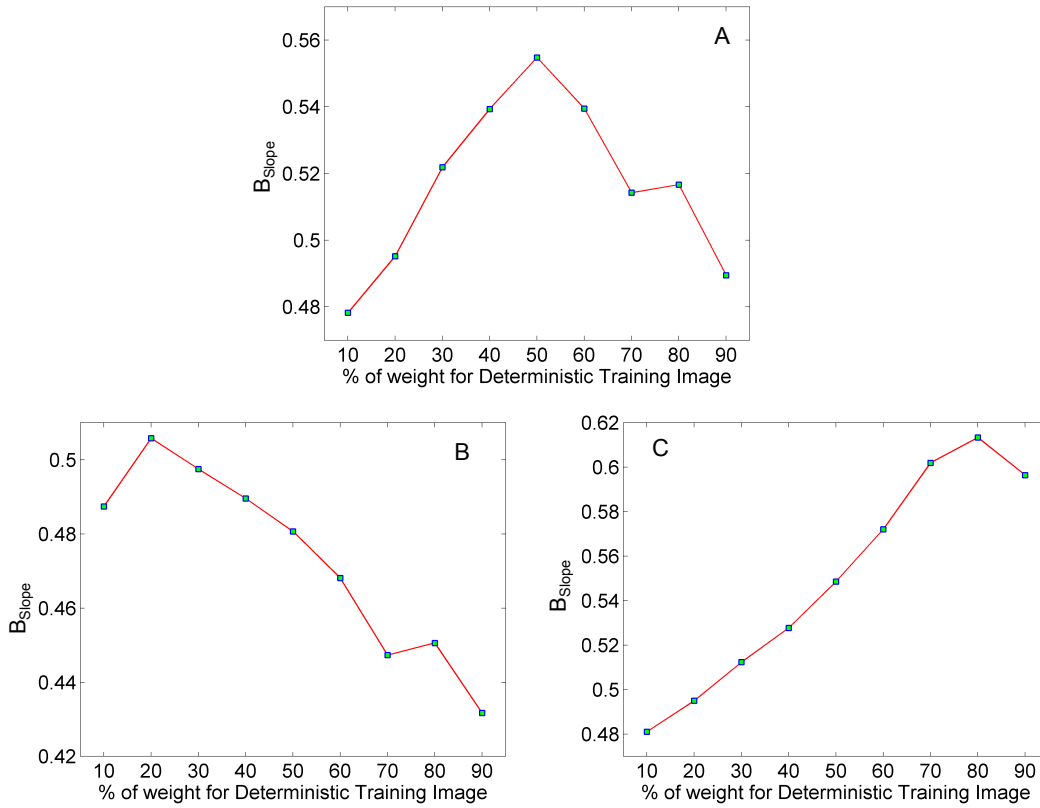
Figure 4.28: B_{Slope} curve for three different reference models. Case A based on 50% DTI and 50% DTI; case B based on 20% DTI and 80% DTI, and C based on 80% DTI and 20% DTI.

The B_{Slope} statistic is computed for the nine configurations of Figure 4.24. Recall that the reference was conceived from a realization combining the training images using configurations of weights of 50% DTI - 50% STI. The resulting curve of B_{Slope} is plotted in Figure 4.28 A. In this case, the B_{Slope} is able to associate the most

accurate and informative geostatistical model to its respective reference model (50% DTI - 50% STI), that is, the B_{Slope} reaches the maximum at 50% DTI. To extends the analysis, curves B and C illustrates the correct association of the MPS geostatistical models when the reference is changed to cases based on a configuration of 20% DTI - 80% STI, and 80% DTI - 20% STI. In both cases, B_{Slope} reaches the maximum at 20% DTI and 80% DTI, predicting correctly the reference model.

The B_{Slope} summarizes the accuracy, precision and predictive ability of the model within a checking framework. However, it does not constitute by itself an absolute measure of judgment. The proposed measure is one among the set of available checking tools.

4.8 Remarks

Although the algorithm was applied for training images constructed using variogram-based techniques, the methodology is extensible to training images of different character. The deterministic training image might be created from a classical hand-drawn approach, some basic geologic sketch, or a object-based technique. On the other hand, any geostatistical method may be used for the training images with greater variability. The combining process may be put into practice with any class of training images. The integration of conditional probabilities by a weighted linear function inevitably leads to features in the images that are a combination of both training images.

The original *snesim* code was modified to process multiple training images and compute conditional probabilities based on the same data event. The processing time is proportional to the number of training images used. The required memory also increases according to the number of training images. For the original *snesim* code for instance, the amount of memory necessary to store and simulate a block model of 256x256x256 using a training image of 256x256x256 is approximately 500 MB, depending on the complexity of the contained patterns in the image. An additional training image doubles the required memory.

The implementation of the method is flexible with respect to the number of images that could be used. The LOP could provide solutions to other geologic modeling problems. For instance, combining multiple training images using the linear opinion pool scheme might address solutions either for non-stationary phenomena or provide alternatives for modeling increasingly complex geologic environments starting from simplified versions. Further research could be undertaken on these paths and alternative algorithms explored.

The proposed methodology is useful for mineral resources estimation since reproducing the correct spatial arrangement variability is important for accurately estimating dilution and lost ore in the mining process. The estimated resources/reserves are directly affected by the large/short scale domains features. An incorrect prediction of dilution and lost ore by classical deterministic or geostatistical techniques would result in a gap between predicted and true economic performance. This will be reviewed in the case study in Chapter 6.

Chapter 5

Special Topics of MPS Simulation with Multiple Training Images

A series of advances related to the integration of multiple training images based on the LOP scheme for MPS simulation are presented in this chapter. **Section 5.1** describes an extension of the LOP where the combination of training images is implemented by distinguishing different scales of geologic features. The assignment of weights to the training images varies throughout the simulation using a function based on data spacing. The approach gains some flexibility to handle geologic variability at specific scales.

Section 5.2 explores the use of the LOP to model non-stationary geological environments; two algorithms are proposed. The first couples multiple training images guided by a range field. This field establishes which training image are combined and the area where the combination takes place. For a correct performance, a certain level of similarity among training images is required. The second algorithm proposes to couple training images by building one search-tree that stores patterns from all the images; this incorporates training images of different character avoiding the requirement of similarity and reproducing correct patterns as needed.

Finally, **Section 5.3** analyses the use of an interpreted geologic model as a training image. The introduction of local proportion models and post-processing methods for treating inconsistencies in the realizations are also discussed.

5.1 Dynamic Training Image Integration based on Data-Spacing

In this section, the method of multiple training images MPS simulation is extended. Instead of considering a fixed configuration of weights for each training image, the weights are determined using a function that depends on the scale of the features simulated. The proposed method offers an alternative for reproducing large-medium-short scale geologic features slightly different than the LOP scheme. The results provides flexibility as to the geologic features reproduced at multiple scales.

5.1.1 Overview of Sequential MPS Simulation

Section 2.5.2 discusses the importance of the conditional probabilities for a reliable reproduction of features from the training image. In particular, the hierarchical and sequential character of MPS simulation ensures the reproduction of large and short scales features.

To illustrate this point, consider the 1-D conceptualization of the MPS simulation process in Figure 5.1. The simulation schematically performs from left to right, and the arrow at the top refers to the random path followed when simulating the nodes along the grid. In the middle, configurations A, B, C and D represent the reduction in the scale of the features in the grid. By directly considering the scale of features, the geologic variability could be treated in a more flexible manner. Under certain circumstances, the modeler could maintain control over large scale geologic features such as those represented by the deterministic training image, but adding more variability at some scales. The geologic variability may be controlled by the type of function adopted; the red and green line at the bottom of the figure represent two different alternatives. For the red function, the influence of the stochastic training
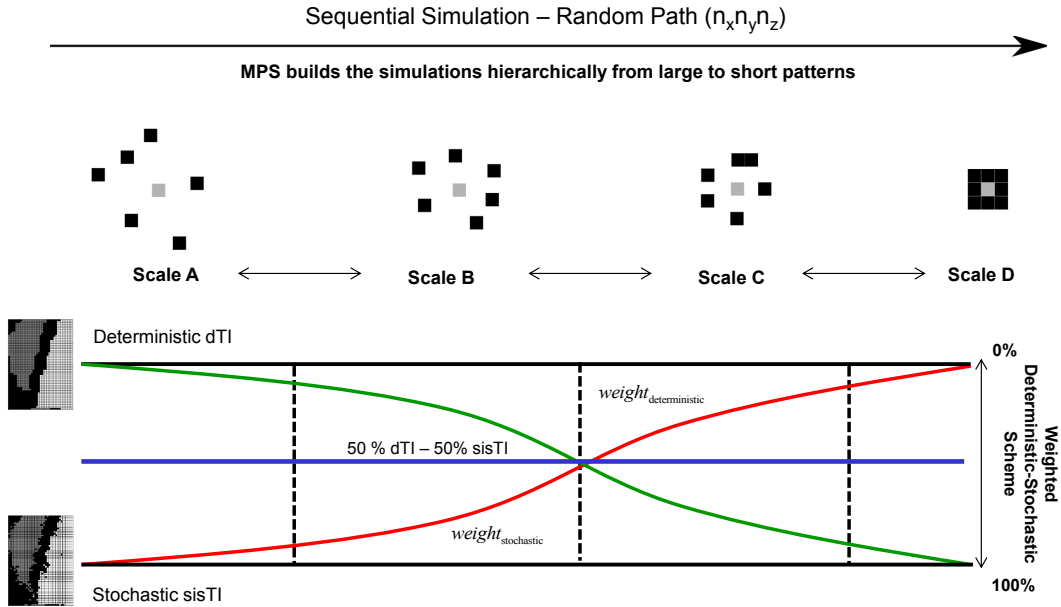
Figure 5.1: Schematic illustration of the sequential MPS simulation. At the top, a representation of the random path. In the middle, an illustration that shows the reduction of data event throughout the simulating path. At the bottom, two functions (red and green) that establish how the training images will be combined depending on the scale of the geologic features.

image is increasingly incorporated into the simulation; the last nodes of the grid are simulated almost exclusively making use of this image. This scheme ensures the increasing addition of randomness from large to short-scale features. If the opposite case is desired, the green function should be used. Here, the influence of the stochastic training image decreases along the simulation path. The resultant realization would be characterized by short scale features extracted almost entirely from the deterministic training image.

5.1.2   The Method

The method considers two modifications to the proposed MPS algorithm for the combination of training images presented in Section 4.4.3. First, the introduction of a measure that accounts for the scale of geologic features; secondly, the selection of functions for the assignment of weights to each training image.

Data-Spacing Measure

A widely used measure for characterizing the spacing between adjacent data consists of determining the number of nearby data around the location $\mathbf{u}$, $n_{samples}(\mathbf{u})$, within a volume V. In particular, the measure is given by the cube root of the ratio $V/n_{samples}(\mathbf{u})$; in the case of two dimensions, the volume is replaced by an area A and a square root is used. A two dimensional illustration is presented in Figure 5.2. Two different areas A and A' are considered; both contain the same number of samples. For the area A', the adjacent samples are closer if compared to the samples of the area A, that is, $A/n_{samples}(\mathbf{u}) > A'/n_{samples}(\mathbf{u})$.

$$data_{spacing}(\mathbf{u}) = \frac{1}{\sqrt[3]{\dfrac{V}{number_{samples}(\mathbf{u})}}} \tag{5.1}$$
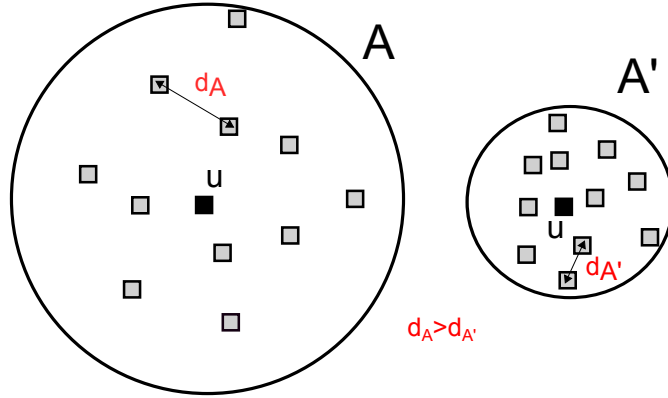
Figure 5.2: Data spacing measure illustration.

For convenience, the data-spacing relation $V/n_{samples}(\mathbf{u})$ is inverted (Eq 5.1), and thus represented as an increasing function, that is, as large its value closer adjacent nodes. In this manner, the data-spacing value can be easily integrated into the weighting function as presented in 5.3. The volume is constant and directly related to the size and configuration of the template used for MPS simulation. The measure goes from zero, when there is no samples found, to one in the case that the template is completely occupied by data.

The Weighting Function

The second modification consists of defining the function that relates the data-spacing $data_{spacing}(\mathbf{u})$ with the weights to be assigned to each training image. An interesting function commonly used in Geostatistics for tail extrapolation modeling is the power model $f(z) = z^w$. The function requires the definition of two parameters: the power w, and the type of model adopted. Both will be reviewed below.

Two Models: The power model establishes the manner that the weights will be assigned to the training images based on the data-spacing $data_{spacing}(\mathbf{u})$ measure. Two configurations could be adopted, an increasing and decreasing power model, see Figure 5.3 for illustration.
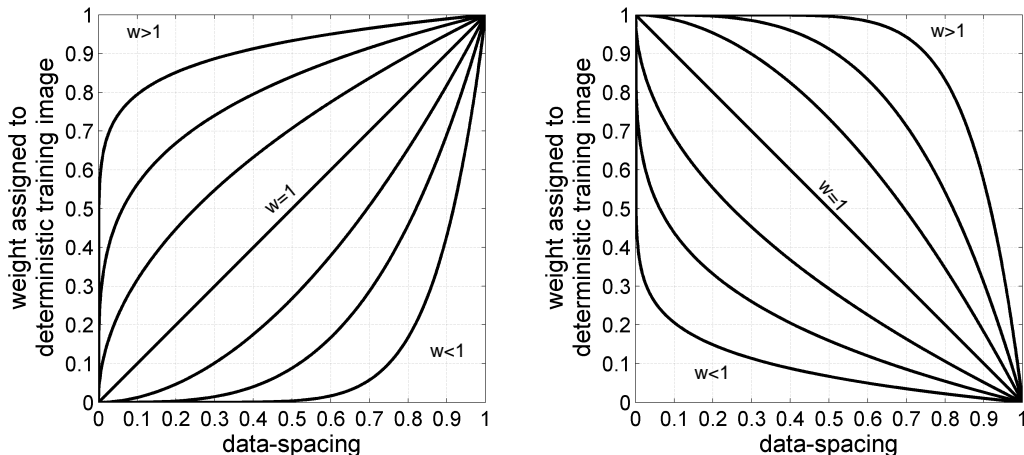
Figure 5.3: Power models proposed for the assignment of weights of the training images. The shape of the function depends on the value of the power w and the data spacing measure given by Eq 5.1.

Note that according to Eq 5.1, the $data_{spacing}(\mathbf{u})$ increases as the grid of simulation becomes denser; in other words, it increases while the scale of geologic features being simulated is smaller. For instance, if the increasing type is adopted, the weights assigned for simulating large scale geologic structures relies on the stochastic training image. The configuration changes when the last nodes of the grid are simulated; in this case, the deterministic training image becomes relevant and provides most of the features for short scale modeling. On the contrary, if the large

127

scale structures are required to be as continuous like the deterministic training image, the decreasing model is adopted. The value of the power w controls the shape of the function and the type of model allows managing the ordering of the geologic features introduced in the models.

Tuning the power w of the Model: The increasing and decreasing power models are presented in Eq 5.2. $w_{DTI}(\mathbf{u})$ corresponds to the weight assigned to the deterministic training image at location $\mathbf{u}$; the weight of the stochastic training image $w_{STI}(\mathbf{u})$ is given by $1\text{-}w_{DTI}(\mathbf{u})$.

$$w_{DTI}(\mathbf{u}) = \begin{cases} data_{spacing}(\mathbf{u})^{w=\frac{1}{w_{DTI}^c}-1}, & \text{if increasing type is adopted} \\ 1 - data_{spacing}(\mathbf{u})^{w=\frac{w_{DTI}^c}{1-w_{DTI}^c}}, & \text{if decreasing type is adopted} \end{cases} \tag{5.2}$$

A relation between the calibrated weight w_{DTI}^c as presented in Section 4.5 and power w of the functions described above must be established. The relation is given in terms of the expected value of the increasing and decreasing functions. Whether the power w is assumed to be equal to $w = \frac{1}{w_{DTI}^c} - 1$ and $w = \frac{w_{DTI}^c}{1-w_{DTI}^c}$ for the increasing and decreasing models respectively, then, $E[w_{DTI}(\mathbf{u})] = w_{DTI}^c$, that is, the expected value of the increasing or decreasing function correspond to the weights assigned after calibration.

5.1.3 Comparison to Original LOP Integration Scheme

Realizations generated based on the increasing or decreasing power functions are compared to the original LOP scheme. Three configurations are evaluated, the cases of 20% DTI - 80% STI, 50% DTI - 50% STI and 80% DTI - 20% STI. It is assumed that these configurations are the result of a calibration process as presented in Section 4.5. For each case, the increasing and decreasing models are considered. The realizations are presented on the right of Figures 5.4, 5.5 and 5.6. Deterministic and stochastic training images are shown at the top of each figure. The three function on the left represent the LOP, the increasing and the decreasing models.

All realizations are characterized by an increment in the geologic variability when
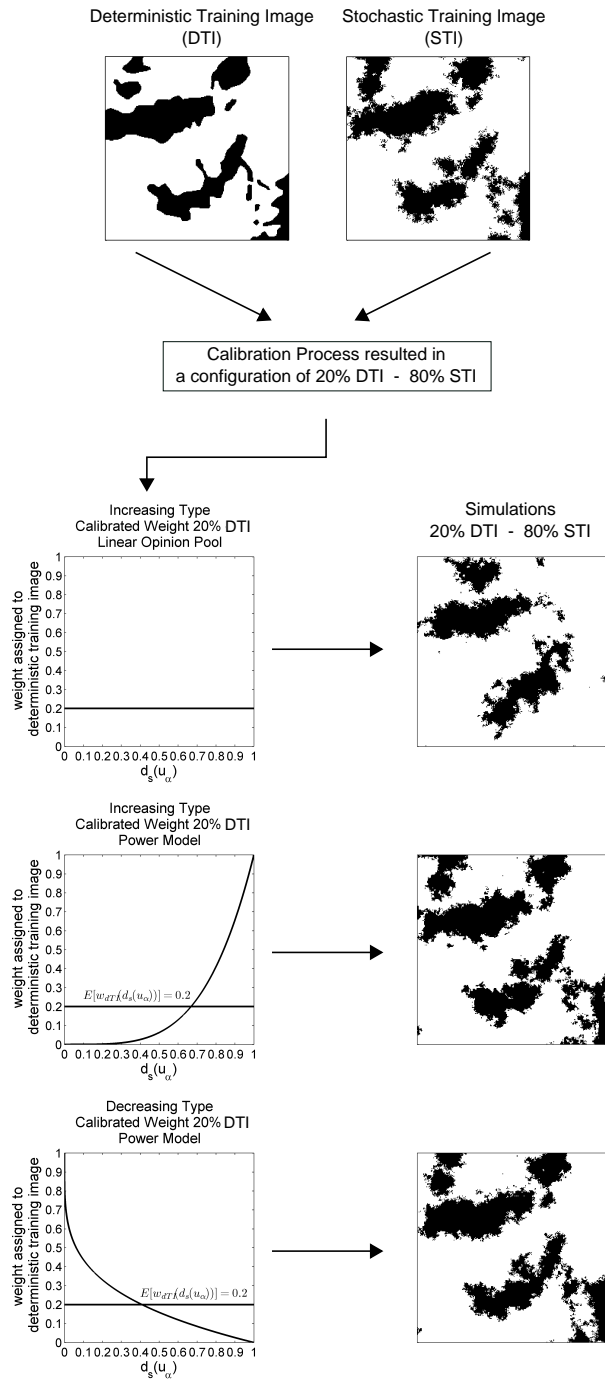
Figure 5.4: Some realizations when applying the increasing, decreasing and LOP functions based on a configuration of 20% DTI - 80% STI. At the top, the training images. On the left, the LOP, the increasing and the decreasing functions.
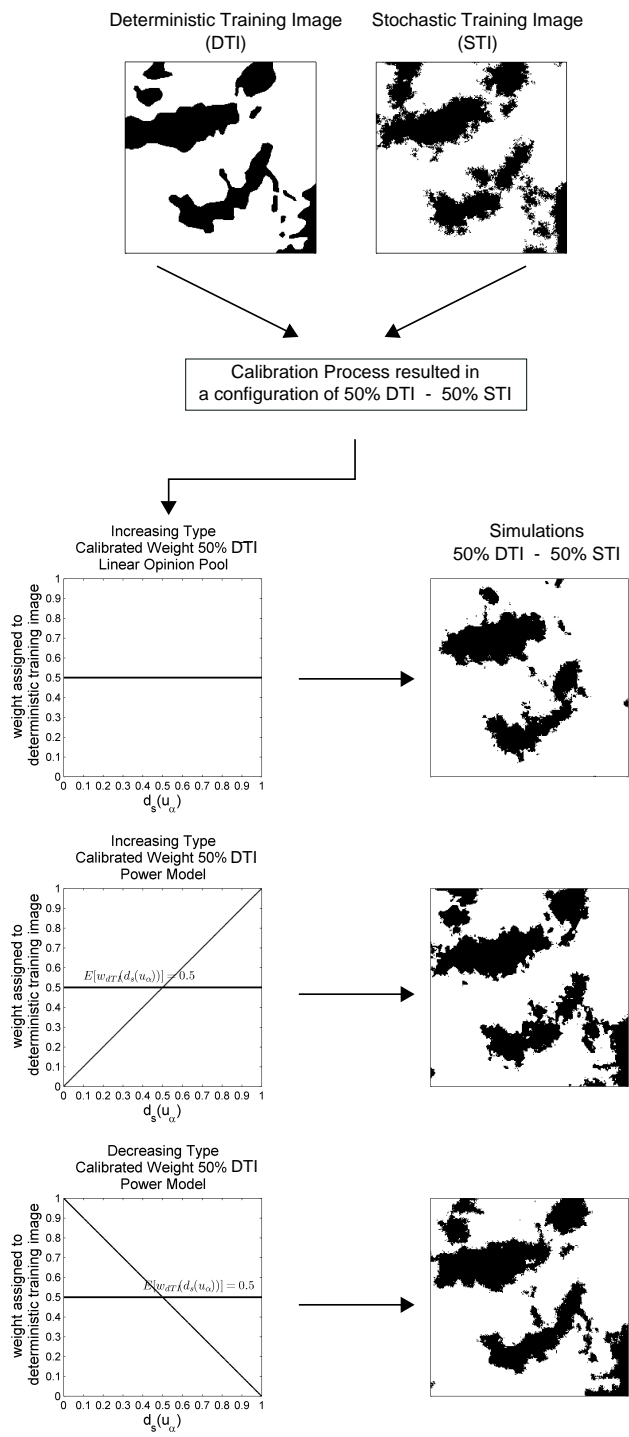
Figure 5.5: Some realizations when applying the increasing, decreasing and LOP functions based on a configuration of 50% DTI - 50% STI. Note that the power model whose expected value is 0.5 is given by linear functions.
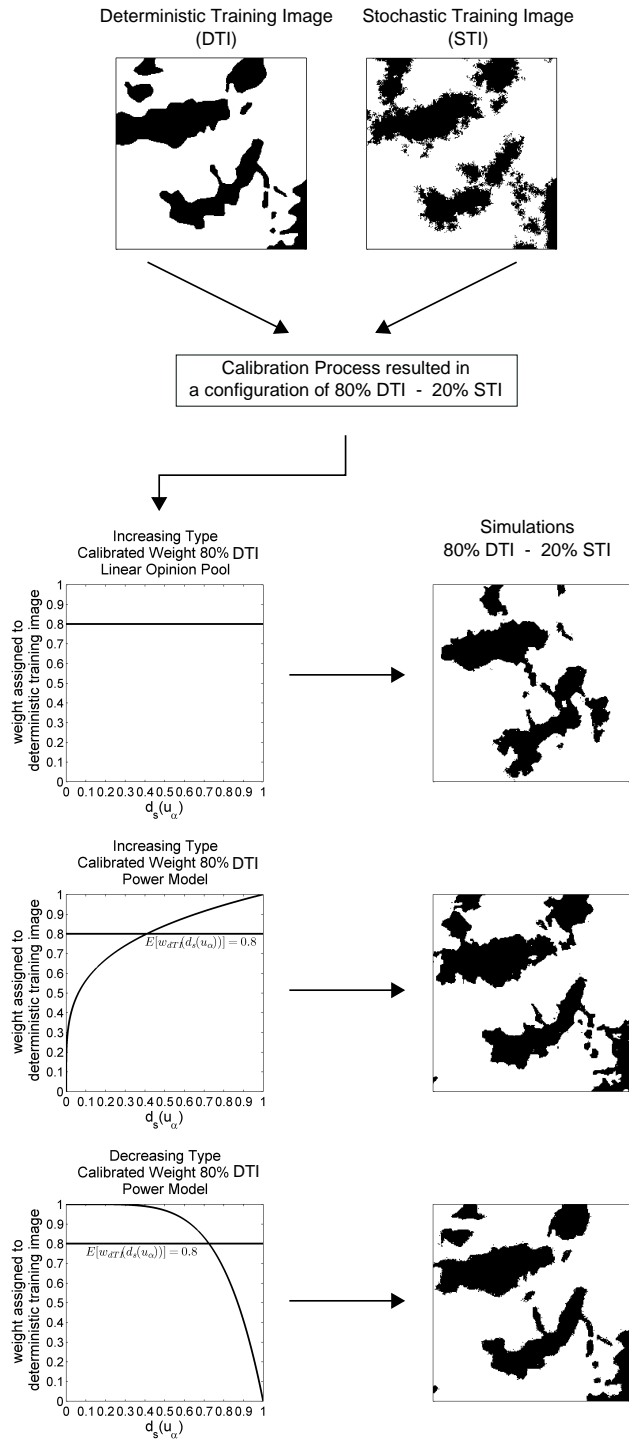
Figure 5.6: Some realizations when applying the increasing, decreasing and LOP functions based on a configuration of 80% DTI - 20% STI.

increasing the proportion of stochastic training image; however, some differences may be observed. In particular, the change of the geologic variability at different scales. In the case of the LOP, the realizations are constructed by combining both large and short scale geologic features of the training images equally at all scales. In the increasing and decreasing models cases, the realizations show how the large and short scale features are subjected to the order of the training images adopted. For instance, consider the decreasing model. Here, the large scale structures are mostly retrieved form the deterministic training image with percentages of 50% DTI - 50% STI, 20% DTI - 80% STI and 80% DTI - 20% STI, the realizations maintain their large scale structures; however, their short scale variability is controlled by the stochastic training images. The opposite occurs when considering the increasing model.

5.1.4 Characterization of Geologic Variability at Multiple Scales

This section quantifies the differences in the realizations generated by the increasing/decreasing power models and the LOP combination scheme. The differences are quantified in terms of entropy. A geologic multiple scale spatial relation may be established by comparing entropies based on configurations of templates of different size.

Two templates are used, one of 4 by 4, and a second template with the same configuration but rescaled eight times, that is, the nodes are separated by eight units. The example presented above is considered, and nine scenarios are evaluated. Each of them is based on an increment in the weights of the training images in intervals of 10%. For each scenario, a total of 10 realizations were generated. Figure 5.7 shows a scatter diagram between the entropies computed using the 4 by 4 template and its upscaled version. The nine scenarios are plotted from left to right, that is, the first correspond to the configuration 90% DTI - 10% STI, and the last to 10% DTI - 90% STI. The blue and cyan curves are associated to the increasing and decreasing power model respectively, the red corresponds to the LOP integration. The increasing and decreasing functions modifies the reproduction of large scale features. This is

132
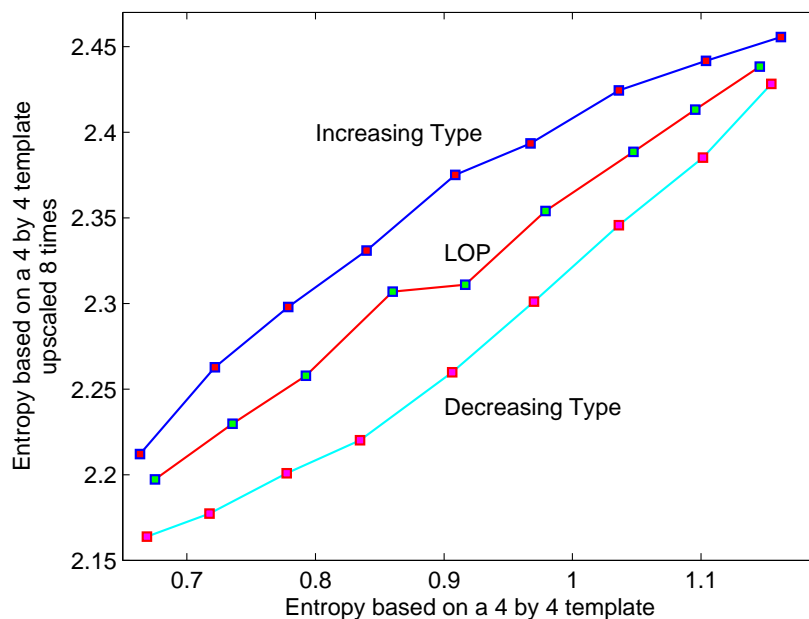
Figure 5.7: A scatter diagram of entropy at short and large scale for the increasing, decreasing and LOP scheme. On the $x-axis$, the entropy using a short four by four template; on the $y-axis$, the entropy of a larger upscaled template.

reflected by the deviations of the curves with respect to the original LOP (diagonal). The short scale entropy does not change much. Thus, the introduction of a weighting scheme based on a data spacing measure provides some control over the large scale features in the realizations by changing the combination of training images at early stages of simulation.

5.2 Non-Stationary Modeling by Coupling Training Images using Range Fields

A number of Multiple Point Statistics simulation techniques have been proposed to deal with non-stationary geologic modeling. Most of them follow the same principle by decomposing the non-stationary phenomenon into stationary regions and transitions zones. Strebelle (2002) and Harding et al. (2005) introduce a probability field as a secondary information for updating the posterior conditional probability.

The field favors the simulation of certain categories at particular locations. Boucher (2009) considers the partition of complex training images into several classes that guarantee smooth transitions between stationary regions. In a similar manner, Wu et al. (2008) investigates an approach that incorporates stationary multiple training images at different regions in space.

The problem with transition zones in a non-stationary framework is that artifacts could appear when training images change quickly. Due to the LOP combination scheme is specially attractive when combining multiple training images, this section focuses on non-stationary modeling of geological phenomena by coupling training images using this scheme. A range field establishes the transition zones and the degree of mixing of the training images.

The effectiveness of the proposed approach depends on a reasonable degree of similarity in the training images. To provide greater flexibility, a second approach is presented in Section 5.2.2. This method considers training images that are quite different.

5.2.1 The Linear Opinion Pool Approach

The linear opinion pool approach for simulating non-stationary geologic models is developed. The methodology considers the integration of multiple stationary training images by a range field. The objective consists of generating a non-stationary realization that adequately reproduces the different heterogeneities presented in multiple training images.

The use of a range model for integrating a sequence of training images was originally proposed by Manchuk and Deutsch (2012). The weights of each training image are associated to a specific location $\mathbf{u}$ within the range field. This field determines the stationary regions and transitions zones

Algorithm 3 describes the steps involved in the simulation of non-stationary models by LOP integration. First, the range field is defined (Line 1); this field and the simulation grid must have the same resolution. Consider M training images, each of them associated to one location within the range field (Line 2). The simulation

starts by visiting all the nodes of the grid (Line 3 to 13). For the current location $\mathbf{u}_\alpha$, the two adjacent training images i and $i+1$ are selected (Line 4). Then, the data event D_n surrounded by n conditioning data is retrieved (Line 5), and the conditional probability computed from training images i and $i+1$ (Line 6 to 8). At this point, the weights of the range field assigned to the training images are standardized (Line 9), and the, the probabilities are integrated linearly (Line 10). Finally, the categorical value is drawn from the updated conditional probability, and the value added to the grid. The simulation ends when all nodes have been visited and simulated.

Algorithm 3 Non-stationary MPS Simulation based on a range model, Manchuk and Deutsch (2012)

Input: Set of training images, range field
Output: Non-stationary realizations
 1: Construct a range model with the same resolution that the grid model
 2: Allocate the set of training images M to locations $i = 1, ..., M$ in the range model
 3: **for** each node $\mathbf{u}_\alpha$ of the random path $\{\alpha = 1, ..., (n_x n_y n_z)\}$ **do**
 4:     Obtain the range value $r(\mathbf{u}_\alpha)$ and select the adjacent training images at consecutive locations r_i and r_{i+1}
 5:     Retrieve the data event $D_n(\mathbf{u}_\alpha)$
 6:     **for** each training image i and $i+1$ **do**
 7:         Compute the conditional probability $P(s_k = k|D_i)$ and $P(s_k = k|D_{i+1})$
 8:     **end for**
 9:     Based on the range value $r(\mathbf{u}_\alpha)$ determine the weights for each training image by the standardization procedure: $w = \frac{r(\mathbf{u}_\alpha)-r_i}{r_{i+1}-r_i}$
10:     Integrate the conditional probabilities by LOP scheme $P(S = s_k|D_i, D_{i+1}) = wP(s = s_k|D_i) + (1-w)P(s = s_k|D_{i+1})$
11:     Draw the categorical variable from $P(S = s_k|D_i, D_{i+1})$
12:     Add the simulated value $s_k(\mathbf{u}_\alpha)$ and move to the next node $\mathbf{u}_{\alpha+1}$
13: **end for**

An example based on ellipsoid-based training images is used to illustrate the process, see Figure 5.8. The example is based on the integration of training images whose ellipsoids are oriented at different angles; the sequence chosen considers ellipsoids at $50°$, $25°$, $0°$, $25°$, $50°$, $75°$ and $90°$ (top of the figure). The training images are placed at equally spaced intervals from left to right along the range field. The resultant simulation is show at the bottom of the figure. The ellipsoids smoothly

change their orientation according to the range field. The convex property of LOP allows generating realizations in between the training images. The linear opinion pool offers a suitable framework to model non-stationary phenomena by combining the training images guided by transition zones.
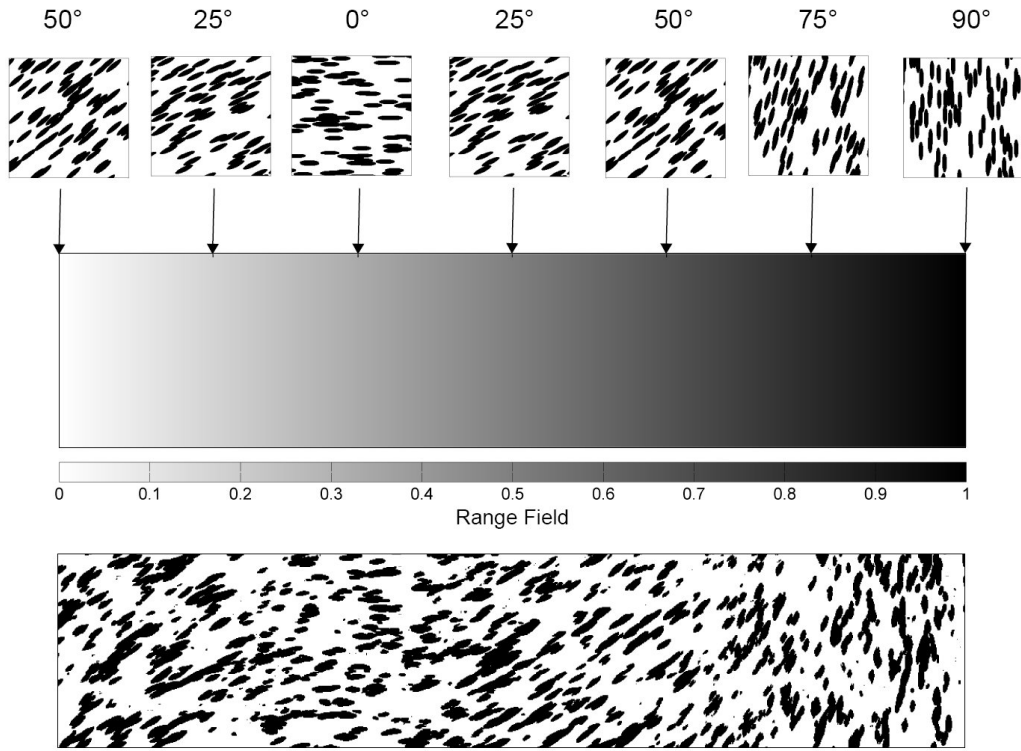
Figure 5.8: A non-stationary realization by LOP integration based on training images with ellipsoids oriented at different angles. At the top, the range field that establishes which training images will be combined. At the bottom, a non-stationary realization.

As discussed in Section 4.6.2, a central element of this LOP implementation is the similarity between adjacent training images. The integration of training images with a low level of similarity lead to unpredictable results. This is shown by an unjustified increment of the spatial disorder in the realizations (see Figure 4.20). A second example illustrating this point is shown in Figure 5.9. As in the previous example, the same range field and ellipsoid-based training images are used. Two training images are placed at the extremes of the range field. These images contain ellipsoids oriented at 0° and 90°. The MPS simulation based on the LOP scheme is performed and the realization plotted at the bottom of the figure. The lack of

136

similarity between the training images introduces features that do not relate to those contained in the training images. The effect is even stronger in the middle of the range field where the training images are combined with equal proportions. The requirement of a minimum degree of similarity between training images could be limiting for certain circumstances. One would desire a MPS method that integrates dissimilar training images and still provides of suitable models. This is discussed in the following section, where a second algorithm for non-stationary modeling is presented.
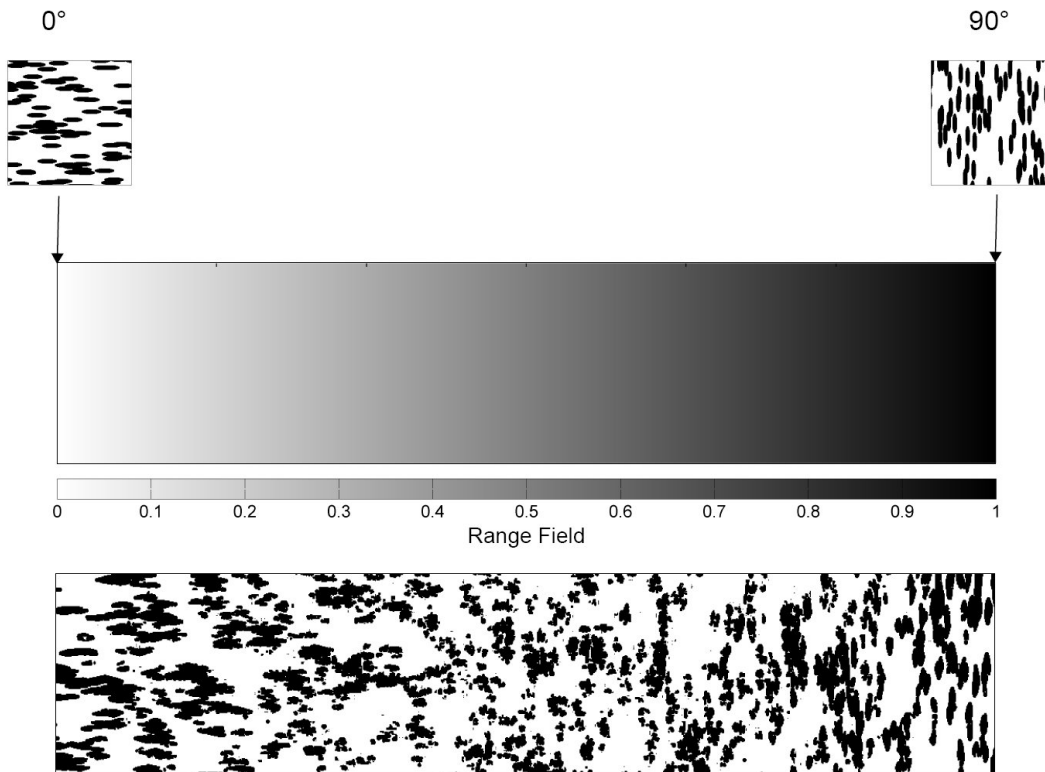
Figure 5.9: Same example as Figure 5.8, but considering training images with ellipsoids oriented at angles separated by 90 degrees. At the bottom, the realization when implementing the LOP.

5.2.2   Search-Tree Replicates Combination Algorithm (STRC)

The algorithm presented in this section differs from the previous LOP method in that the training images are integrated before computing their conditional probabilities. The algorithm is based on the search-tree storage procedure of the *snesim* method;

detailed insights on its construction and processing may be found in Appendix A. The idea is to store the multiple point configurations from both training images in the same search-tree, but instead of integrating the conditional probabilities, the number of replicates are combined first.

Consider the data event $D_{m,n}$ of n conditioning data associated to the training image m from the set of images $\{m = 1, ..., M\}$. The data event may take $\{s_k, k = 1, ..., K\}$ values. Now, consider that the training image m is scanned and the search-tree constructed. The notation $c^m_{s_k}(D_{m,n})$ denotes the node of the search-tree that stores the number of replicates of the data event $D_{m,n}$ found in the training image m whose center location $\mathbf{u}$ takes the state s_k. Then, the total number of replicates in the node considering all the categorical values $\{s_k, k = 1, ..., K\}$ is given by:

$$c^m(D_{m,n}) = \sum_{k=1}^{K} c^m_{s_k}(D_{m,n})$$ (5.3)

Instead of using the weights for the linear combination of the conditional probabilities, the weights are used for combining the number of replicates in each node of the search-tree. In the case that the pattern is shared by both training images, the final number of replicates will be given by Eq 5.4. Whether the patterns belong to just one of the training images, the weights are still applied, even though the number of replicates for the remainder training images is zero.

$$c(D_{m,n}) = \left\lceil \sum_{i=1}^{M} w_i c^i(D_{m,n}) \right\rceil$$ (5.4)

with $w_i > 0$ and $\sum_{i=1}^{M} w_i = 1$. Note that what is being combined is the number of replicates so that the result may not be an integer number, for this reason the ceiling function $\lceil \cdot \rceil$ is used. Algorithm 4 describes the methodology in a step-by-step manner. First, the range field is constructed and the training images allocated within it (Line 1 to 2). Each training image m is scanned, and the patterns found stored in

138
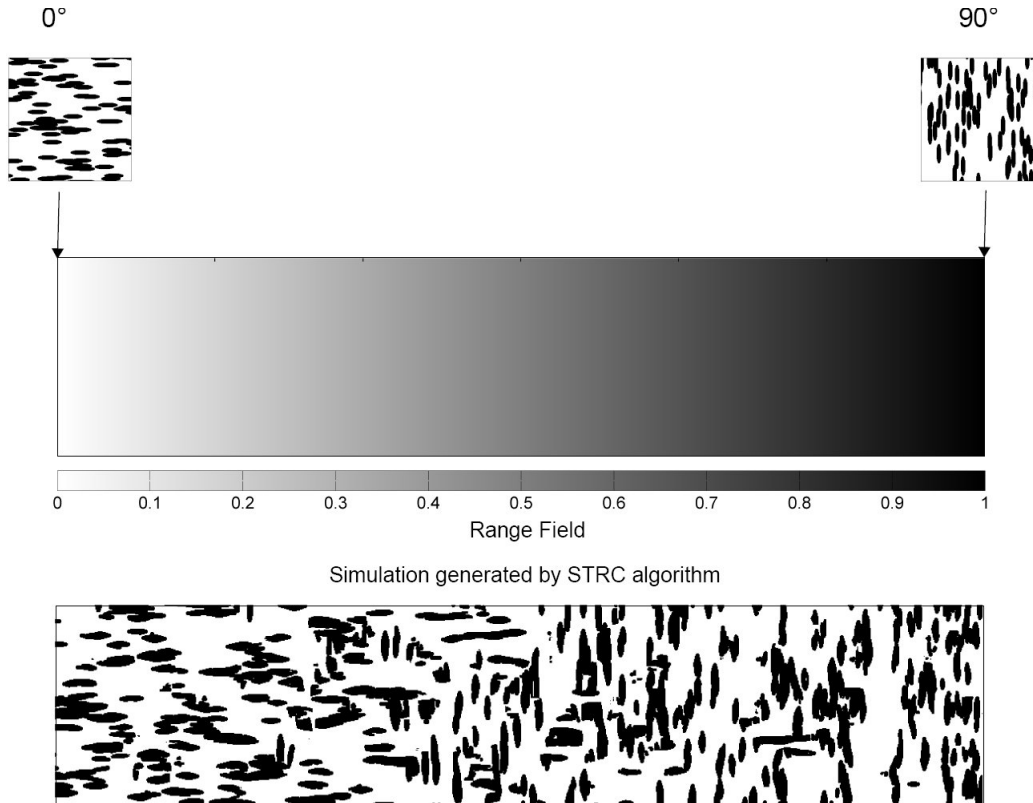
Figure 5.10: Same training images of the Figure 5.9 are used; then, the search-tree replicate combination algorithm is implemented. At the bottom, a realization guided by the range field at the top.

an individual search-tree (Line 3). Each node of the tree contains M values (one for each scanned training image). These values correspond to the number of replicates $c^m(D_{m,n})$ found for the data event $D_{m,n}$ in each training image m. Similarly to Algorithm 3, the two adjacent training images i and $i+1$ are selected, and the weights of the range field standardized (Line 5 to 6). Then , the data event D_n is retrieved (Line 7). On Line 8 the number of replicates of the search-tree are linearly combined. Based on the updated tree, the conditional probability $P(S = s_k|D_n)$ is computed (Line 9), the categorical value drawn from it (Line 10), and the next location is visited (Line 11). Special care must be taken if the training images are quite different and particularly large; in this case, the resultant search-tree may be excessively large leading to reduced computer speed and increased memory.

The example of Figure 5.9 is recalled. The same range field and training images

139

are adopted. A realization of the proposed algorithm is presented at the bottom of Figure 5.10. In this case the ellipsoids are not combined, but the ellipsoids of both image are simulated and overlapped over the same area. The weights are directly related to the frequency of appearance of 0° and 90° oriented ellipsoids in the grid. Note how the number of ellipsoids oriented at 0° and 90° increases to the extremes of the range field, that is, where the weights tend to 0 and 1 respectively. Thus, the major advantage of the algorithm is that the simulated features mostly belong to one of the training images.

Algorithm 4 Non-stationary MPS Simulation by Search-Tree Combination

---

Input: Set of training images, range field
Output: Non-stationary realizations
 1: Construct a range model with the same resolution that the grid model
 2: Allocate the set of training images M to locations $i = 1, ..., M$ in the range model
 3: Construct an individual search-tree by scanning the M training images. Each node of the search-tree contains the number of replicates $c^m(D_{m,n})$ of the data event D_n found in the training image m.
 4: **for** each node $\mathbf{u}_\alpha$ of the random path $\{\alpha = 1, ..., (n_x n_y n_z)\}$ **do**
 5:    Obtain the range value $r(\mathbf{u}_\alpha)$ and select the adjacent training images at consecutive locations r_i and r_{i+1}
 6:    Based on the range value $r(\mathbf{u}_\alpha)$ determine the weights for each training image by the standardization procedure: $w = \frac{r(\mathbf{u}_\alpha) - r_i}{r_{i+1} - r_i}$
 7:    Retrieve the data event D_n
 8:    Combine linearly the number of replicates found in each training image
$$c(D_{m,n}) = \left\lceil w_i c^i(D_{m,n}) + w_{i+1} c^{i+1}(D_{m,n}) \right\rceil$$
 9:    Compute the conditional probability $P(S = s_k | D_n)$ from the updated node $c(D_{m,n})$
 10:    Drawn the categorical variable from $P(S = s_k | D_n)$
 11:    Add the simulated value $s_k(\mathbf{u}_\alpha)$ and move to the next node $\mathbf{u}_{\alpha+1}$
 12: **end for**

The algorithm is suitable for generating non-stationary realizations based on quite different training images. Two examples based on different geologic depositional environments are developed. The first combines training images that represent deposits of fluvial origin, see at the top of Figure 5.11. Both images share three facies: Flood Plain (code 0), Channel Fill (code 1), and Channel outline (code

Channel TI

Leeve TI

Transition Zone

Zone of exclusive use of Channel

Zone of exclusive use of Leeve

0 Proportion TI 1

Two Realizations based on
the implementation of STRC

Two Realizations based on
the implementation of LOP
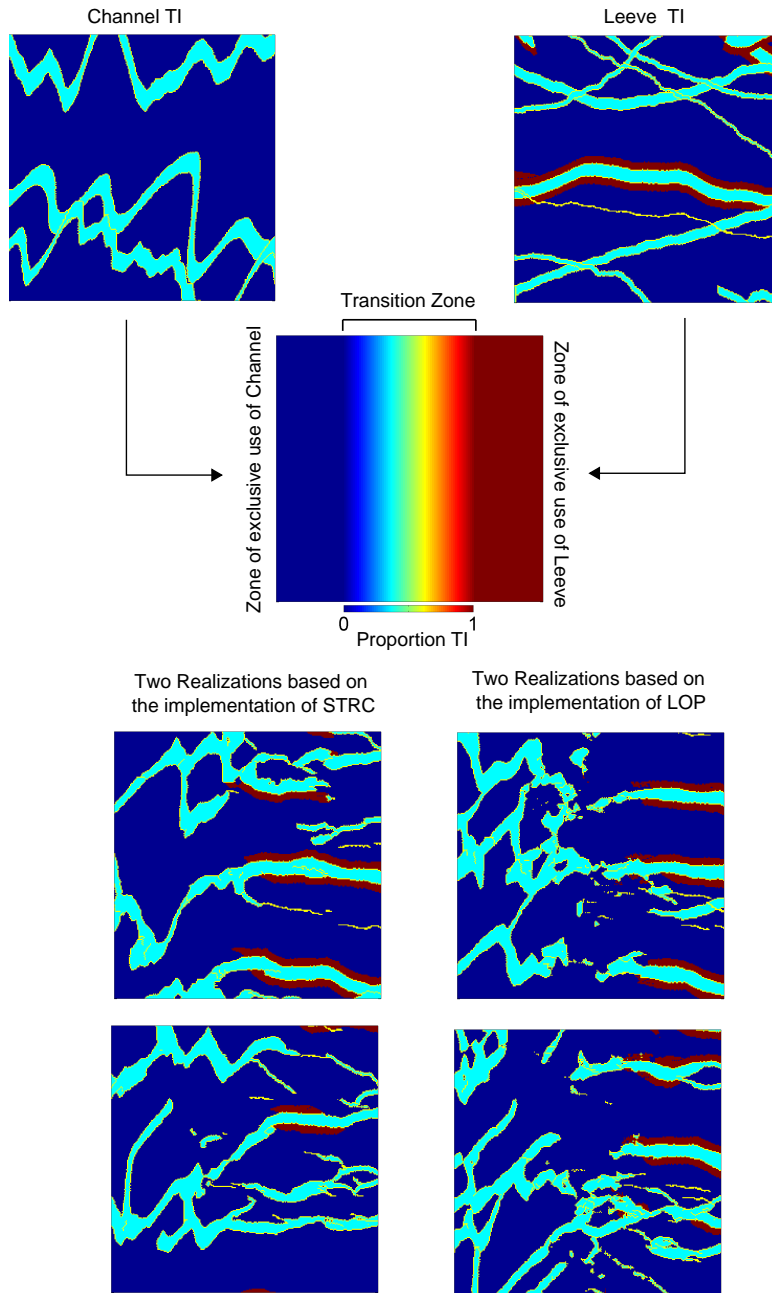
Figure 5.11: At the top, the training images; at the bottom on the left, two realizations by implementing the search-tree combination algorithm, and on the right, two realizations based on the LOP.

2), a fourth facies is added in the second training image on the right: Leeve (code 3). Beneath them, the range field that conditions the mixing process. At the bottom, four realizations are presented. On the left column, two realization based on
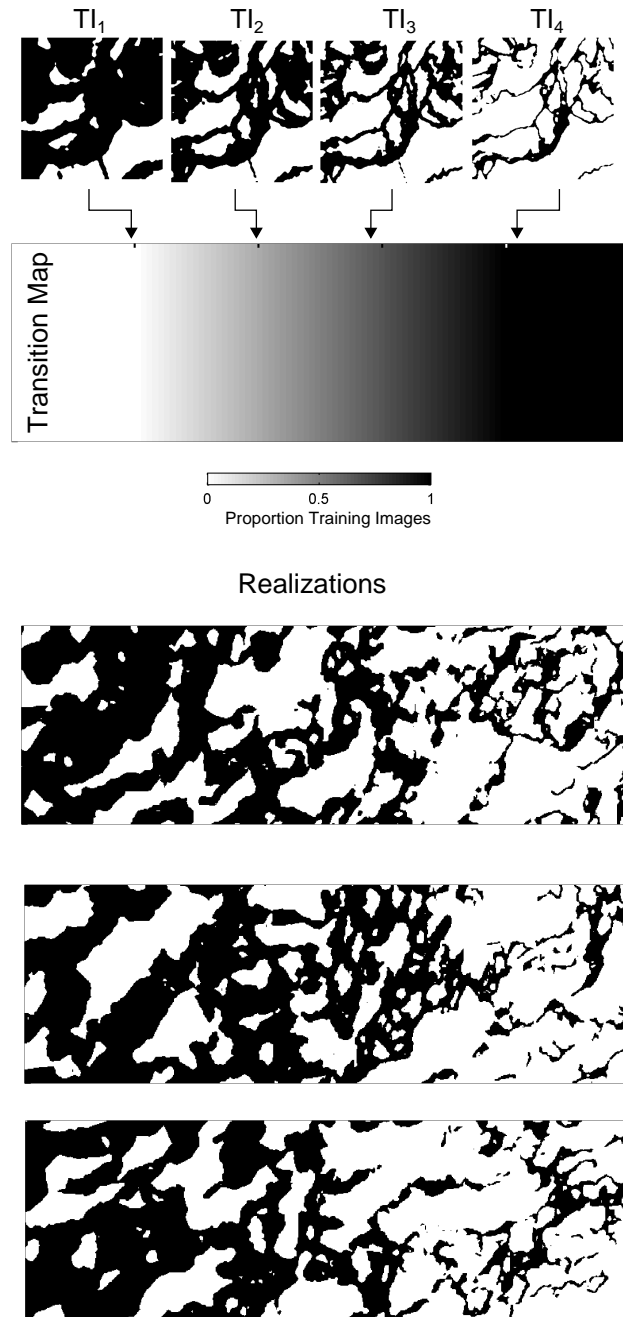
Figure 5.12: An example of a non-stationary realization using STRC of a turbidite reservoir. At the top, four training images; at the bottom, three realizations guided by the range field in the middle.

the proposed search-tree combination algorithm. Here, the features of the training images are coupled by the range field; this can be noted by the continuity of the

channels. As a comparison, the column on the right shows two realizations based on the LOP. In this case, the training images are well-reproduced close to the edges of the range field. However, the LOP scheme gives rise to features that are not originally contained in the training images, which is seen in a lack of continuity in the channels.

The second example presented in Figure 5.12 simulates a reservoir of deep water depositional origin. The training images characterize turbidite lobes of different geometry and size, seen at the top of the figure. The four images are placed at 0.2, 0.4, 0.6 and 0.8 along the range field. Three realizations are shown at the bottom. A smooth transition of the turbidite lobes through the range field is observed. There are no abrupt changes in between, and the non-stationarity of the phenomenon is reproduced.

5.3 MPS Non-Stationary Modeling using a Local Conditioned Training Image

The idea of using an individual data-driven training image for modeling mineral deposits is revisited, see Section 2.5.3 for a brief introduction. A series of practical modifications to remedy inconsistencies in the models due to non-stationarity are presented. The purpose of the current section is to explore the use of a local conditioned training image for the improvement of non-stationary MPS realization.

Two approaches are investigated. The first uses local proportion models to control the location of geologic domains in space. The second approach introduces a post-processing methodology based on a simulated annealing technique with the objective of improving the reproduction of multiple point statistics in the realizations. An example based on a sedimentary gold deposit is used for illustrating the advances.

5.3.1 Implementation of *snesim* in a Sedimentary Folded Gold Deposit

The deposit is characterized by a set of sedimentary folded structures. Figure 5.13 shows an isometric view of the geologic model; in orange and yellow, the sedimentary upper and mean sill domains, in red, the intrusive unit called sage basalt. Both drillholes dataset and the geologic model are provided by mine professional personnel.
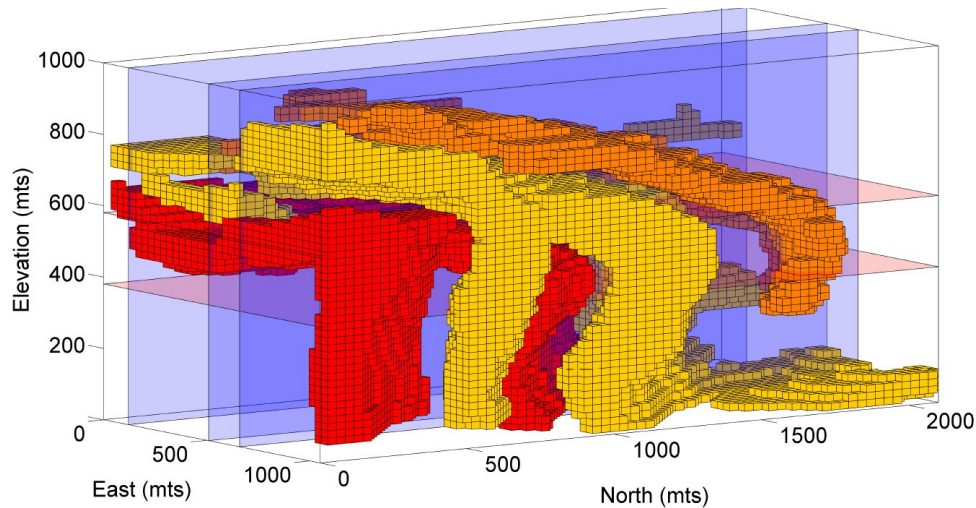
Figure 5.13: Geologic model of the gold deposit. Folded sedimentary structures and intrusive domains control the mineralization process. In orange and yellow, the sedimentary upper and mean sill domains, in red, the intrusive unit called sage basalt.

This geologic model is used as the training image for the implementation of the snesim algorithm. A maximum number of samples of 128, a neighborhood of 750 by 750 by 750 meters and template of 5 by 5 by 5 were used for simulating. Neither servosystem for correcting global proportions of domains nor local proportion models were implemented. Figure 5.14 shows three cross sections and two plan views of a resultant realization. Even though the folding features are transfered from the training image, a series of inconsistencies such as some unforeseen noise and a lack of continuity in certain locations may be observed. This is because some configurations in the realizations are not necessarily contained in the training images. For handling these issues, the use of local proportions models and a post-processing technique is

144

developed.
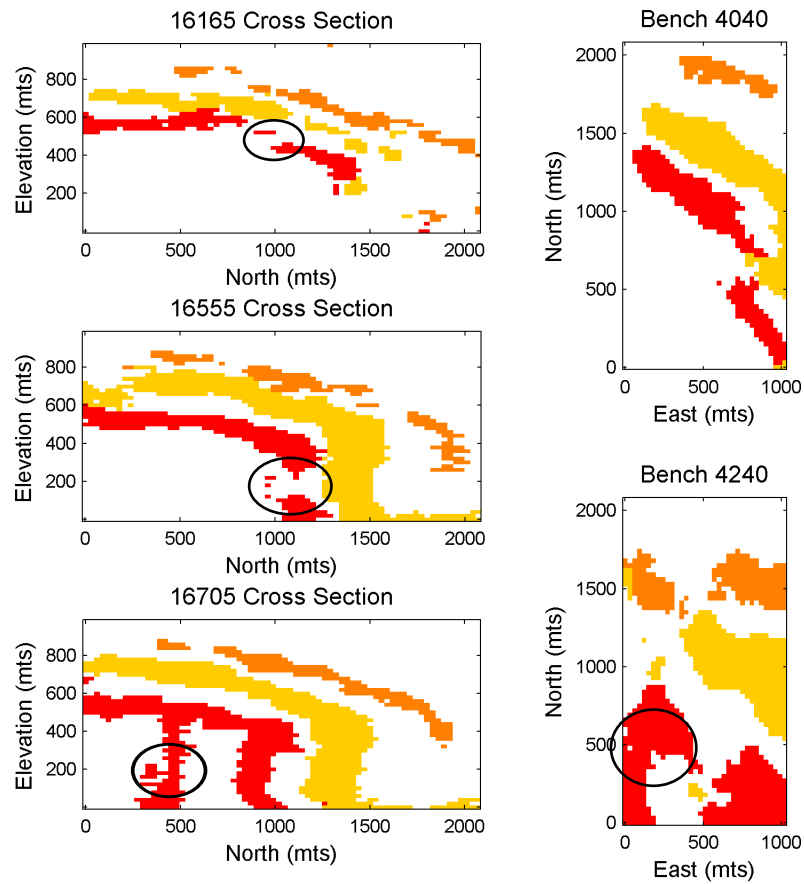
Figure 5.14: An MPS realization using the geologic model of Figure 5.13 as a training image. The black circles denote two major problems: unwanted noise and lack of continuity for certain domains.

5.3.2 Integration of Local Proportion Models

Local proportions models provide of secondary information to deal with the lack of continuity. The models establish a level of knowledge about the presence and absence of particular domains in the deposit. The information of these domains is contained in the form of proportions and integrated into the conditional probabilities when simulating.

The local proportion models may be constructed by computing the proportions of domains of surrounding data within a moving window. The user must define

the size and orientation of the window. Figure 5.15 shows three cross sections for the upper sill, mean sill, sage basalt and host rock domains. The proportions were computed based on an isotropic local mobile window of 45 by 45 by 45 meters.

A heuristic method for integrating the local proportion models into the MPS framework is proposed. The method is similar to the servosystem correction technique reviewed in Section 3.3.3. The updated posterior conditional probability is based on the prior probability computed from the MPS and the likelihood information assumed as the ratio between the local proportion model and the target marginal distribution of domains, see Eq 5.5. $P^{MPS}(\mathbf{u}; k)$ corresponds to the conditional probability computed from the training image and $P^{Local}(\mathbf{u}; k)$ the local proportion at location $\mathbf{u}$; $P^{Target}(k)$ is the target proportion of domain k established by the user. The power factor γ increases or decreases the influence of the local proportion model in the updated conditional probability $P^{Updated}(\mathbf{u}; k)$. When γ is equal to zero no correction is applied.

$$P^{Updated}(\mathbf{u}; k) = P^{MPS}(k)\Big(\frac{P^{Local}(\mathbf{u}; k)}{P^{Target}(k)}\Big)^{\gamma} \qquad (5.5)$$

This updating scheme offers an accurate manner to handle local proportions models since there is no assumption among sources as in the case of Bayesian Updating Deutsch (2002), permanence of ratios or Tau model Journel (2002). These methods show a reliance on the global proportion of domains, that is, the posterior conditional probability $P^{Updated}(\mathbf{u}; k)$ is updated considering the global proportion instead of the local model as the prior, which could lead to biased proportions in the realizations.

Figure 5.16 illustrates an example of the integration process. On the left column, the same three sections showed in Figure 5.13; on the right, a realization considering a local proportion model. The use of this secondary information allows overcoming the lack of continuity of certain domains; see for instance the sage basalt (red domain) in the middle. Due to the introduction of a local model, the realization on

146
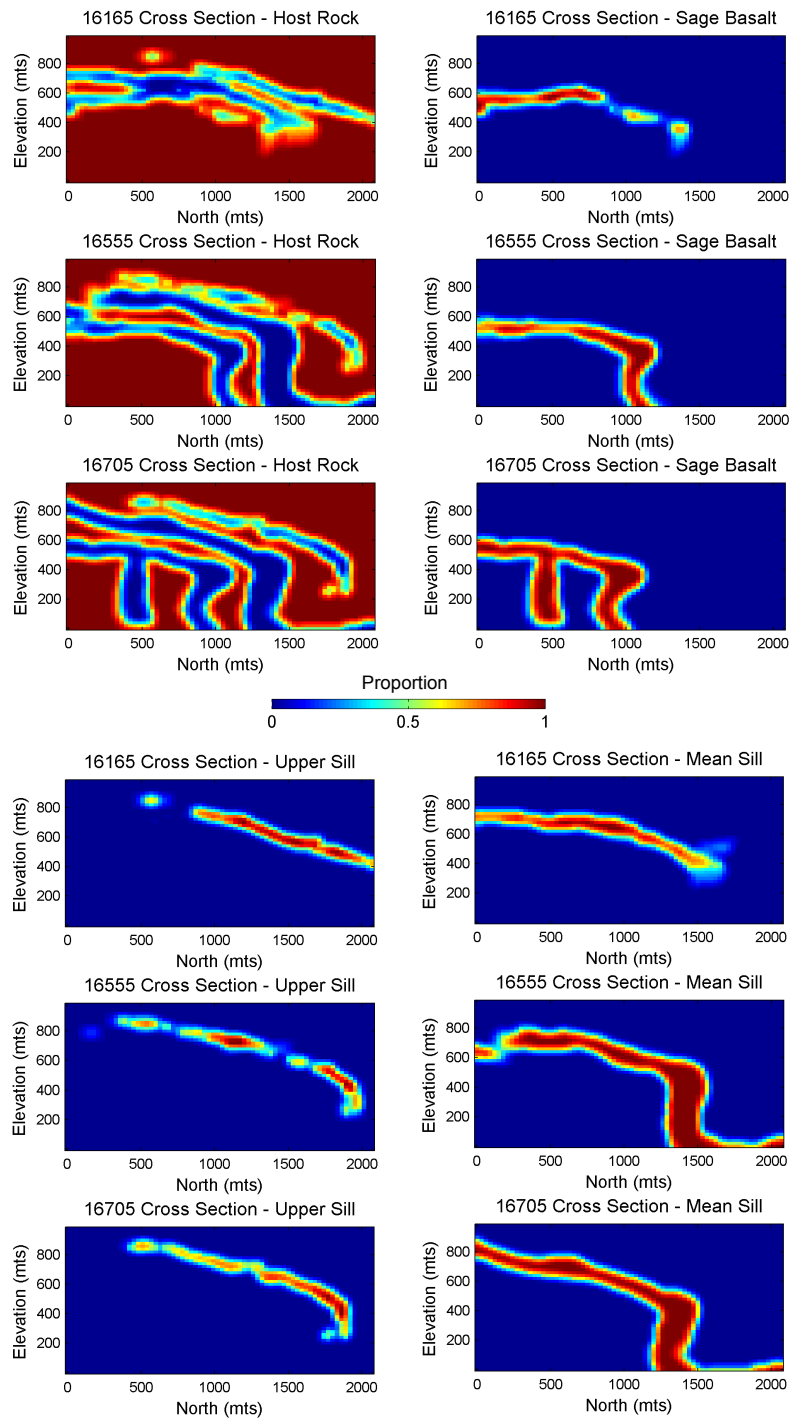
Figure 5.15: Local proportion models based on a mobile window of 45 mts by 45 mts by 45 mts. Three cross-sections show the proportions for all domains.
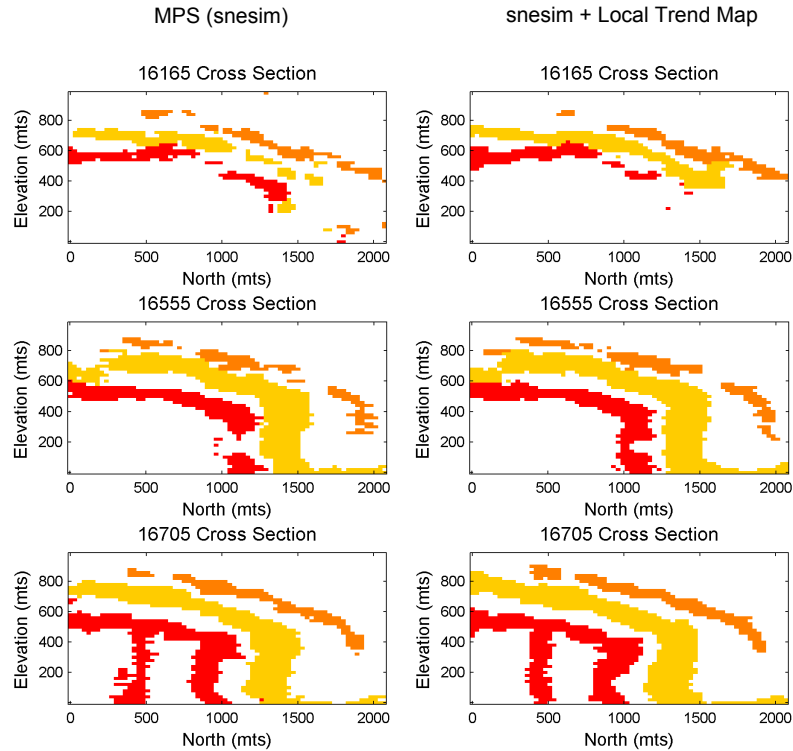
Figure 5.16: On the left, the same realization of the Figure 5.14. On the right, a snesim realization using the local proportion model of the Figure 5.15 as secondary information.

the right looks closer to the features of the training image. The correction factor has a direct consequence on the global proportions of domains; its application must be checked.

5.3.3 Multiple Point Statistics Random Perturbation (MPS-RP) Post-processing

The objective of a post-processing implementation is to eliminate inconsistencies and unwanted noise from realizations. Once snesim is performed, a second MPS simulation based on simulated annealing is implemented. The MPS annealing technique proposed by Deutsch (2011) is considered. This algorithm is straightforward and requires a low level of parameterization. The method minimizes an objective function (OF) based on the difference of three statistics retrieved from the training image and the realization: a local proportion model, two-point transition probabil-

ities and multiple point statistics.

Local Proportion Model: The first component of the objective function relates to the reproduction of a local category proportion model. Introduction of local models is relevant in the non-stationary modeling since they provide a framework to generate realizations with different proportions than the training image.

$$O_p = \sum_{k=1}^{K} \sum_{i=1}^{n_p} (n_{k,i}^{TrainingImage} - n_{k,i}^{Realization})^2 \qquad (5.6)$$

The local proportion model is divided into n_p equally spaced proportions. K corresponds to the number of geologic domains considered. Thus, the objective function ensures to reproduce the correct number of cells by classifying all the nodes of the model into one of the n_p classes.

Two-Point Transition Probabilities: This function ensures the reproduction of the K^2 two-point transitions probabilities for each lag considered. The function is given by:

$$O_{tp} = \sum_{l=1}^{n_l} \sum_{k=1}^{K} \sum_{k'=1}^{K} (p_{l,k,k'}^{TrainingImage} - p_{l,k,k'}^{Realization})^2 \qquad (5.7)$$

where l correspond to oriented lag vectors. This transition probabilities are especially advantageous when considering large distances and different anisotropies.

Multiple Point Statistics: The last component of the objective function relates to the correct reproduction of multiple point configurations in the realizations.

$$O_m = \sum_{k=1}^{K} \sum_{m=1}^{M} (p_{k,m}^{TrainingImage} - p_{k,m}^{Realization})^2 \qquad (5.8)$$

where M correspond to the number of locations of the multiple point configuration considered. Recall that the number of possible configurations is given by K^M, which depending on the complexity of the training image and the number of

149

locations may be extremely large. In order to reduce the dimension of the problem, Deutsch (2011) proposed to not consider the location of the category within the configuration, but simply observe the frequency of appearance within the M locations of each category; this reduces the dimension of the problem to KxM.

Each component of the objective function is weighted to ensure their contribution to the ultimate convergence.

$$O = \sum_{i=1}^{n} w_i O_i \tag{5.9}$$

The weights may be set up differently to provide more relevance to one particular component, but commonly, they are configured so that all component contribute equally to the objective function. Different strategies can be adopted, and additional components may be considered if desired.

Algorithm 5 MPS Random Perturbation (MPS-RP) Post-Processing

Input: snesim MPS realization
Output: Post-processed MPS realization
 1: **for** each loop. Number of times the program will visit all grid nodes in the realization **do**
 2:     **for** each node $\mathbf{u}_\alpha$ of the grid $\{\alpha = 1, ..., (n_x n_y n_z)\}$ **do**
 3:         Perturb the category k of the node
 4:         Compute each component of the objective function: O_p, O_{tp} and O_m
 5:         Accept the change if the objective function value $O = \sum_{i=1}^{n} w_i O_i$ is lesser than before
 6:     **end for**
 7: **end for**

Algorithm 5 describes the MPS random perturbation post-processing methodology. The method proceeds as follows: each node of the grid is visited, and the corresponding simulated category perturbed several times; the objective function is evaluated and the perturbation accepted if the value of the function decreases. The iteration continues visiting the nodes up to the number of loops is completed. Due to the nodes may be visited several times, the time to reach the minimum may be

considerable, especially when dealing with grid of several millions of nodes.

Figure 5.17 shows a resultant realization after the annealing post-processing is applied. The same local proportion model of Figure 5.15 is used. The annealing post processing focuses on local scale improvements by filtering the noise, making the features of the post-processed realization more similar to the training image.
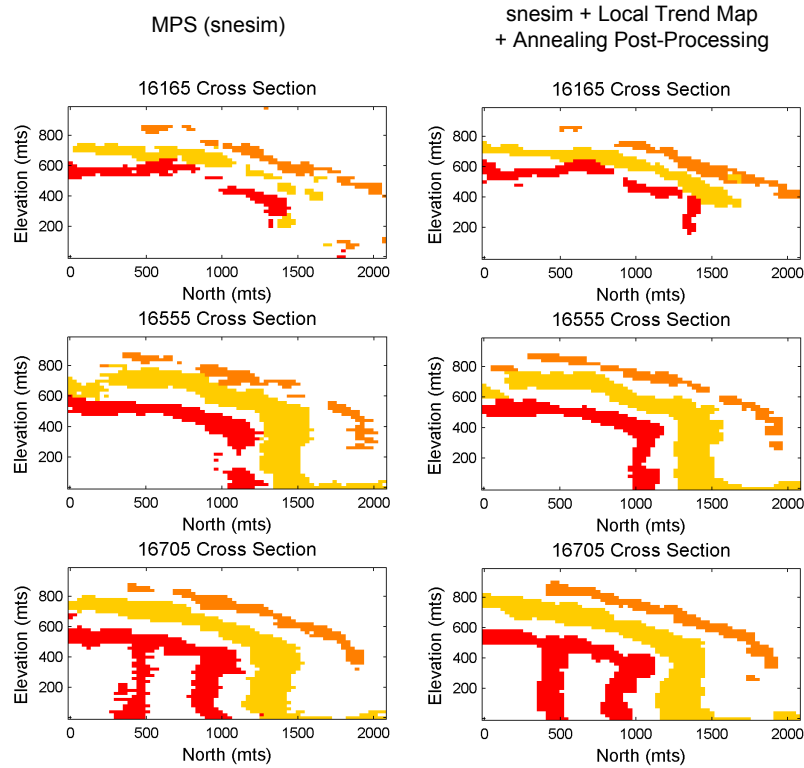
Figure 5.17: On the left, the snesim realization of the Figure 5.14. On the right, a realization considering the local proportion model of Figure 5.15 along with the application of the simulated annealing post-processing.

The annealing post-processing partially addresses unwanted noise arising in the snesim algorithm. The method is practical and easily embedded within the snesim algorithm due to its straightforward parameterization. Post-processing techniques based on the local averaging of domains have the risk of altering and filtering the essence of the multiple point configurations in the models. In this sense, the MPS annealing is statistically consistent since the multiple point configurations of the training image are still reproduced after post-precessed the realization. Besides, the

151

requirement of conditioning of data in the realizations is satisfied, that is, no change of category at drillholes locations.

5.4 Remarks

The proposed multiple training image MPS algorithm was modified to control the geologic variability to specific scales. The method incorporates flexibility in the models by allowing the user to determine the ordering and the sequencing of the reproduction of geologic features. Even though the realizations look slightly different relative to the original LOP algorithm, the contribution of the approach relies in the dynamic assignment of weights subject to a measure of information content. In the past, the characterization of the sequential path based on information from the data has shown improvements in the reproduction of large scale structures (Liu and Journel, 2004). A similar approach could be adopted, and characterize the sequential path for the local conditioning of training images. Resultant models can be extended in terms of the spatial arrangements of the patterns reproduced. The use of functions, local maps and a structured path have the potential to extend the application into different geological environments.

A local map for the assignment of the weights is introduced in the second part. This map is used as a range field to guide the use of multiple training images for non-stationary modeling. In order to generate suitable realizations, the training images have to be similar. A second algorithm was proposed to deal with training images with different heterogeneities. It is based on the construction of an individual search-tree that contains multiple point configurations from multiple training images. There are fewer unrealistic transitions. Special care must be taken when dealing with several training images with a high level of complexity; under these circumstances, the search-tree may become too large.

Finally, the last approach considers the use of local trend maps and post-processing for improving realizations of mineral deposits. The use of trend maps help to maintain the continuity in the models and the post-processing corrects un-

wanted noise. The local trend constrains the final realizations in terms of spatial variability. There is a risk the models are overly constrained.

Chapter 6

Case Study: Red Dog Mine

This chapter shows a case study based on data from the Red Dog mine, Alaska. The objective is to evaluate the forecasting of mineral resources with uncertain geologic domains through the combination of multiple data-driven training images.

Section 6.1 provides a description of the data used, the geology and the geologic domains of the Red Dog mineral deposit.

Section 6.2 reviews the implementation of the proposed multiple data-driven training image MPS simulation; while **Section 6.3** shows the procedure for calibrating the geologic variability.

The geostatistical models are evaluated in terms of accuracy, precision and prediction of the geologic domains in **Section 6.4**. In turn, **Section 6.5** presents the improvements in mineral grade prediction. Results are compared to production data.

Finally, **Section 6.6** provides remarks of the proposed methodology for modeling of mineral deposits with respect to classical geostatistical techniques.

6.1   Overview of Geology and Available Data

The Red Dog deposit consists of a number of sulphide ore zones of sedimentary exhalative and replacement origin (Ayuso et al., 2004; De Vera et al., 2004; Moore et al., 1986). Structural and sedimentary characteristics define four mineralized

plates: the upper, median, lower and sub-lower units. Four geologic domains will be modeled, all of them, contained the upper and median plates. These domains are the low iron siliceous (LIS), high iron siliceous (HIS), baritic (BA), and their associated overlying waste rock denominated as HRU.
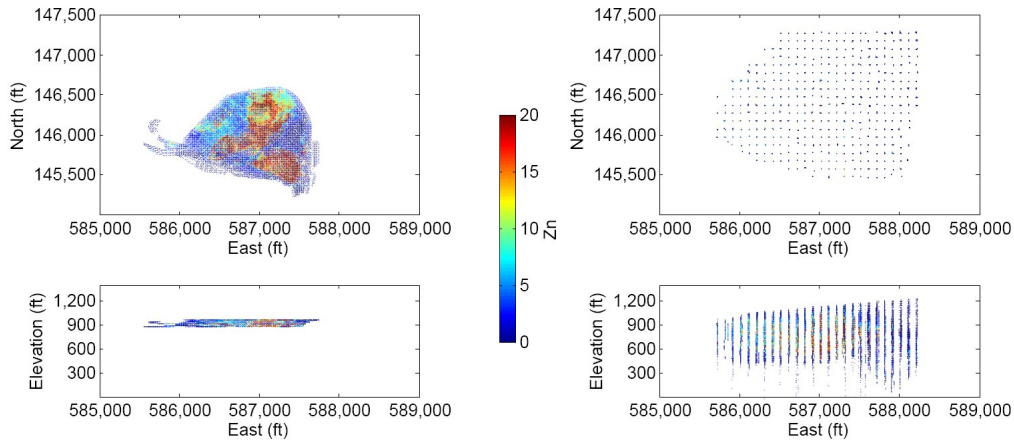
Figure 6.1: Available drillhole and blasthole data displayed on a plan view and an East-West cross section.

The case study is centered in the area where Teck Resources Limited (Teck) has provided production data (blastholes). Figure 6.1 shows projections of the drillholes and blastholes onto the North-East plan and Elevation-East section. The volume of interest is approximately 1,250 ft by 1,250 ft by 100 ft. The drillholes are composed of samples of 5 ft support spaced about 100 ft. For the blastholes, the spacing is 10 ft by 12.5 ft placed on 4 benches separated by 20 ft.

Direction	Origin Block Model	Number of Blocks	Size of Blocks (fts)
Northing (ft)	586,350	100	12.5
Easting (ft)	145,350	95	12.5
Elevation (ft)	877	20	5

Table 6.1: Dimension parameters of the block model utilized for the construction of the training images and the implementation of the MPS modeling algorithm.

Table 6.1 presents the parameters of the block model used in the construction of the training images and MPS simulation. Since drillholes and blastholes are sampled at different supports, this model is later upscaled in the vertical from 5 ft to 20 ft.

The models of Zn grades are constructed in a two-stage process. First, realiza-

155

tions of the geologic domains are generated by implementing the proposed multiple data-driven training image MPS simulation. Secondly, the Zn grades are simulated independently for each domain. Finally, the Zn grades are merged in a unique model on a by domain basis.
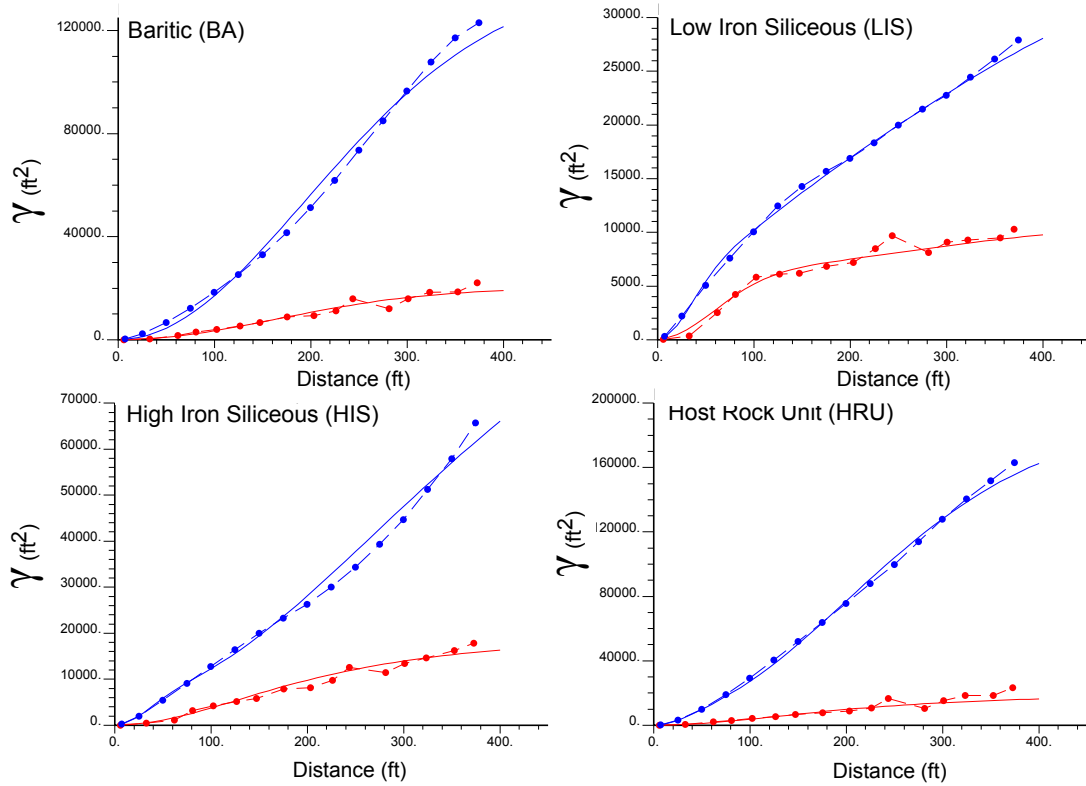
Figure 6.2: Variogram analysis of the distance function variable for each domain. In blue, the onmihorizontal experimental variogram, in red, the vertical direction. All variograms are fitted by Gaussian structures.

6.2 Implementation of MPS Simulation using Multiple Data-Driven Training Images

6.2.1 Construction of Deterministic Training Image (DTI)

Four ore type domains are modeled: low iron siliceous (LIS), high iron siliceous (HIS), baritic (BA) and the host rock (HRU). For each domain, the experimental signed distance values variograms and their models are presented in Figure 6.2. No

particular direction of continuity is observed in the horizontal plane; in red, the omnihorizontal variogram, in blue, the vertical direction. Note the difference between the vertical and horizontal directions; this is an indicative of the sedimentary nature of the mineralization. Gaussian models are used for fitting the experimental variograms.
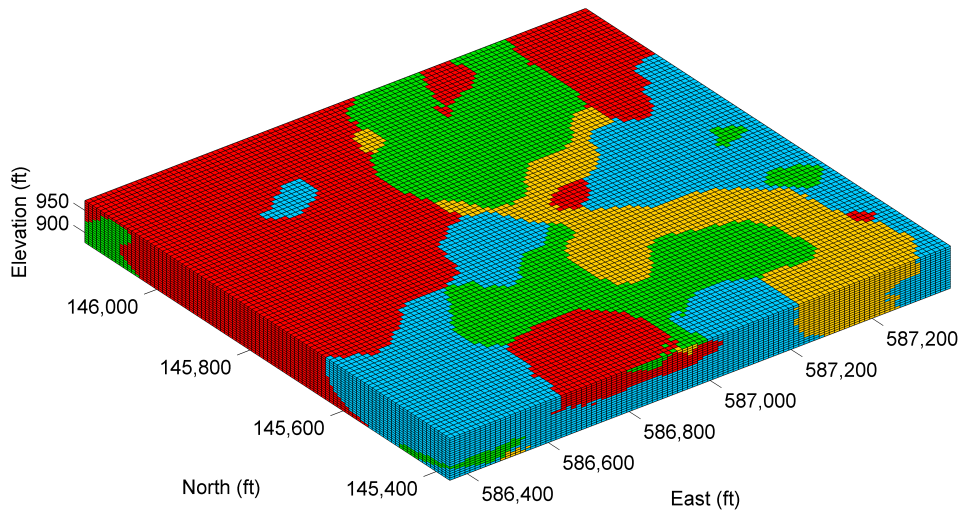
Figure 6.3: Isometric illustration of the deterministic training image. In red the baritic domain (BA), in green the Low Iron Siliceous (LIS), in yellow the High Iron Siliceous (HIS), and in cyan the Host Rock Unit (HRU).

The multiple distance function algorithm presented in Section 3.3 is used to generate the deterministic training image, see Figure 6.3 for illustration. No blast-hole information was used. In red the baritic domain (BA), in green the Low Iron Siliceous (LIS), in yellow the High Iron Siliceous (HIS), and in cyan the Host Rock Unit (HRU).

6.2.2   Construction of Stochastic Training Image (STI)

The stochastic training image was constructed with Sequential Indicator Simulation. The indicator variograms are presented in Figure 6.4. The LIS and HIS domains follows the same directions of anisotropy as the signed distance values. In the case of BA and HRU domains, a zonal anisotropy is detected at an azimuth of 90 degrees; however, it is not particularly significant. A trend is identified in the baritic

157

domain, which is depicted by the indicator variogram values above the sill. All the experimental indicator variograms are modeled using spherical structures.
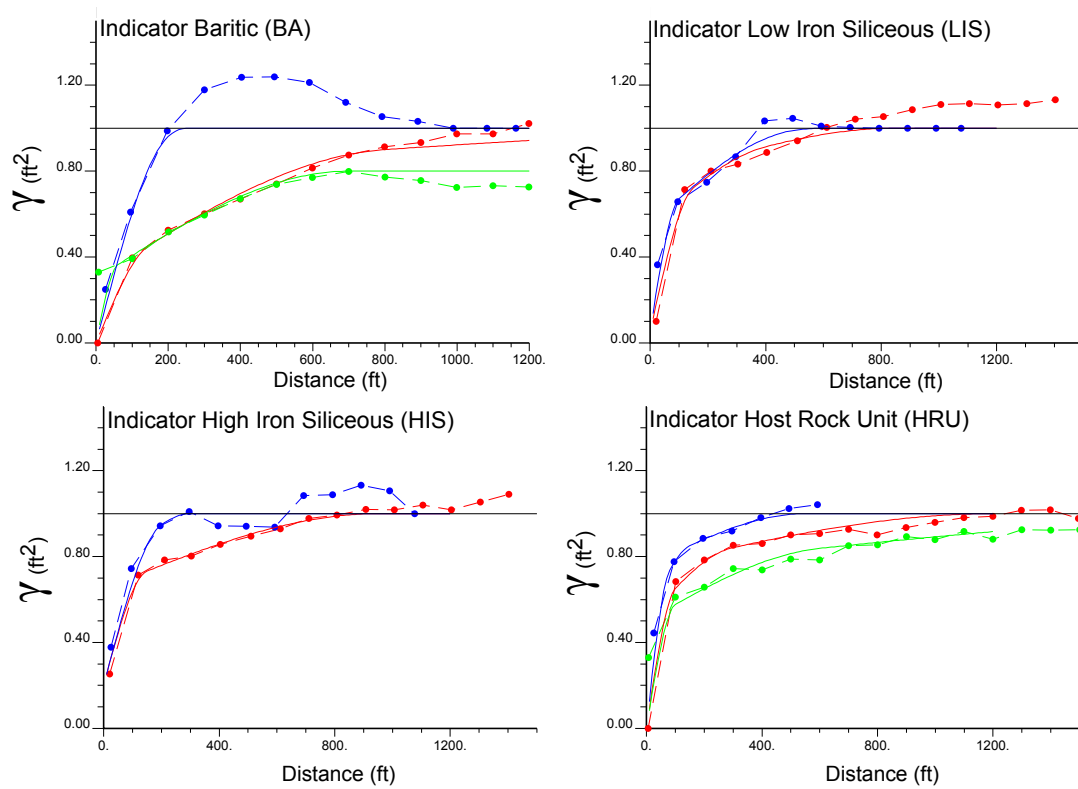
Figure 6.4: Indicator experimental variograms and their respective models. In blue and red the onmihorizontal and vertical directions for HIS and LIS. For the Baritic and Host Rock a zonal anisotropy between the N0E (red line) and N90E (green line) directions is observed.

The training image is shown in Figure 6.5. This image maintains the main large-scale structures observed in the drillhole data and the deterministic training image, but also, shows characteristics short scale variability.

6.3   Geologic Variability Calibration

This section is devoted to determining the correct weights for the integration of both training images. The proposed calibration method which includes the sampling process and the computation of the entropy are described in Section 4.5.
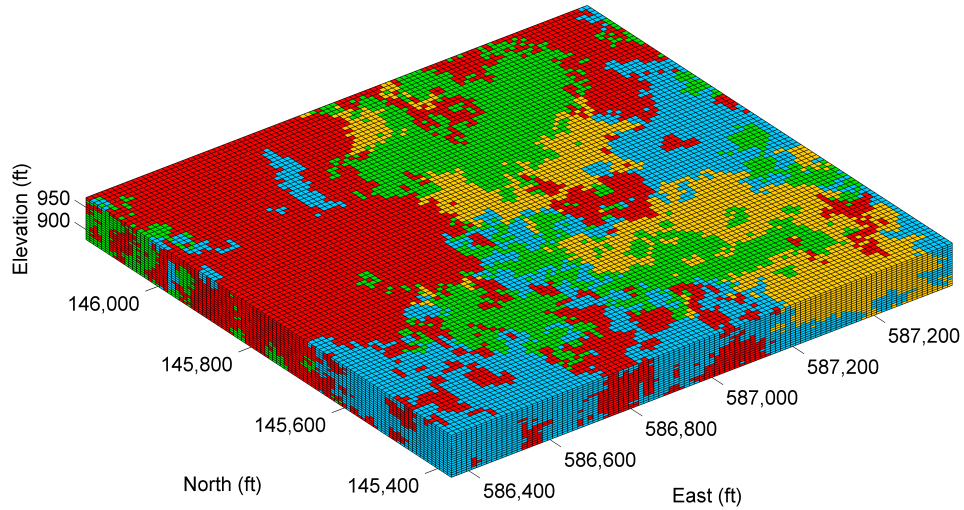
Figure 6.5: Isometric illustration of the stochastic training image. The model corresponds to an individual realization of the Sequential Indicator Simulation algorithm. In red the baritic domain (BA), in green the Low Iron Siliceous (LIS), in yellow the High Iron Siliceous (HIS), and in cyan the Host Rock Unit (HRU).

Generation of Multiple-Weighted MPS Simulations

A sequence of realizations that integrate both training images are generated. Each realization is constructed by considering weights in intervals of 5% when combining the DTI and STI training images. A total of 19 realizations were generated.

The set of realizations is presented in Figure 6.6. In general, the global geologic structures such as the spatial arrangements of domains and their proportions remain the same in all the realizations. Expected geologic structures from both training images are preserved, keeping large scale structures of the deterministic but including variability of the stochastic.

Entropy Sampling Process and Geologic Variability Matching

A sampling process is implemented as part of the variability calibration. Figure 6.7 provides an illustrative image of the sampling process for an individual drillhole. At the top, the placement of one drillhole (just at the intersection of the two panels); in the middle, the locations of its replicates. A mesh of 50 ft by 50 ft is chosen. At the bottom of the figure, a representation of the replicates of the drillhole with

159
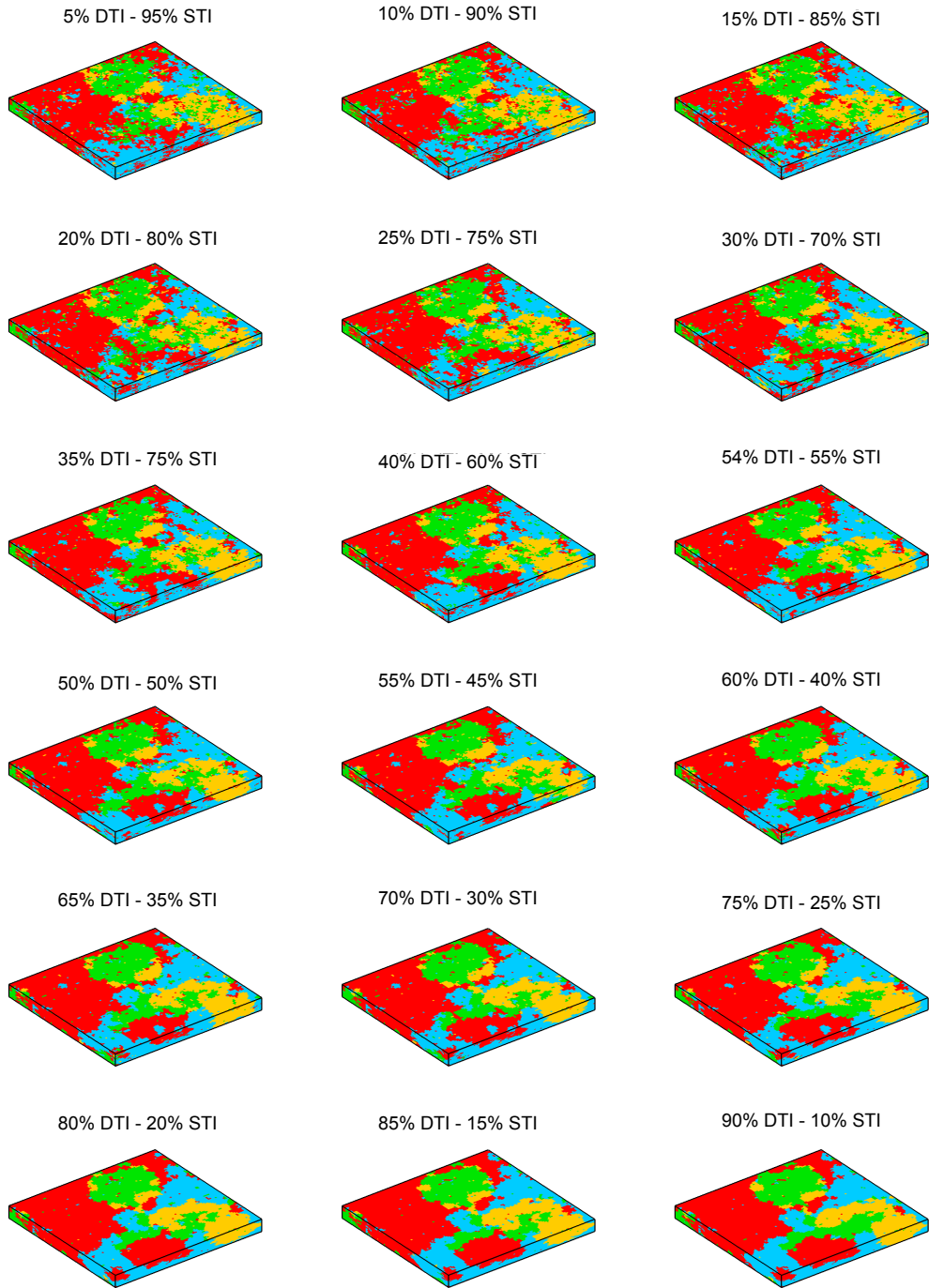
Figure 6.6: Some realizations based on the proposed MPS algorithm. The training images are combined using weights in intervals of 5%. Models present the same scale as Figures 6.3 and 6.5, that is, 1,250 fts from East to West, 1,190 fts from South to North and 100 fts in the vertical.
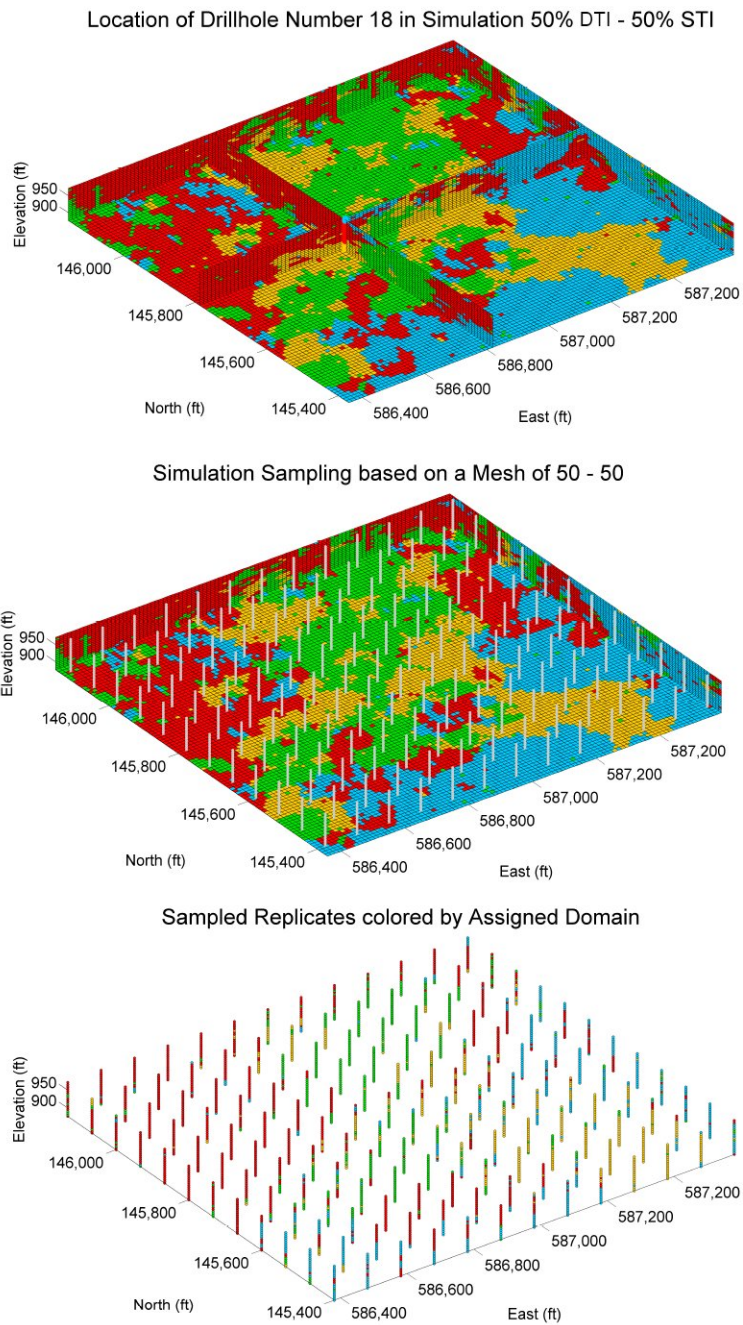
Figure 6.7: Overview of the sampling process utilized for the training image calibration. An example based on an individual drillhole is presented. At the top, the location of the drillhole. In the middle, the sampling over the training image on a regular basis each 50ft. At the bottom, The replicates resulting from the sampling process colored by their respective domains. The entropy associated to the drillhole is computed from the set of sampled replicates. The calibration ends when all drillholes have been sampled and their entropies averaged.

161

the geologic domains assigned. Then, a 1-D template of two locations is used to compute the entropy of the set of replicates and its averaged value assigned to the corresponding drillhole at the top. The process is repeated for each drillhole, and the entropy of the realization is computed as the average over the set of drillholes, Eq 4.6.

The calibration curve resulting from the 19 configurations is shown on the left of Figure 6.8. The entropy computed from the drillholes reaches 0.62, it is plotted by the blue line on the ordinate axis. Then, the calibration consists of establishing the realization that provides the same geologic variability shown by the drillholes. This configuration corresponds to 80% from DTI and 20% from STI. On the right, a simulation based on this configuration.
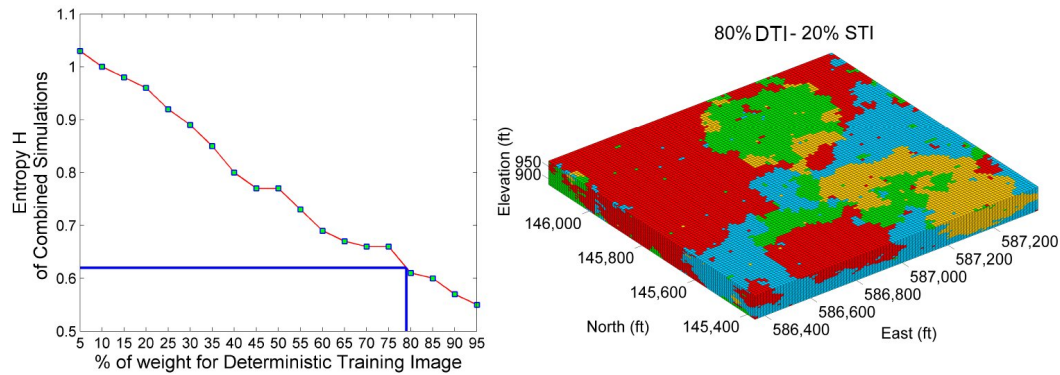
Figure 6.8: The calibration curve used for determining the correct degree of mixing of the training images. On the right, a realization whose weights were set out to 80% DTI - 20% STI, which corresponds to the same geologic variability present in the drillholes.

Comments on the Geologic Calibration Process

The calibration leads to a deposit-specific result. The assumption of stationarity is important, that is, the geologic variability of the set of drillholes is assumed spatially homogeneous in the area being evaluated. The results obtained appear reasonable but the use of local entropy could also be applied. A future path of research could be to adapt the training image integration based on local geologic conditions.
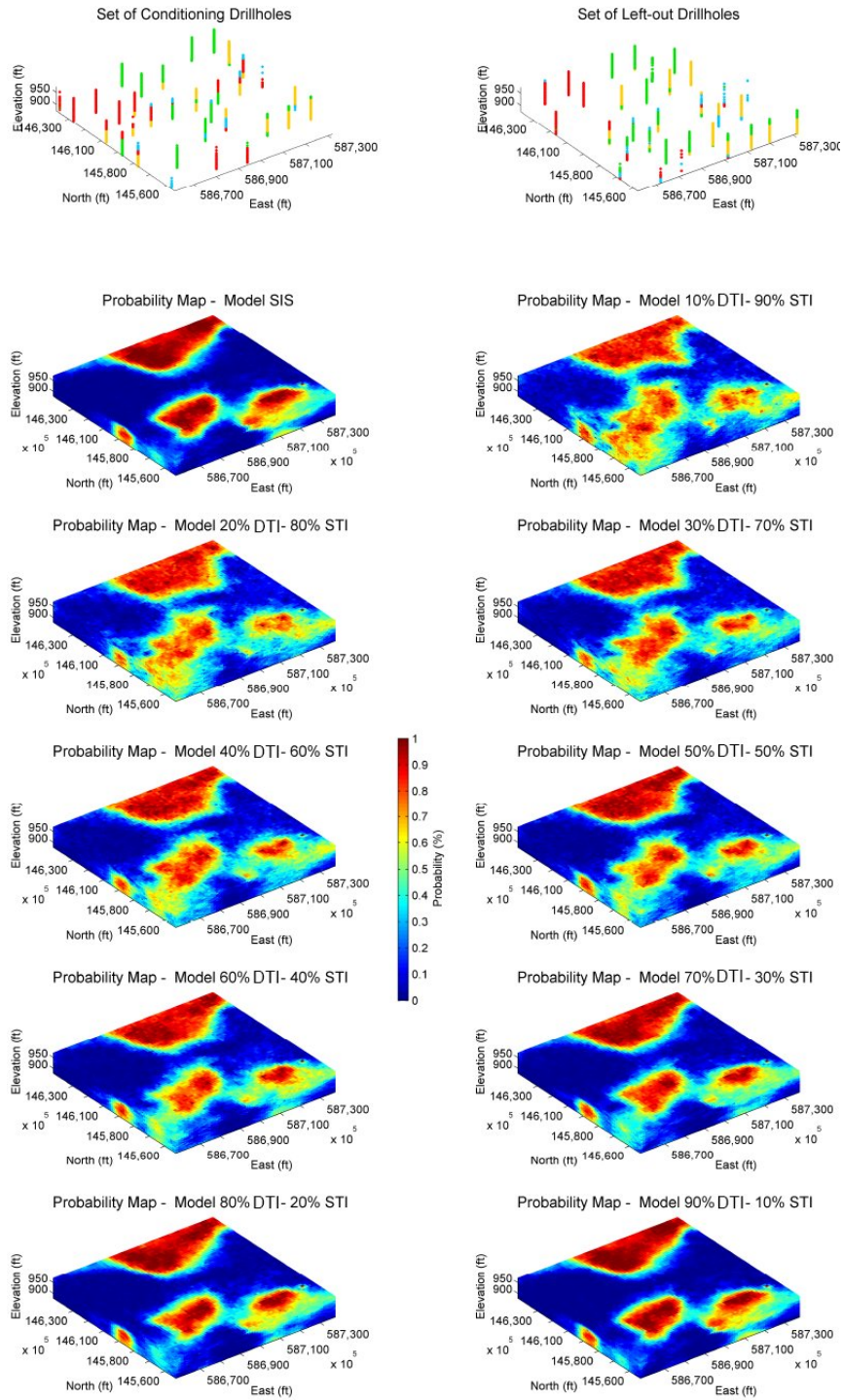
Figure 6.9: At the top, two columns representing the spatial configuration of the data used for cross-validation. Below them, 10 probability maps associated to the LIS domain for 10 different configuration of weights.
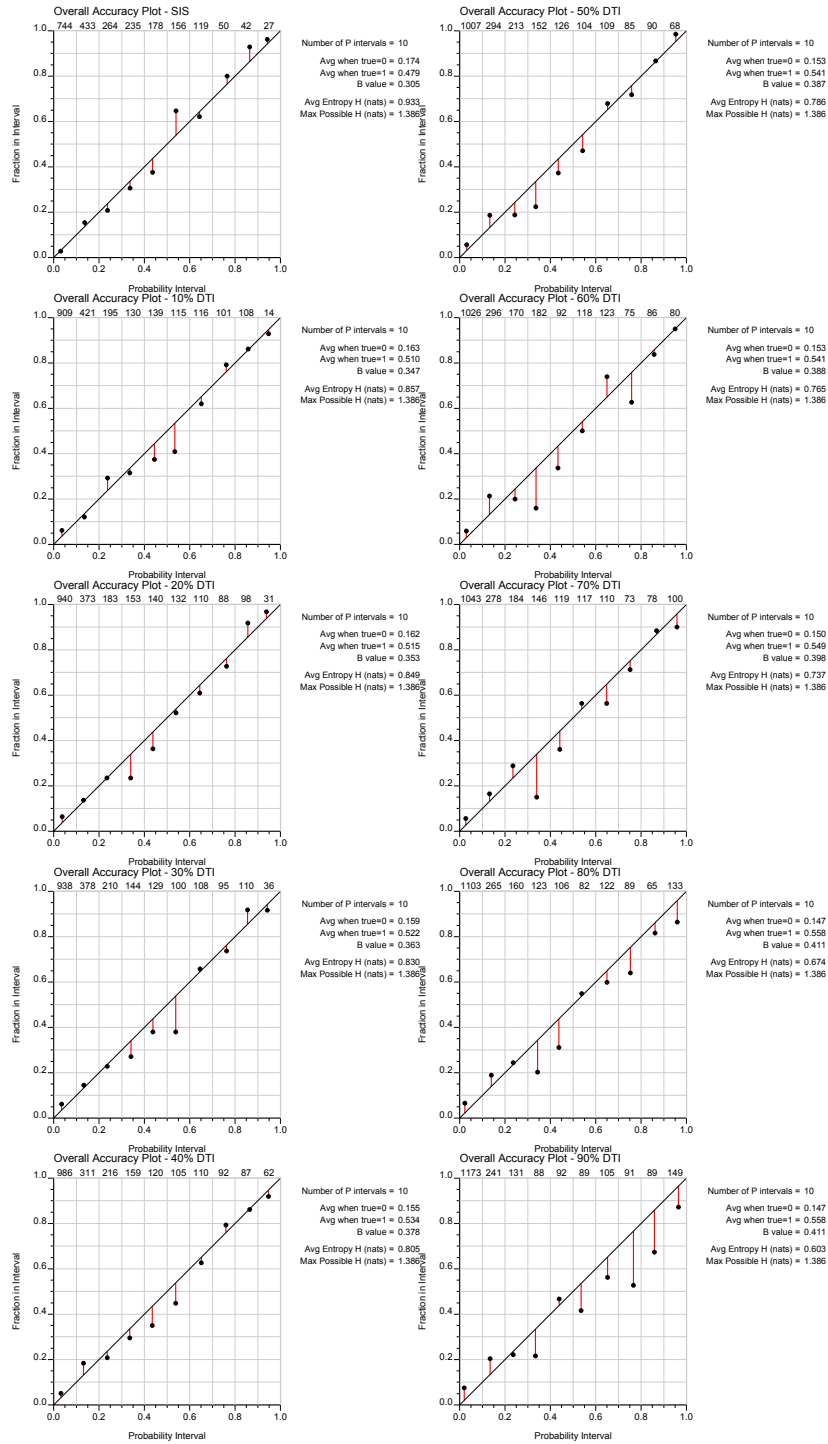
Figure 6.10: Categorical accuracy plots for evaluating the capacity of prediction of the probabilistic models. Each plot is determined by comparing the predicted domain with the set of data kept aside for validation.

6.4 Probabilistic Assessment of the Method

One aspect of checking geostatistical model is to compare the predicted values with the known data in cross-validation mode. From the set of 1,260 samples considered in the study area, 646 correspond to samples utilized as conditioning data and 614 are left-out for comparison purposes. Nine cases are evaluated, considering setting weights in intervals of 10 %. Figure 6.9 at the top shows scattered illustrations of the reference and left-out data; beneath them, the probability maps derived from 50 realizations.

The method was applied on a grid basis while the set of reference data are located at scattered positions. In order to make the results comparable, the closest nodes of the grid are assigned to the locations of the reference data.
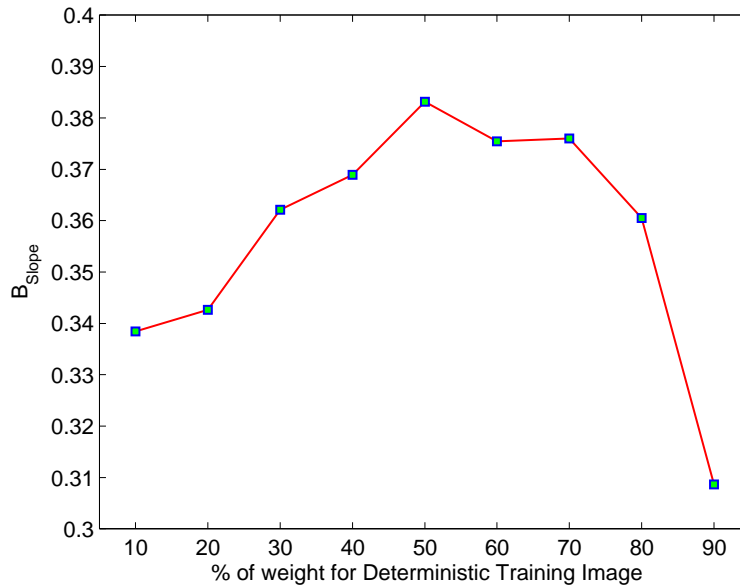
Figure 6.11: B_{Slope} curve. The optimum result is obtained when the data-driven training images are combined using the configuration of weights 50%-50%.

Local distributions of uncertainty of the reference data are assessed through the analysis of accuracy plots, see Figure 6.10. As discussed in Section 4.7, if the estimated probabilities are above the 45 degrees line, the model is accurate; the model is precise if the probabilities are close enough to the diagonal. Two issues are observed. The first relates to a systematic loss of accuracy as the weight of the

deterministic training image increases. The second corresponds to the increment of the B measure (Eq 4.9) - difference between the averaged conditional probabilities when the reference data is true and false. These issues were previously discussed and an alternative dimensionless measure proposed in Section 4.7.3.

The measure, named B_{Slope}, combines the information of the accuracy plots and the B measure, see Eq 4.10 for details. The B_{Slope} for each combined model is presented in Figure 6.11. The configuration of 50% DTI and 50% STI provides models that maximize its accuracy, precision and predict the true values correctly. These models improve the estimates of local conditional probabilities, and perform better than implementing the SIS algorithm or the deterministic geologic model individually. Even though the deterministic/stochastic approaches look geologically unrealistic, the combination of both incorporates elements that allow improving the geologic representation of the models.

6.5   Validation of the Grade Model: Measuring the Performance of Mineral Resource Forecasting

To further understand of the proposed method, the uncertainty characterization of grades of Zn must be also assessed. The objective is to demonstrate that the suitable characterization of the geologic variability in the models leads to enhanced predictions of mineral resources. The validation considers the univariate analysis of Zn grade; no multivariate relations among grades of other minerals are considered. A set of 50 realizations of Zn grades are constructed. The validation consists of comparing the realizations with the grade of blastholes (production data). A limited number of benches were used in the analysis; further, market prices and operational costs do not correspond to those used by Red Dog mine personnel.

6.5.1   Geostatistical Simulation of Zn Grade

Sequential Gaussian Simulation (SGS) is used to simulate the Zn grades. Further theoretical and practical insights of the methodology may be found in Chilès and

Delfiner (1999); Deutsch and Journel (1998). The method relies on the assumption that the data may be fully characterized by the Multivariate Gaussian distribution. Commonly, the grades do not present a Gaussian distribution so a normal score transformation is applied first. The normal scores values are used for variogram modeling. The data are simulated in Gaussian units and then back-transformed to original values.
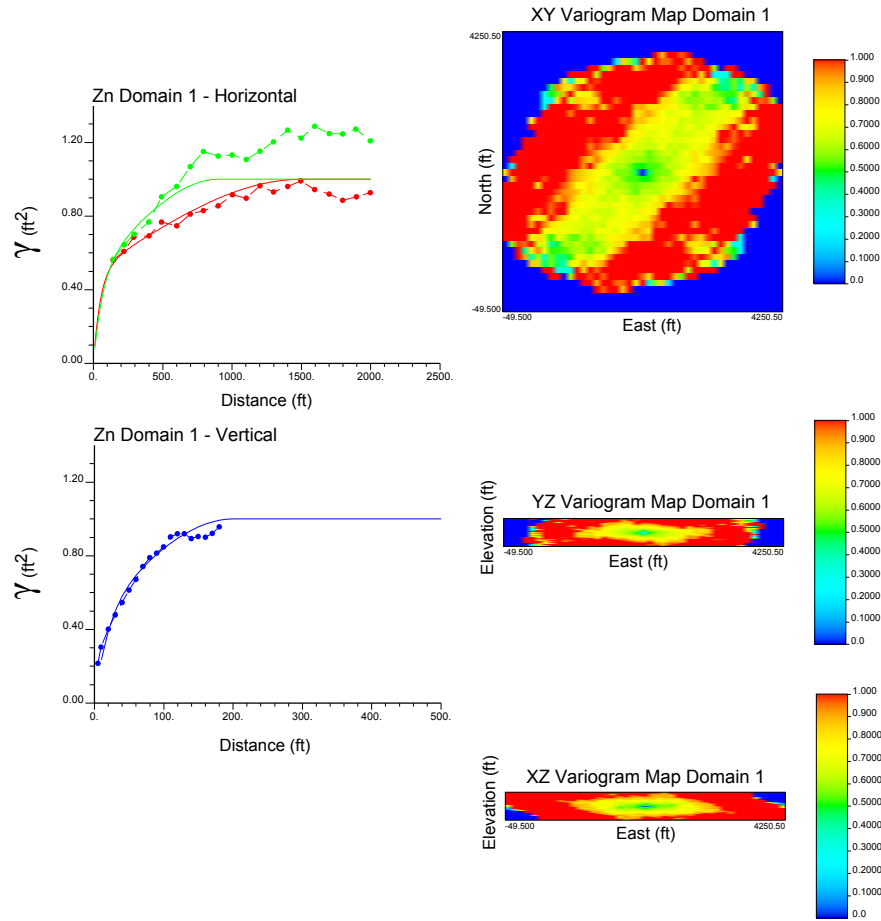
Figure 6.12: Variogram analysis for Zn grade. A zonal anisotropy at N45E with a slight dipping of 10 degrees is detected from the variogram maps. The experimental variograms were fitted by exponential and spherical models.

Normal Score Transformation

Normal score transformation is a one-to-one relation that consists of converting values in the original distribution of Zn grades into a standard Gaussian distribution.

Depending on the distribution of original units, a despiking procedure may be needed in the case of multiple data with the same values (below detection limit grades for example).

Variogram Modeling of Normal Scores Values

Figure 6.12 on the right presents the variogram map for the planes XY, YZ and XZ. A strong zonal anisotropy is observed along the azimuth $45°$ dipping at $10°$. The experimental and model variograms are plotted on the left. In red and green, the directions of $45°$ and $135°$, respectively; in blue, the vertical variogram.

Parameters for Gaussian Simulation

The realizations are generated using the GSLIB program *sgsim*, Deutsch and Journel (1998). The set of parameters required for the simulation relate to defining the dimension of the search and the number of samples used. The selected search radii was set up to 250 x 250 x 100, and the minimum and maximum number of samples to 8 and 64 respectively.
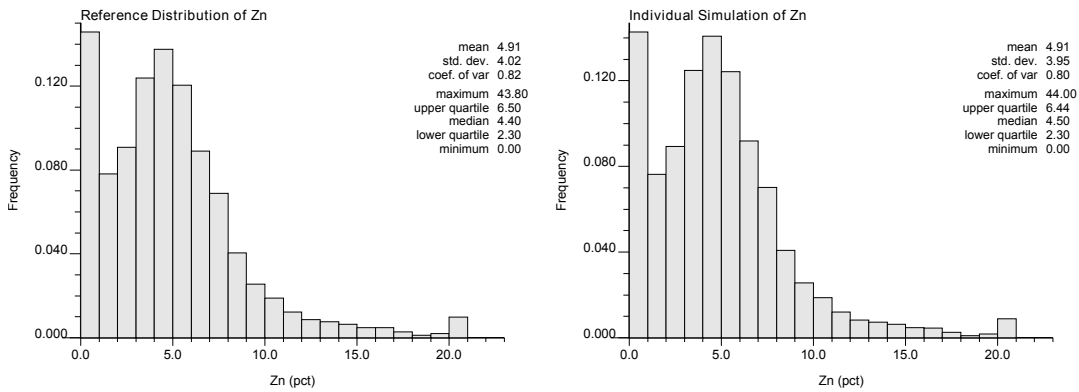
Figure 6.13: Histogram reproduction checking of Zn grades. On the left, the reference distribution; on the right, the histogram of an individual realization.

Validation of Simulated Zn Grades Models

Non-stationarity issues and the incorrect selection of parameters might generate biased realizations. The realizations must reproduce the conditioning data as well as

168
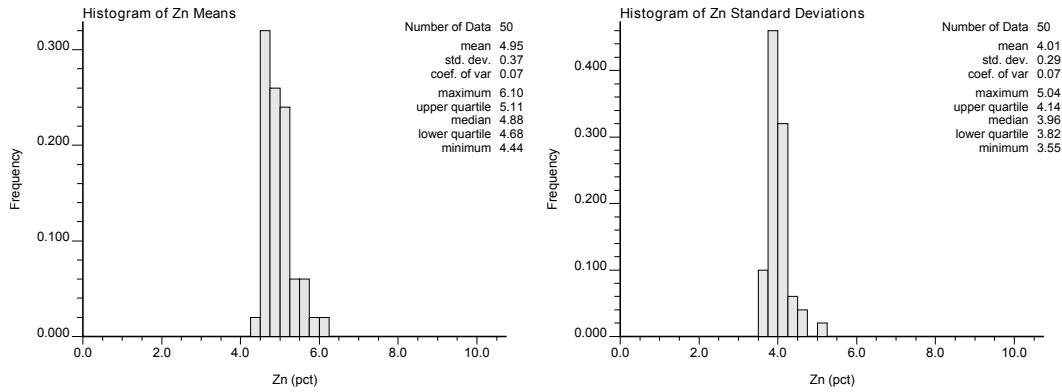
Figure 6.14: Histograms of the mean and standard deviation for all realizations.

the histogram and variogram. These checks are standard practice. The data are reproduced exactly.

Reproduction of Histogram: Histograms of simulated values after back transformation should be close to the reference distribution of conditioning data. Figure 6.13 compares the reference distribution with one obtained from an individual realization. Both the shape of the histograms and its basic statistics including the mean and the standard deviation are well reproduced. A more general comparison is shown in Figure 6.14. Here, the histogram of the means and standard deviation for the set of realizations show that the distribution of the conditioning data is reproduced in the realizations.

Reproduction of Variograms: Experimental variograms are computed over the set of realizations for the same directions found in the exploratory analysis of data. Variogram reproduction is checked in Figure 6.15. Three directions of continuity are plotted; in red, the variograms of each realization and in black the corresponding model for that direction. A slight bias may be observed for the first and the last figures, but the effective reproduction of the variogram within expected deviations.

169
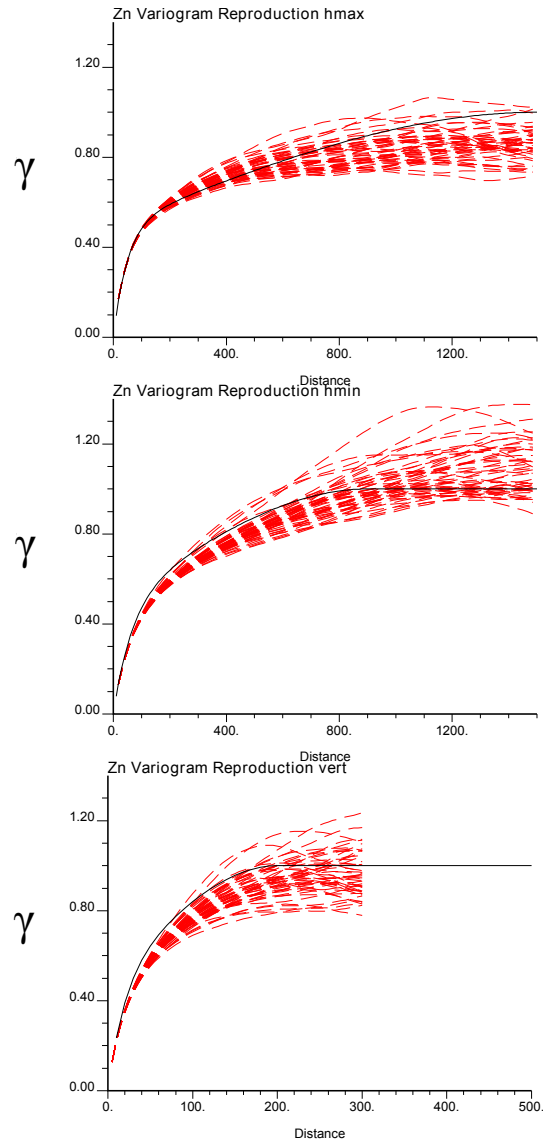
Figure 6.15: Variogram reproduction checking. The major, minor and vertical directions are considered. The set of red dash lines correspond to the experimental variograms of the realizations; the model is presented by the continuous black line.

6.5.2 Local Mineral Resources Forecasting Performance

The analysis focuses on assessing the local prediction of Zn grades. The simulated grades are compared with the grades of blastholes. Two measures are implemented, the correlation coefficient, and the accuracy and precision of the estimated conditional distributions obtained from the set of Zn realizations.

The grades are originally simulated using the block model parameters described

170
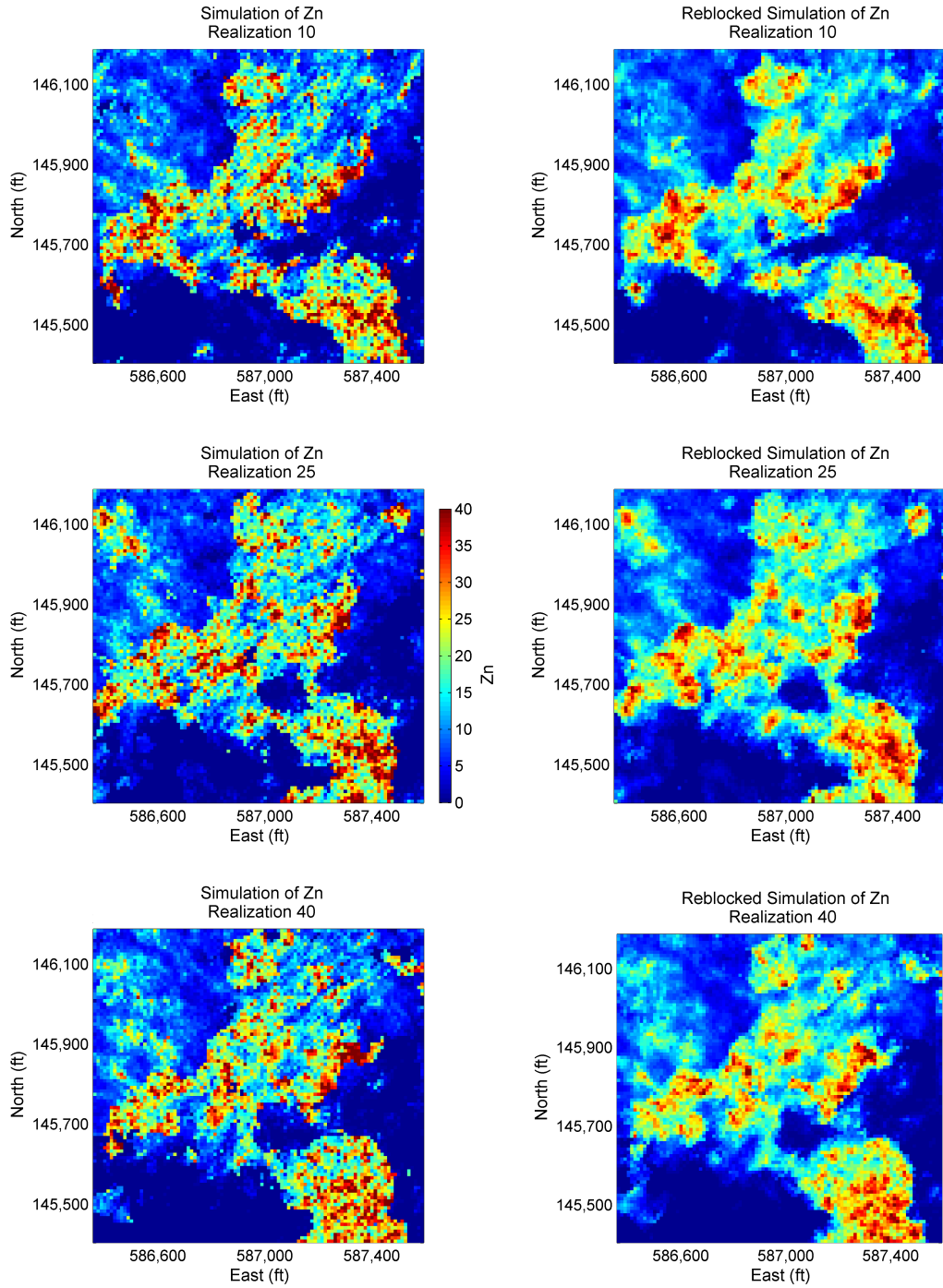
Figure 6.16: Three realizations of Zn grade for an individual bench. On the left, the model at a resolution of 12.5ft by 12.5ft by 5ft; on the right, the upscaled model at 12.5ft by 12.5ft by 25ft (same scale that the blasthole data).
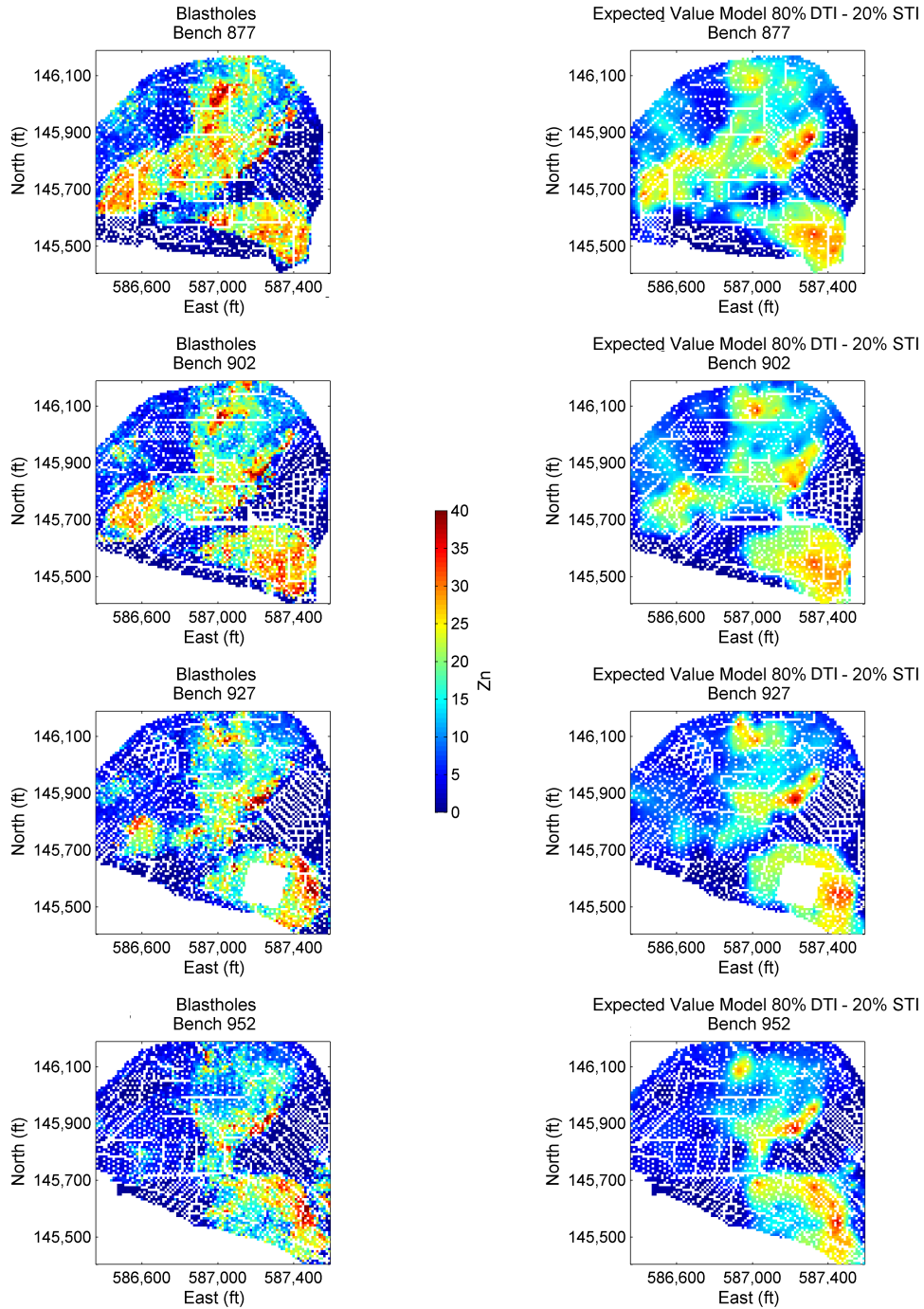
Figure 6.17: Blastholes colored by Zn grade. On the left, the original values; on the right, the expected value of Zn of all realizations. For the latter, the simulated values in the grid were paired to the closest blastholes.

in Table 6.1. The calibrated model of 80 % DTI - 20% STI is implemented. The column on the left of Figure 6.16 illustrates three realizations for the bench at level 877 ft. The model is upscaled from the sample to the blastholes support, that is, the realizations are upscaled from (12.5ft,12.5ft,5ft) to (12.5ft,12.5ft,25ft) support. The column on the right shows the same upscaled 877ft bench. Note that the variability of the Zn grades decreases due to the upscaling in the vertical direction.

Correlation of Simulated Zn Grades and Blastholes

The blastholes are commonly drilled on a regular mesh independent of the block model mesh, both the realizations and the blastholes are paired according to nearest neighbor. A total of 21,347 blastholes distributed on four benches have been considered for the analysis. The column on the left of Figure 6.17 displays the spatial configuration of the blastholes. The column on the right corresponds to the average of grades of Zn over the set of 50 realizations.
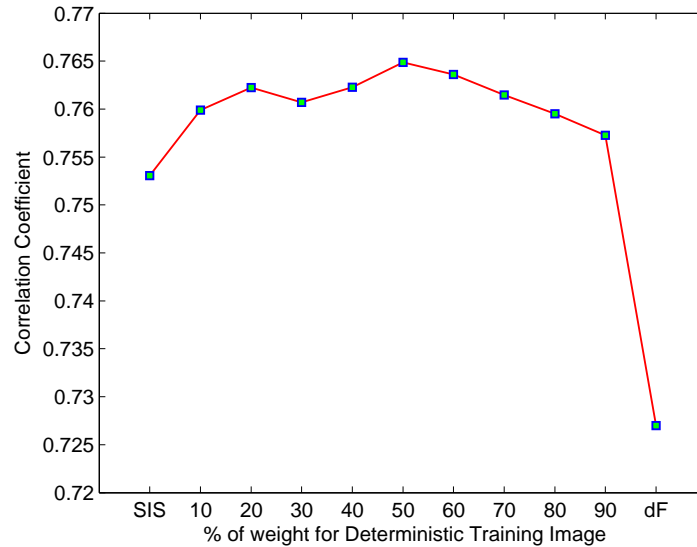
Figure 6.18: Correlation coefficient between the blastholes of the Figure 6.17 and the set of expected values of the set of realizations.

Given that the set of realizations offers 50 simulated values to be compared against the blastholes, their expected value is used as an unique representation that summarizes the information contained in the realizations. The correlation coefficient

between the blastholes and averaged Zn grades is presented in Figure 6.18. At the extremes there are also plotted the coefficients for the cases when SIS and the deterministic model are implemented individually. The maximum correlation corresponds to the configuration of 50% DTI - 50% STI. It is interesting to note that the combined models present better correlations than considering the SIS and DTI models individually. The increase of the correlations coefficient confirm that combining multiple point configurations from two training images allow a relative small improvement in the local prediction of mineral resources.

Local Accuracy and Precision of Models

Another measure of local performance relates to the goodness of the local uncertainty distributions of the probabilistic models of grades of Zn. The same definition when evaluating a categorical model, detailed in Section 4.7.1, applies, but for the continuous variable Zn. Recall that a statistical model is deemed accurate if the fraction of true values falling within a given p symmetric interval exceeds or is equal to p. Precision, in turn, relates to the closeness to the p interval; as the fraction of true values equals p, the model is considered more precise.
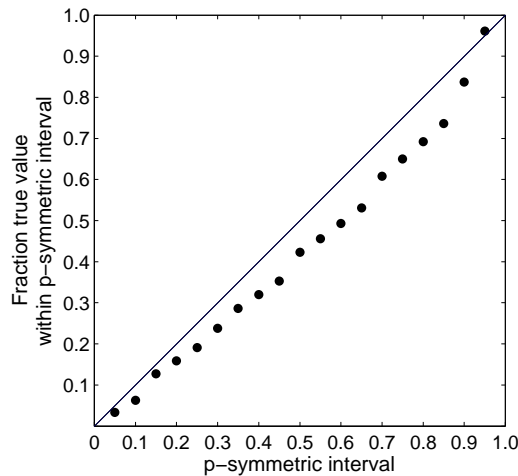
Figure 6.19: Accuracy plot of Zn grades.

174

A scatter plot using p symmetric increasing interval of 5% against the true fraction of values is presented in Figure 6.19. The difference in average is around 7% being close enough to consider that the Zn grades were estimated with accuracy and precision.

6.5.3 Assessment of Global Forecasting of Recoverable Resources

This subsection evaluates the ability of the proposed method for predicting mineral recoverable resources in global terms.

The ore tonnage and grades is crucial for decision-making. A series of operational and economical parameters are involved including the grade of zinc, the metallurgical recovery, the price in the international market for one tonnage of zinc, and mining the cost associated to the extraction process for one tonnage of material. All these elements are embedded in a profit-oriented function to set the minimum grade (marginal cut-off) needed for an estimated block of Zn to be considered profitable. The marginal cut-off corresponds to the minimum value for the grade of Zn from which one may obtain zero revenues. Then, all material above this cut-off may be considered as economically profitable. To maintain the reserve of the economic and operational parameters used in the Red Dog mine, the cut-off grade was arbitrarily established to 15% of Zn.

The marginal cut-off is applied over the set realization individually; then, a distribution function of the global amount of ore is obtained.

The distribution of global recoverable resources that result from the calibrated model (80% DTI - 20% STI) is evaluated. To complement the analysis other two models are added, the one based on SIS and that that uses the geologic interpretation provided by the DTI model. For the latter the geologic model remains unique. The Figure 6.20 provides the distribution of global mineral resources in terms of quantity of metal for each case; fitted distributions curve and general statistics are also plotted as a reference. To facilitate the evaluation, the red line plots the mean of the distribution and the blue the content of metal computed from the blastholes. The calibrated model is considered accurate since the mean of the distribution is
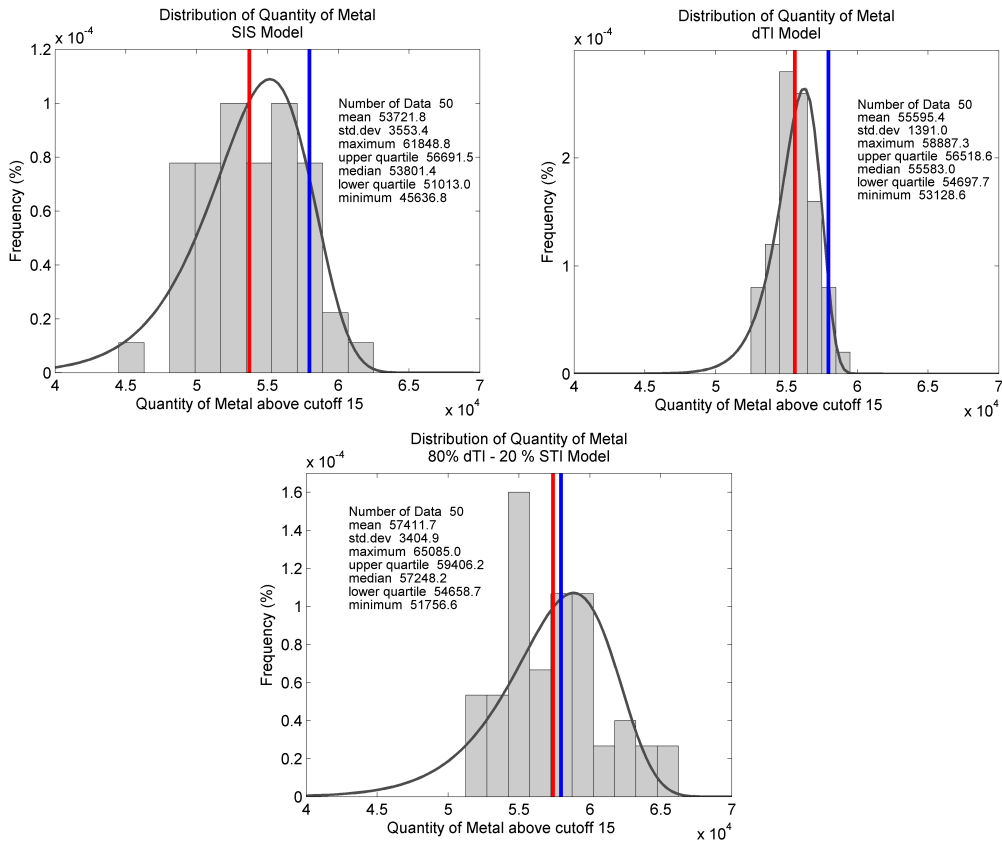
Figure 6.20: Distributions of the quantity of metal for the SIS, DTI and 80% DTI - 20% STI models. The blue line represents the quantity of metal computed directly from the blastholes. The red line corresponds to the mean of the distribution. As close these two lines more accurate the model.

close to the amount of metal obtained from the blastholes, but also precise due to the distributed values are fairly concentrated around the mean. The model based on the DTI interpretation cannot be assessed for accuracy. The SIS model appears neither accurate nor precise.

6.5.4 Economic Impact on Decision-Making

The last assessment consists of attempting to quantify the economic impact of the proposed methodology. The ore-waste classification of the blastholes after applying the marginal cut-off is compared with conventional mineral resources evaluation approaches. Since all models have the same simulated grades, the ore-waste classification is only affected for the geologic variability of simulated domains. Figure
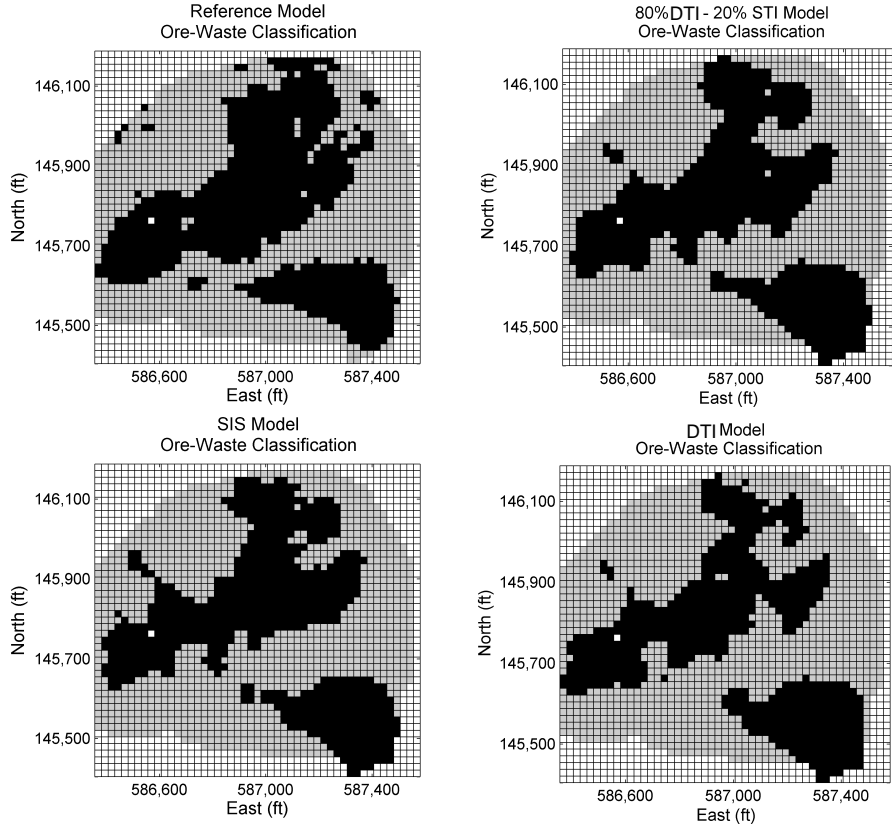
Figure 6.21: Ore-waste classification maps for SIS, DTI and 80% DTI - 20% STI models. The true classification based on blastholes (reference model) is also plotted on the top left corner.

6.21 displays the ore-waste classification for an individual bench; in black the blocks considered as the ore, and in gray those send to waste dumps. Note how the 80% DTI - 20% STI model inherits geologic features from DTI and STI that results in an improved version of ore-waste classification.

| | Sisim | | Model 80%-20% | | DTI | |
Classification	Ore	Waste	Ore	Waste	Ore	Waste
Ore$_{Blastholes}$	1,863	757	1,917	703	1,833	787
Waste$_{Blastholes}$	397	4,336	412	4,321	403	4,330

Table 6.2: Ore-waste classification results for SIS, DTI and 80% DTI - 20% STI models.

The Table 6.2 provides a quantification of the classification process. The comparison is based on the classification of the true model relative to the number of

blocks classified as ore and waste for each methodology. The misclassification of material for the SIS and DTI methods resulted in 15.7% and 16.2% respectively; a lesser percentage of 15.1% is achieved when the proposed method is applied. Another important measure consists of the percentage of ore correctly predicted. From the total of true profit, the SIS approach forecasts adequately 71.1%, the DTI 69.9%, while the MPS proposed method predicts 73.2%. Depending on the area evaluated, the grades of Zn involved and the metallurgical treatment this 2% of difference may translate to millions of dollars in revenue. In this case, the improvement of 2% amounts to an improvement of $14M and $13M in the predicted profit (based on current Zn price) of the proposed methodology relative to the DTI and SIS approaches.

6.6   Remarks

The MPS integration of data-driven training images leads to an improvement in the classification of ore and a reduction of misclassified material. A comparison of the multiple training image MPS approach relative to classical geostatistical implementation approaches show an improvement in profit. The technique integrates information in a manner that better captures the intrinsic geologic variability of the deposit. Improved mine design and planning may be undertaken by mining engineers.

Chapter 7

Concluding Remarks

Resource and reserve estimation of mineral deposits requires high resolution models of the spatial configuration of geologic domains. A realistic representation of the variability of these domains may not be obtained either by currently used geologic deterministic method or recent developments in geostatistical simulation. Deterministic models are overly continuous, whereas geostatistical approaches are characterized by a lack of large-scale continuity and an exaggerated random component. These issues lead to biased predictions on the evaluation of mineral resources (Richards and Sides, 1991; Srivastava, 2005).

Geostatistical techniques can be improved upon. This thesis proposes a methodology that improves the realism of the geologic contact variability in mineral deposits. Numerical models are constructed based on a MPS simulation approach that integrates two geologic interpretations represented by multiple data-driven training images. The resulting models lead to an improvement in modeling the geologic variability, providing of better estimates of dilution and lost ore, and an improved forecasting of mineral resources and reserves.

7.1 Summary of Contributions

The main contribution of this thesis consists of a methodology for the correct modeling of the geologic variability in mineral deposits. Specifically, the original contri-

butions include: (1) a procedure for constructing data-driven training images, (2) the integration of these multiple data-driven training images based on the linear opinion pool scheme, and (3) the calibration of the geologic variability observed in conditioning data.

7.1.1 Construction of Multiple Data-Driven Training Images

The use of training images derived from available information is an essential element of the proposed methodology; the introduction of conceptual geological models is avoided. One image comes from modeling the categories by a deterministic approach and another comes from the application of stochastic method. The first is too continuous and the second too random. Both training images represent interpretation of the mineral deposit; the difference relies in the principles assumed for their construction.

In the case of the stochastic training images, the sequential indicator simulation was used; its theoretical and practical aspects have been widely documented (Chilès and Delfiner, 1999; Deutsch and Journel, 1998; Goovaerts, 1997; Journel, 1983; Journel and Isaaks, 1984), and do not constitute a contribution of this thesis. Its relevance relies on the resultant characteristics of these models when integrated with the deterministic training image.

On the contrary, for the deterministic training image, an implicit geologic modeling approach based on the estimation of signed distance values was developed. The method provides a quick manner to generate deterministic geologic models and training images. It is straightforward and does not require heavy parameterization; it is reasonably objective, works with multiple geologic domains simultaneously, reduces the time for interpretation and alleviates the workload when doing auditable revisions. It allows geologists to focus on understanding the geological process instead of tedious manual digitalization of cross-sections and planviews. Introduction of the *softmax* transformation function allows obtaining a heuristic representation of the uncertainty of the boundaries. This heuristic measure is the basis for the implementation of a post-processing technique that corrects the global proportions

of the geologic domains. These modifications provide flexibility and consistency in the resulting models.

7.1.2 Integration of Multiple Data-Driven Training Images

The proposed methodology integrates information from the training images based on the linear opinion pool scheme. The theoretical support relies on its convex property, that is, the combined probability falls inside the range of probabilities computed individually from each training image. The convexity ensures the construction of a mixture model with the correct amount of continuity and variability. The linear opinion pool is a straightforward and an appropriate framework to combine probabilities in our context of finding a reasonable geologic variability.

The MPS integration of data-driven training images reduces misclassified ore/waste material. A mining case study served as the basis to compare the performance of the method relative to classical geostatistical implementation approaches. Results show an improvement in the predicted profit. The technique integrates information in a manner that models honor the intrinsic geologic variability of the deposit. Then, improved decisions related to mine design and planning may be undertaken by mining engineers.

Although the algorithm was applied for training images constructed using variogram-based techniques, the methodology is extensible to training images of different nature. The deterministic training image might be created, for instance, from the classical hand-drawn approach, some basic geologic sketch, or an object-based technique. On the other hand, any geostatistical method may be used for generating the training image with larger variability. The combining process may be put into practice with any class of training images. The integration of conditional probabilities by a weighted linear function leads to combined features from both training images.

7.1.3 Calibration of Data-Driven Training Images by a Measure of Geologic Variability

The calibration procedure of training images is an important contribution that aims to generate geologic models with a similar variability as shown by the conditioning data. The weighting of the two training images is established by calibrating the weights to match the multipoint entropy of the realizations and the drillholes. There is no need for a conceptual training image from another modeling technique. Resulting realizations reflects the degree of spatial disorder found in the drillholes.

The entropy is used to characterize the spatial arrangement of geologic domains. This measure serves as a structural tool for calibrating mixture model when integrating multiple training images. A major issue arises when dealing with two/three-dimensional training images, and the limited one-dimensional information of the drilling campaign. As a solution, a sampling process allows assessing the entropy of the training images and calibrating the weights to match the entropy of the drillholes. The idea takes advantage of the abundant drillholes involved in most mining projects. Even though the result depends on the particular arrangement of drillholes, its simplicity makes it an efficient measure.

Analysis on multiple point distributions related to expected geologic features in the models after combination has shown the benefits of the convex property. Realizations are balanced within the range of variability established by the training images. The linear combination guarantees generating realizations with an adjustable amount of geologic variability.

Reproduction of the geologic variability depends on the randomness of the stochastic training image, which in turn, corresponds to a realization of sequential indicator simulation. The use of multiple realizations as training images was also considered. Further analysis showed minor differences in terms of the reproduced patterns. The use of only one stochastic training image is acceptable to produce the target level of geologic variability in the realizations.

The method is a step forward in the generation of models that incorporate multiple sources and add realistic geological elements to the models. Resulting

simulated realizations have a level of determinism that is warranted by the data and provides a model of uncertainty that is checkable and reliable for decision making

7.2 Areas of Applications

Besides the analysis on the reproduction of the geologic variability of mineral deposits, the idea of integrating multiple training images may be extended. Two areas of applications emerged throughout this thesis: (1) the use of the LOP to model non-stationary geological environments, and (2) modifications to MPS simulation to use local conditioned geologic model as a training image.

Integration of Multiple Training Images for Modeling Non-Stationary Geological Phenomena: The integration of multiple training images based on the LOP combination scheme has the potential to provide an alternative for non-stationary geologic modeling as demonstrated by Manchuk and Deutsch (2012). The implementation of local maps that controls the integration of multiple training images is an important benefit of the LOP in this context.

The similarity between training images is fundamental for a suitable construction of non-stationary models. To provide a method that does not need similar training images, a more generalized algorithm is proposed. It is based on the construction of an individual search-tree that contains multiple point configurations of the training images. Here, the training images are loaded into the same search-tree before the computation of conditional probabilities. Results showed a significant reduction of artifacts when training images are coupled within non-stationary transition zones.

Practical MPS Simulation Using Data-Driven Geologic Models as Training Image: Applications of MPS methods for modeling mineral deposits based on data-driven training images is not common practice due to non-stationarity issues such as lack of continuity and unwanted noise in realizations. Practical modifications as local trend maps and MPS post-processing mitigate these issues.

Local proportions models provide of secondary information to deal with the lack of continuity. The models establish a level of knowledge about the presence

and absence of particular domains in the deposit, and are embedded into the MPS framework by a heuristic method. The method is similar to the servosystem correction technique. This updating scheme offers an accurate manner to handle local proportions models since there is no assumption among sources, and show a reliance on the local proportions. However, there is a risk of generating models overly constrained.

Inconsistencies and unwanted noise are eliminated from realizations by introducing a second MPS simulation based on simulated annealing. The annealing postprocessing partially addresses unwanted noise produced when using snesim based on local conditioned training images. The method is practical and easily embedded within the snesim algorithm due to its straightforward parameterization. The MPS annealing is statistically consistent since the multiple point configurations of the training image are still reproduced after post-precessed and does not present problems about conditioning data.

7.3 Future Work

Future applications are grouped based on the design of an implicit geologic modeling technique, and the integration of data-driven training images.

Multiple Signed Distance Function Implicit Geologic Modeling: The geologic implicit distance function method may be extended to a grid-free framework. This would facilitate the construction of geologic wireframes at arbitrary resolution. A wireframing method could be incorporated directly. The use of dual kriging for estimating the distance values would alleviate the estimation process in terms of memory and computer speed.

Geologic models are non-stationary. Further elements including local trend maps could address the conditioning of specific geologic features in the deposit. Specifically, a data integration technique that account for secondary variables and signed distances would be an interesting avenue of future research.

Implicit modeling of geologic domains does not always offer a satisfactory result when specific trends and continuity of certain domains are desired. In this case, the understanding of a trained professional becomes essential for obtaining improved results. Developing hybrid techniques considering methodologies that incorporate this understanding also remains an attractive path of investigation.

Even though the implicit distance function method provides reasonable models, the resultant geologic domains depends on the volume used during estimation. In other words, the global proportion changes with the extension of the model considered in the analysis, despite the fact that the configuration of data remains the same. This dependency might lead to edge effects and bias in the proportions of geologic domains. The lack of an effective methodology for representing global proportions of domains calls for further research in declustering techniques for categorical variables and modifications to signed distance modeling for unbiased models.

Geologic modeling relies on the analysis and interpretation of data. Data is usually represented by samples of an arbitrary dimension. In the case of continuous variables as mineral grades, this discretization makes sense since zones of high-medium-low grades need to be identified and separately modeled. The resolution of such samples has an impact upon the prediction of mineral resources and reserves. However, the case for categorical variables is different; the discretization for characterizing the geologic domains is unnecessary. Geologic description based on both classical indicator formalism and signed distance could be fully described by continuous functions. There is no need for discretized samples or compositing. Further investigation could establish the principles required to change the data representation from arbitrary point or composite samples to functions defined in space.

MPS with Multiple Data-Driven Training Images: In the current thesis, the weights are determined to reflect the geologic variability of the drillholes in the resultant realizations. An interesting research alternative consists of considering uncertainty in the weights, that is, their assignment is the result of a random function model. The weights could be drawn from this model before the integration of the

training images. This double stochastic process would allow a wider assessment of the space of uncertainty since the variability in the weights would extend to the geologic domains in the models.

In general, future research in geologic modeling should aim for the construction of hybrid models that integrate multiple interpretations. In this respect, the linear opinion pool is a suitable framework for this task. The LOP could be extended to more than just two training images. Two approaches could be conceived: 1) assign the weights to multiple training images or 2) select training images as the result of a stochastic function prior to their combination. The use of multiple training images adds flexibility and extends the range and type of geologic environment that can be modeled. Perhaps a scenario based scheme would be appropriate.

More information could be considered in the calibration of weights for the training images. The introduction of continuous variables such as grades and geometallurgical attributes could help incorporate more geology into the numerical models.

With respect to the computation of the measure of entropy, evidence has shown that the entropy depends on direction. This encourages the development of algorithms that consider local measures of geologic variability. In this scheme, the combination of training images could vary locally improving the models.

Adjustments to the simulation path has shown advantages for generating more flexible geologic models. The use of information from conditioning data to control this is another interesting area of future research. Models can be adapted to certain geological environments. Optimization technique would aid to determine phases within the simulation path for combination of training images based on particular conditions.

Finally, both the computer memory and speed are proportional to the number of training images used. Even though there is no theoretical restriction as to the number of training images, there are computer constrains when treating a large number of them, and efficient algorithms would be required in this situation. It may be also considered to adapt the proposed method to other MPS techniques as Impala (Straubhaar et al., 2011) or direct MPS sampling method (Mariethoz

et al., 2010) that do not demand for memory-expensive storage procedure as the search-tree.

Bibliography

Alabert, F. (1987). Stochastic imaging of spatial distributions using hard and soft infonilation. Msc, Stanford University, Palo Alto, California.

Allard, D., Comunian, A., and Renard, P. (2012). Probability aggregation methods in geoscience. *Mathematical Geosciences*, 44(5):545–581.

Ayuso, R., Kelley, K., Leach, D., Young, L., Slack, J., Wandless, J., Lyon, A., and Dillingham, J. (2004). Origin of the Red Dog Zn-Pb-Ag deposits, Brooks Range, Alaska: Evidence from Regional Pb and Sr isotope sources. *Economic Geology*, 99(7):1533–1553.

Bak, P. R. and Mil, A. J. (1989). Three dimensional representation in a geoscientific resource management system for the minerals industry. In Raper, J. F., editor, *Three-dimensional applications in geographical information systems*, pages 155–182. Taylor and Francis, London.

Bezier, P. (1967). Definition numerique des courbes et surfaces. *Automatisme*, pages 625–632.

Bezier, P. (1974). Mathematical and practical possibilities of unisurf. *Computer-Aided Geometric Design*, pages 12–152.

Bloomenthal, J. and Bajaj, C. (1997). *Introduction to Implicit Surfaces*. Morgan Kaufmann series in computer graphics and geometric modeling. Morgan Kaufmann Publishers, Incorporated.

Boisvert, J. (2007). Mineral deposit modeling with pseudo-genetically constructed training images. Msc, University of Alberta, Edmonton, AB.

Boisvert, J., Pyrcz, M., and Deutsch, C. (2010). Multiple point metrics to assess categorical variable models. *Natural Resources Research*, 19(3):165–175.

Boucher, A. (2009). Considering complex training images with search tree partitioning. *Computers & Geosciences*, 35(6):1151 – 1158.

Bridge, J. S. and Leeder, M. R. (1979). A simulation model of alluvial stratigraphy. *Sedimentology*, 26(5):617–644.

Caers, J. (2001). Geostatistical reservoir modelling using statistical pattern recognition. *Journal of Petroleum Science and Engineering*, 29(34):177 – 188. Soft Computing and Earth Sciences.

Carrasco, P., Ibarra, F., Rojas, R., Le Loc'h, G., and Seguret, S. (2007). Application of the truncated gaussian simulation method to a porphyry copper deposit. In *APCOM*, pages 31–39, Santiago, Chile.

Chilès, J. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley series in probability and statistics. Wiley.

Christakos, G. (1990). A bayesian/maximum-entropy view to the spatial estimation problem. *Mathematical Geology*, 22(7):763–777.

Christakos, G. (2012). *Random Field Models in Earth Sciences*. Dover Earth Science. Dover Publications.

Chugunova, T. and Hu, L. (2008). Multiple-point simulations constrained by continuous auxiliary data. *Mathematical Geosciences*, 40(2):133–146.

Cowan, E. R., Beatson, R. K., Ross, H. J., and Fright, W. R. (2003). Practical implicit geological modelling. In *5th International Mining Geology Conference*. The Australasian Institute of Mining and Metallurgy.

David, M. (1977). *Geostatistical Ore Reserve Estimation.* Developments in geo-mathematics. Elsevier Scientific Pub. Co.

De Vera, J., McClay, K., and King, A. (2004). Structure of the Red Dog District, Western Brooks Range, Alaska. *Economic Geology*, 99(7):1415–1434.

de Vries, L., Carrera, J., Falivene, O., Gratacs, O., and Slooten, L. (2009). Application of multiple point geostatistics to non-stationary images. *Mathematical Geosciences*, 41(1):29–42.

Deutsch, C. (1992). *Annealing Techniques Applied to Reservoir Modeling and the Integration of Geological and Engineering (Well Test) Data.* PhD thesis, Stanford University, Palo Alto, CA.

Deutsch, C. (2002). *Geostatistical Reservoir Modeling.* Applied geostatistics series. Oxford University Press.

Deutsch, C. (2011). Simulation of categorical variables with multiple point statistics and random perturbations: MPS-RP. Technical report, CCG Annual Report 13, University of Alberta, Edmonton, AB.

Deutsch, C. and Wang, L. (1996). Hierarchical object-based stochastic modeling of fluvial reservoirs. *Mathematical Geology*, 28(7):857–880.

Deutsch, C. V. (2006). A sequential indicator simulation program for categorical variables with point and block data: Blocksis. *Computers & Geosciences*, 32(10):1669 – 1681.

Deutsch, C. V. and Journel, A. G. (1998). *GSLIB : Geostatistical Software Library and User's Guide.* Oxford University Press.

Deutsch, C. V. and Wilde, B. J. (2013). Modeling multiple coal seams using signed distance functions and global kriging. *International Journal of Coal Geology*, 112(0):87 – 93.

Deutsch, J. and Deutsch, C. (2014). A multidimensional scaling approach to enforce reproduction of transition probabilities in truncated plurigaussian simulation. *Stochastic Environmental Research and Risk Assessment*, 28(3):707–716.

Deutsch, J. L. and Deutsch, C. (2012). Accuracy plots for categorical variables. Technical report, Centre for Computational Geostatistics, University of Alberta.

Dimitrakopoulos, R. (1998). Conditional simulation algorithms for modelling orebody uncertainty in open pit optimisation. *International Journal of Surface Mining, Reclamation and Environment*, 12(4):173–179.

Dimitrakopoulos, R., Mustapha, H., and Gloaguen, E. (2010). High-order statistics of spatial random fields: Exploring spatial cumulants for modeling complex nongaussian and non-linear phenomena. *Mathematical Geosciences*, 42(1):65–99.

Doob, J. (1953). *Stochastic Processes*. Wiley Publications in Statistics. John Wiley & Sons.

Fontaine, F. and Beucher, H. (2006). Simulation of the Muyumkum uranium roll front deposit by using truncated plurigaussian method. In *6th international mining & geology conference*, Darwin.

Franke, R. (1982). Scattered data interpolation: Tests of some method. *Mathematics of Computation*, 38(157):pp. 181–200.

Galli, A., Beucher, H., Le Loch, G., Doligez, B., and Group, H. (1994). The pros and cons of the truncated gaussian method. In Armstrong, M. and Dowd, P., editors, *Geostatistical Simulations*, volume 7 of *Quantitative Geology and Geostatistics*, pages 217–233. Springer Netherlands.

Garcia, J. and Camus, J. (2011). Value creation in the resource business. *Journal of the Southern African Institute of Mining and Metallurgy*, 111(11):801–808.

Gomez-Hernandez, J. (1991). *A Stochastic Approach to the Simulation of Block Conductivity Fields Conditioned Upon Data Measured at a Smaller Scale*. PhD thesis, Stanford University, Palo Alto, CA.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* Applied geostatistics series. Oxford University Press.

Guardiano, F. (1993). Use of multiple-point statistics for simulation of spatial phenomena. Technical report, Stanford Center for Reservoir Forecasting, Stanford University.

Guardiano, F. and Srivastava, R. (1993). Multivariate Geostatistics: Beyond bivariate moments. In Soares, A., editor, *Geostatistics Tria 92*, volume 5 of *Quantitative Geology and Geostatistics*, pages 133–144. Springer Netherlands.

Harding, A., Strebelle, S., Levy, M., Thorne, J., Xie, D., Leigh, S., Preece, R., and Scamman, R. (2005). Reservoir facies modelling: New advances in MPS. In Leuangthong, O. and Deutsch, C., editors, *Geostatistics Banff 2004*, volume 14 of *Quantitative Geology and Geostatistics*, pages 559–568. Springer Netherlands.

Hardy, R. L. (1990). Theory and applications of the multiquadric-biharmonic method 20 years of discovery 1968–1988. *Computers & Mathematics with Applications*, 19(8):163–208.

Hong, S. (2010). *Multivariate Analysis of Diverse Data for Improved Geostatistical Reservoir Modeling.* PhD thesis, University of Alberta, Edmonton, AB.

Hosseini, A. (2009). *Probabilistic Modeling of Natural Attenuation of Petroleum Hydrocarbons.* PhD thesis, University of Alberta, Edmonton, AB.

Houlding, S. W. (1991). Computer modelling limitations and new directions. *Can. Inst. Min. Metall. Bull*, pages 46–49.

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

Journel, A. (1988). *Fundamentals of Geostatistics in Five Lessons.* Stanford Center for Reservoir Forecasting, Applied Earth Sciences Department.

Journel, A. (1989). Imaging of spatial uncertainty: a non-gaussian approach. In *Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling. Proceedings.*

Journel, A. (2002). Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical Geology*, 34(5):573–596.

Journel, A. (2004). Beyond covariance: The advent of multiple point geostatistics. In *Geostatistics 2004*, Banff, Canada. Springer Netherlands.

Journel, A. and Deutsch, C. (1993). Entropy and spatial disorder. *Mathematical Geology*, 25(3):329–355.

Journel, A. and Huijbregts, C. (1978). *Mining Geostatistics.* London - New York - San Francisco Academic Press.

Journel, A. and Isaaks, E. (1984). Conditional indicator simulation: Application to a Saskatchewan uranium deposit. *Journal of the International Association for Mathematical Geology*, 16(7):685–718.

Journel, A. and Zhang, T. (2006). The necessity of a multiple-point prior model. *Mathematical Geology*, 38(5):591–610.

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

Karonen, O. and Akatemia, T. T. (1985). *Geometric Mine Modeling: Modeling of Three-dimensional Objects Based on Incomplete Information.* Acta Polytechnica Scandinavica: Mathematics and Computer Science Series. Finnish Academy of Technical Sciences.

Krige, D. G. (1951). A statistical approach to some mine valuations and allied problems at the Witwatersrand. Master's thesis, University of Witwatersrand.

Liu, Y. (2006). Using the snesim program for multiple-point statistical simulation. *Computers & Geosciences*, 32(10):1544 – 1563.

Liu, Y. and Journel, A. (2004). Improving sequential simulation with a structured path guided by information content. *Mathematical Geology*, 36(8):945–964.

Lyster, S. and Deutsch, C. V. (2008). MPS simulation in a Gibbs Sampler algorithm. In *VIII International Geostatistics Congress*, Santiago, Chile.

Manchuk, J. and Deutsch, C. (2012). Modeling of the McMurray formation using MPS tied to tidal range. Technical report, CCG Annual Report 14, University of Alberta, Edmonton, AB.

Mariethoz, G., Renard, P., and Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, 46(11):1–14.

Matheron, G. (1962). *Traite de Geostatistique Appliquee*. Technip, Paris, France.

Matheron, G. (1965). *Les Variables Regionalisees et Leur Estimation*. Masson et Cie Editeurs, Paris, France.

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

Matheron, G. (1976). A simple substitute for conditional expectation : The disjunctive kriging. In Guarascio, M., David, M., and Huijbregts, C., editors, *Advanced Geostatistics in the Mining Industry*, volume 24 of *NATO Advanced Study Institutes Series*, pages 221–236. Springer Netherlands.

Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guerillot, D., and Ravenne, C. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs. In *SPE Annual Technical Conference and Exhibition*, Dallas, Texas. Society of Petroleum Engineers.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models (Second edition)*. London: Chapman & Hall.

McLennan, J. (2007). *The Decision of Stationarity*. PhD thesis, University of Alberta, Edmonton, AB.

McLennan, J. and Deutsch, C. (2006). Implicit boundary modeling (boundsim). Technical report, CCG Annual Report 8, University of Alberta, Edmonton, AB.

Mirowski, P., Tetzlaff, D., Davies, R., McCormick, D., Williams, N., and Signer, C. (2009). Stationarity scores on training images for multipoint geostatistics. *Mathematical Geosciences*, 41(4):447–474.

Moore, D., Young, L., Modene, J., and Plahuta, J. (1986). Geologic setting and genesis of the Red Dog zinc-lead-silver deposit, Western Brooks Range, Alaska. *Economic Geology*, 81(7):1696–1727.

Munroe, M. J. (2012). A methodology for calculating tonnage uncertainty in vein-type deposits. Master's thesis, University of Alberta.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147):641–646.

Ortiz, J. (2003). *Characterization of High Order Correlation for Enhanced Indicator Simulation, PhD. Thesis.* PhD thesis, University of Alberta, Edmonton, AB.

Osher, S. and Fedkiw, R. (2003). *Level Set Methods and Dynamic Implicit Surfaces.* Applied Mathematical Sciences. Springer.

Perez, C. (2011). Mine-scale modeling of lithologies with multiplepoint. In *IAMG 2011*, Salzburg.

Polyakova, E. and Journel, A. (2007). The nu expression for probabilistic data integration. *Mathematical Geology*, 39(8):715–733.

Porter, M. (2008). *Competitive Advantage: Creating and Sustaining Superior Performance.* Free Press.

Pyrcz, M. J., Catuneanu, O., and Deutsch, C. V. (2005). Stochastic surface-based modeling of turbidite lobes. *AAPG Bulletin*, 89(2):177–191.

Richards, D. G. and Sides, G. (1991). Evolution of ore-reserve estimation strategy and methodology at Neves-Corvo copper mine. *Trans, Inst. Min. Metall. (Sect. B: Appl. earth sci.)*, pages B192–B208.

Rondon, O. (2009). A look at plurigaussian simulation for a nickel laterite deposit. In *7th international mining & geology conference*, Perth, WA.

Rossi, M. and Parker, H. (1994). Estimating recoverable reserves: Is it hopeless ? In Dimitrakopoulos, R., editor, *Geostatistics for the Next Century*, volume 6 of *Quantitative Geology and Geostatistics*, pages 259–276. Springer Netherlands.

Savchenko, V., Pasko, E. A., Okunev, O. G., and Kunii, T. L. (1995). Function representation of solids reconstructed from scattered surface points and contours. *Computer Graphics Forum*, 14:181–188.

Sichel, H. S. (1952). New methods in the statistical evaluation of mine sampling data. *Transactions of the Institution for Mining and Metallurgy*, pages 261–288.

Silva, D. and Deutsch, C. (2012). Multiple point statistics with multiple training images in geostatistics. Technical report, CCG Annual Report 14, University of Alberta, Edmonton, AB.

Srivastava, R. (1992). Iterative methods for spatial simulation. Technical report, Stanford Center for Reservoir Forecasting, Stanford University, Palo Alto, USA.

Srivastava, R. (2005). Probabilistic modeling of ore lens geometry: An alternative to deterministic wireframes. *Mathematical Geology*, 37(5):513–544.

Stone, M. (1961). The opinion pool. *Annals of Mathematical Statistics*, 32(4):1339–1342.

Straubhaar, J., Renard, P., Mariethoz, G., Froidevaux, R., and Besson, O. (2011). An improved parallel multiple-point algorithm using a list approach. *Mathematical Geosciences*, 43(3):305–328.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21.

Strebelle, S. and Journel, A. G. (1998). Stochastic reservoir simulation using neural networks trained on outcrop data. In *Society of Petroleum Engineers*, New Orleans, USA.

Strebelle, S. and Journel, A. G. (2000). Sequential simulation drawing structures from training images. In *12th International Geostatistics Congress*, Cape Town, South Africa.

Verly, G. (1983). The multigaussian approach and its applications to the estimation of local reserves. *Journal of the International Association for Mathematical Geology*, 15(2):259–286.

Wilde, B. and Deutsch, C. (2011). A new way to calibrate distance function uncertainty. Technical report, CCG Annual Report 13, University of Alberta, Edmonton, AB.

Winkler, R. L. (1968). The consensus of subjective probability distributions. *Management Science*, 15(2):pp. B61–B75.

Wu, J., Boucher, A., and Zhang, T. (2008). A SGeMS code for pattern simulation of continuous and categorical variables: FILTERSIM. *Computers & Geosciences*, 34(12):1863 – 1876.

Appendix A

The *snesim* Algorithm

The following appendix describes the practical steps to implement the snesim algorithm. A detailed review will be presented including the scanning procedure of training images, which involves the search-tree construction and template selection, pseudo-random path generation, conditioning to hard data and data event selection criteria, reproduction of large and short scale features by the implementation of multigrids, and correction of marginal probability of categorical proportions.

A.1   Multigrids

Before describing the snesim algorithm the basis of multigrids are explained. The snesim algorithm applies the same essential steps for each multigrid. The concept of multigrid was introduced by Gomez-Hernandez (1991) and consists of a discretization of the grid into M increasingly nested finer grids. The multigrids may range from $1 \leq m \leq M$ with $m = 1$ corresponding to the final simulation grid, and the multigrid m to grid of nodes separated by 2^{m-1} nodes of the final simulation grid.

An example of 4 nested multigrids is presented in Figure A.1. It is a nested structure since the nodes of the current multigrid m fills in the empty spaces of the previous M-m multigrids. The multigrids allow reproduction of spatial features in the training image at different scales, offering an organized scheme that links the large and short scale patterns.

At multigrid m, the template is rescaled according to the scale of 2^{m-1} nodes. The scanning of the training image, computing of conditional probability and simulation of nodes is performed for each multigrid independently. The simulated nodes of the previous multigrids are added as conditioning data for the next multigrids; in this manner, the large multiple point events are related to the shorter ones facilitating the reproduction of large scale features.
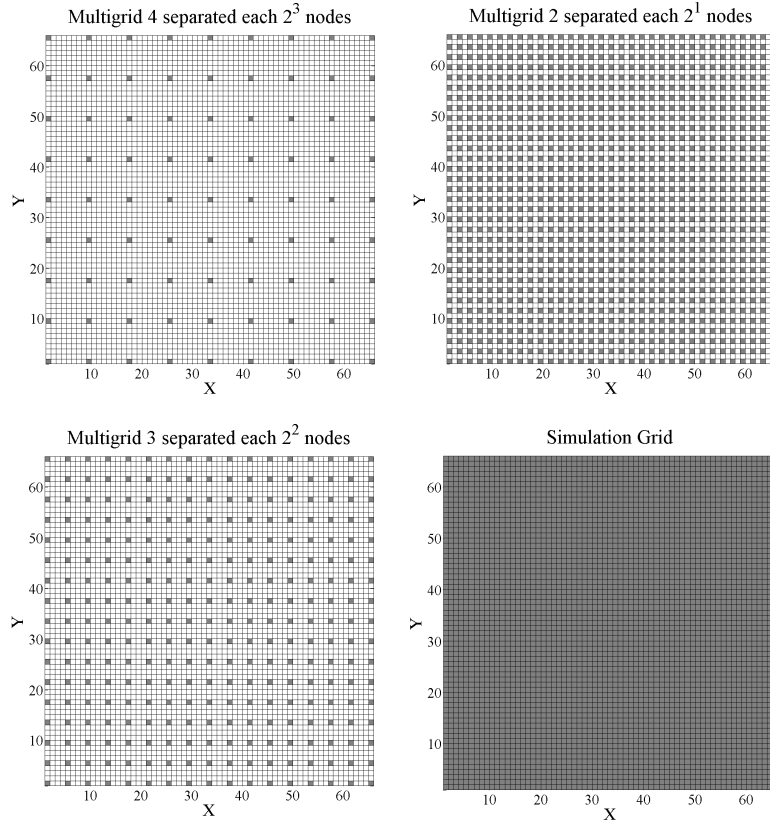
Figure A.1: Illustration of four nested 2D multigrids.

A.2 Algorithm

The snesim algorithm was devised to infer the conditional probabilities from training images and simulate categorical attributes given specific multiple point data event. The snesim algorithm is based on sequential simulation paradigm where locations

199

are simulated one at a time and each simulated values becomes a conditioning data for the next locations to be simulated. At each unsampled location, the surrounding conditioning data give rise to a data event D_n. Given the data event, the conditional probability is computed, and the Monte Carlo method simulates the categorical attribute. The simulated value is added to the conditioning data, and the algorithm goes to the next unsampled location. The process finishes when all the nodes of the grid have been simulated.

The Figure A.2 shows a flowchart of the MPS simulation based on the snesim algorithm. Further details are provided in following subsections, but the main steps are briefly described as follows:

Build Search Tree: This step is done before the simulation starts. The training image is scanned once based on the adopted template configuration. Every single node of the training image is visited and the multiple data events stored in a search-tree.

Multigrid Nodes Set up: All conditioning data are temporarily translated to their closest multigrid node. This maximizes the use of the conditioning data and enforces their reproduction in the final realizations.

Generate a Pseudo-random Path: All the nodes of the grid are visited randomly with some modification to visit the most informed nodes first without altering substantially the randomness of the process.

Data Event Selection: When a data event is not found sufficiently often in the search-tree, the farthest conditioning data is dropped and the updated data event is considered.

Compute Conditional Probability: The conditional probability of the data event is computed from the proportions extracted from the search-tree.

Draw Simulated Attribute: The categorical attribute is simulated by Monte Carlo Simulation and added to the conditioning data.

When all nodes of the multigrid have been simulated, the algorithm moves to the next multigrid and repeats all previous steps. The simulation ends when the finest multigrid has been completely simulated. In the following sections, each step
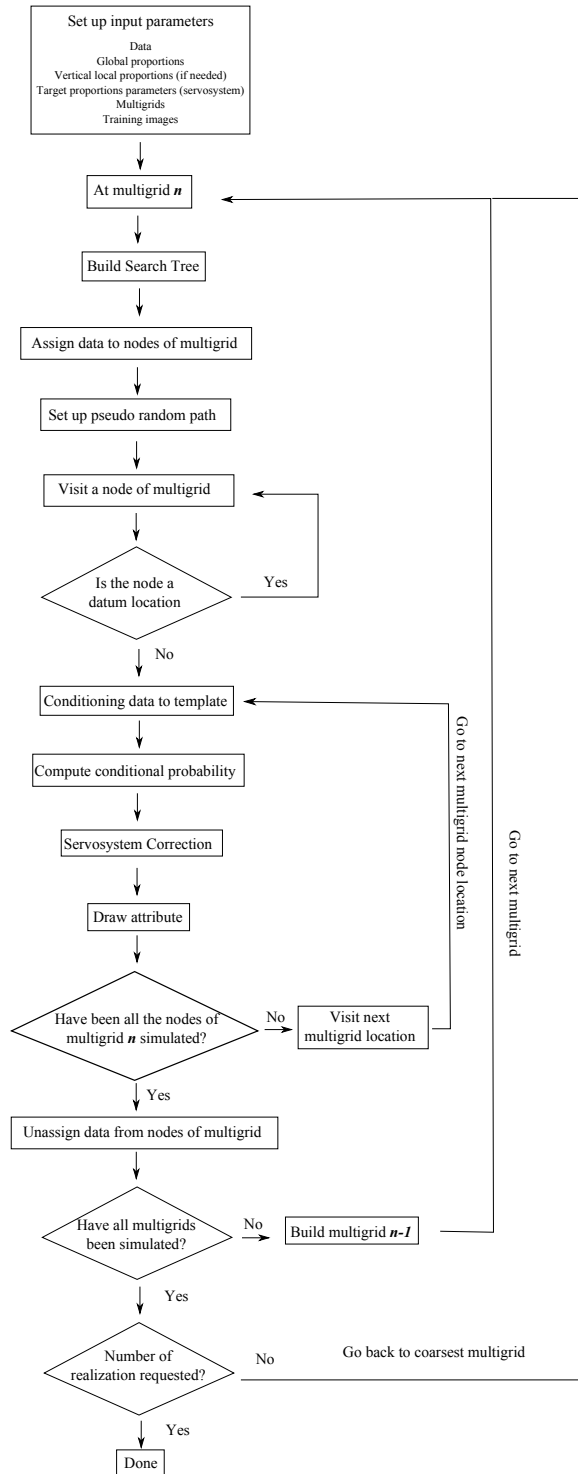
Figure A.2: Flowchart of snesim algorithm.

of the snesim flowchart will be described and explained in details.

The snesim algorithm was implemented in Fortran following the Gslib format. The snesim package is composed by the executable file "snesim.exe" and its parameter file "snesim.par". The input parameters will be related to the following subsections for having a full comprehension of the snesim algorithm. These parameters include global and vertical target proportions for servosystem correction, training image, template set up, the minimum number of data events, definition of multigrids and conditioning data searching parameters.

A.3   Storing Patterns of Training Image

A.3.1   Template

The template is an essential component for conditional probability estimation and pattern reproduction in the final models. The template consists on an arrangement of nodes in 2-D and 3-D with respect to a central location. Its size and design may take any shape. The design of the template is a user-defined parameter, and its size may range from one individual node to the size of the training image. Figure A.3 shows several examples of templates configuration.
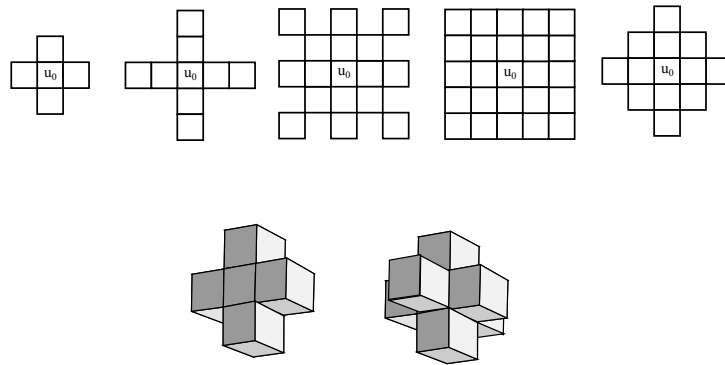
Figure A.3: Some examples of templates' configuration.

The size of the template along with the complexity of the multiple point events and the size of the training image affect the computer requirements including mem-

ory and processing time.

$$Prob(S(u) = s_k | template) \tag{A.1}$$

The template affects directly the reproduction of multiple point events from the training image. A small template might not retrieve the underlying spatial arrangements contained in the training image; on the contrary, a large template might not find even one single replicate given the size of the template, affecting the reproduction of the patterns. In the original version, Guardiano and Srivastava (1993)considered scanning the training image before simulating every unsampled location. The introduction of the search-tree algorithm by Strebelle (2002) speeds up the simulation process and reduces computer requirements.
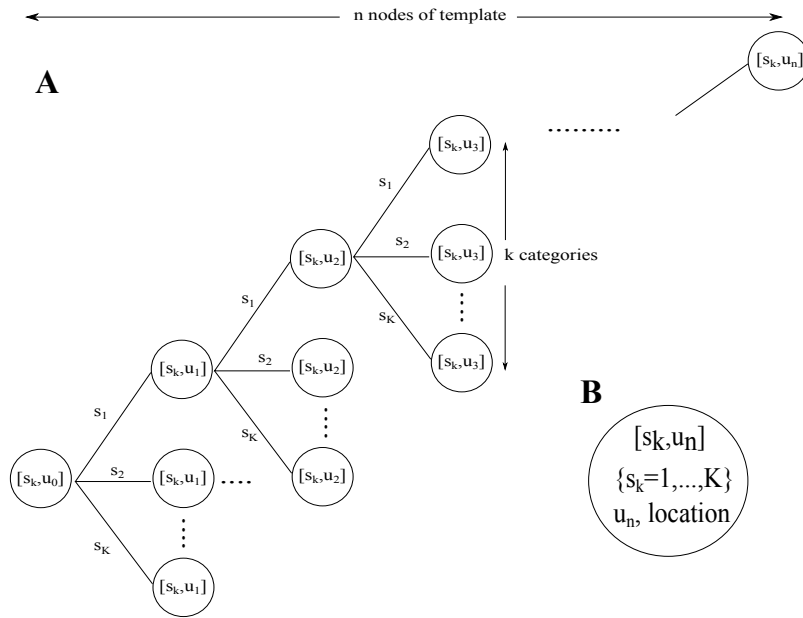
Figure A.4: Search tree general scheme. The number of nodes along the horizontal depends on the size of the template. The amount of children of each node is subjected to the categories to be modeled. The complexity of the training image determines the size of the search tree, and consequently the memory RAM requirements.

A.3.2 Search-Tree

The search-tree is an efficient dynamic storage algorithm that allows storing the multiple point events contained in the training image. It consists of a series of linked nodes that relate the multiple and sub-multiple point events according to the location of data in the template. Each node represents a unique data event configuration. Figure A.4 shows a schematic configuration of a search-tree. Each column of nodes from 1 to n (Figure A.4.A) represent the location of a node of the template. The n column nodes are sorted from the closest to farthest with respect to the center location. Each node is linked to one previous single node, but might be connected to K following nodes depending on the number of states s_k that the attributes may take. The search-tree is a dynamic structure whose size will depends on the number of categorical attributes, the size of the template and the complexity of the geologic relations in the training image. On each node, an array of size k stores the number of replicates founded in the training image according to the state s_k of the center location, see Figure A.4B.

The Figure A.5 presents a detailed flowchart of the search-tree building process in the snesim algorithm. For any multigrid n the construction starts by sorting and rescaling the template nodes to the multigrid size. At each multigrid node, retrieve the state s_k of the center location of the template. Then, go to the next closest node, if the node is not within the training image then move to the next multigrid node; otherwise, check if that multiple point event is in the search tree. If not, add a node to the search-tree, in the other case, add a replicate to the search-tree node. Later, move to the next closest template node and repeat the previous updating process. After visiting the last node with information, go to the next multigrid node. The search-tree construction ends when all nodes of the multigrid have been visited. For the next multigrid, an independent search-tree must be built.

Figure A.6 shows a small example of search-tree construction based on training image of 5 by 5, with two categories (green and yellow) and the cross-shape-template of five nodes. First, the nodes of the template are labeled from 1 to 4 according to the closeness to the center location u. Suppose that the scanning process starts
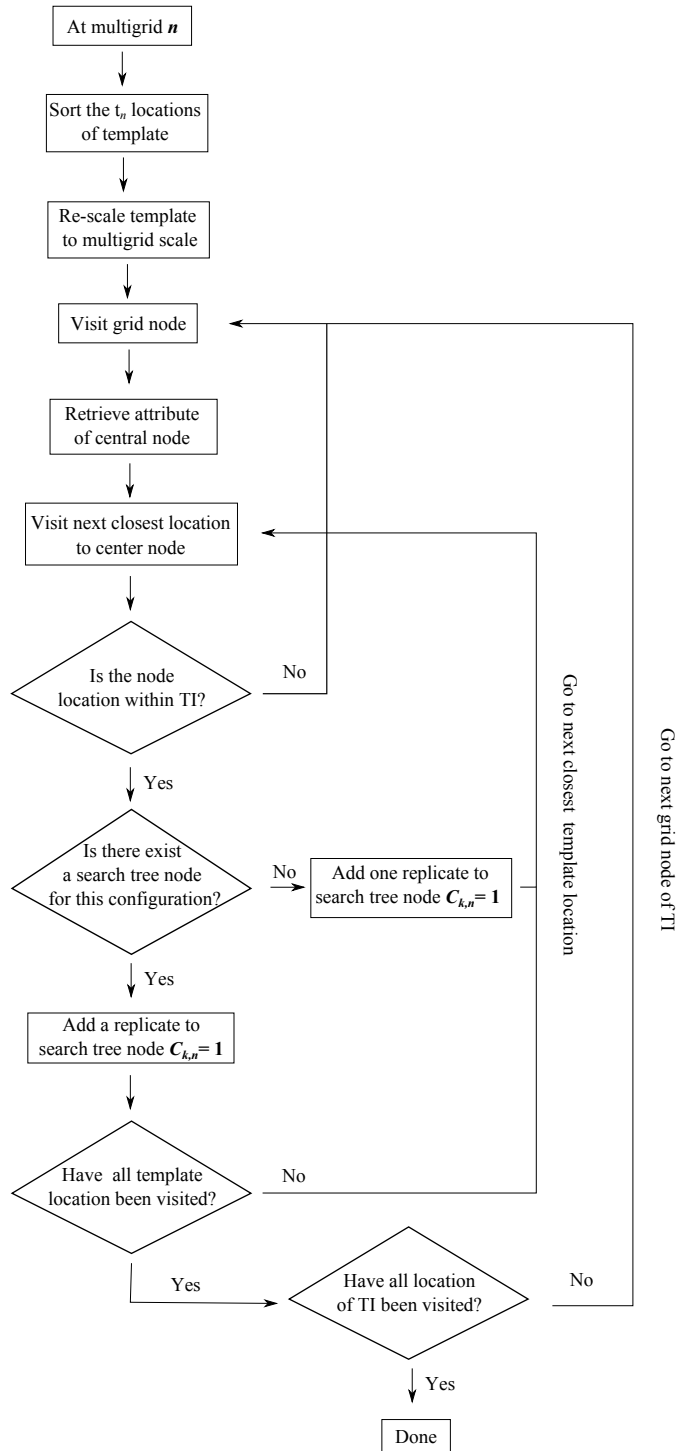
Figure A.5: Flowchart of search tree storage procedure of training image features.
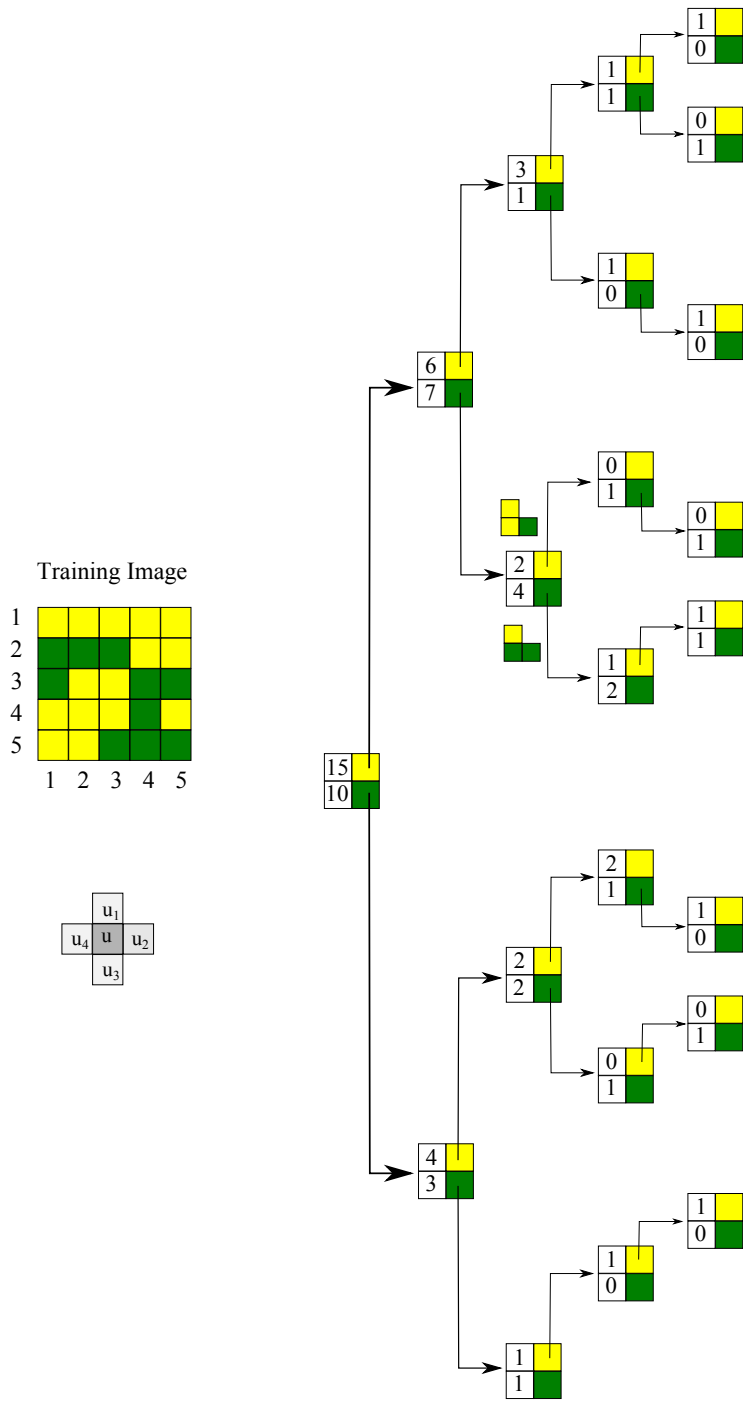
Figure A.6: Brief example of search tree construction.
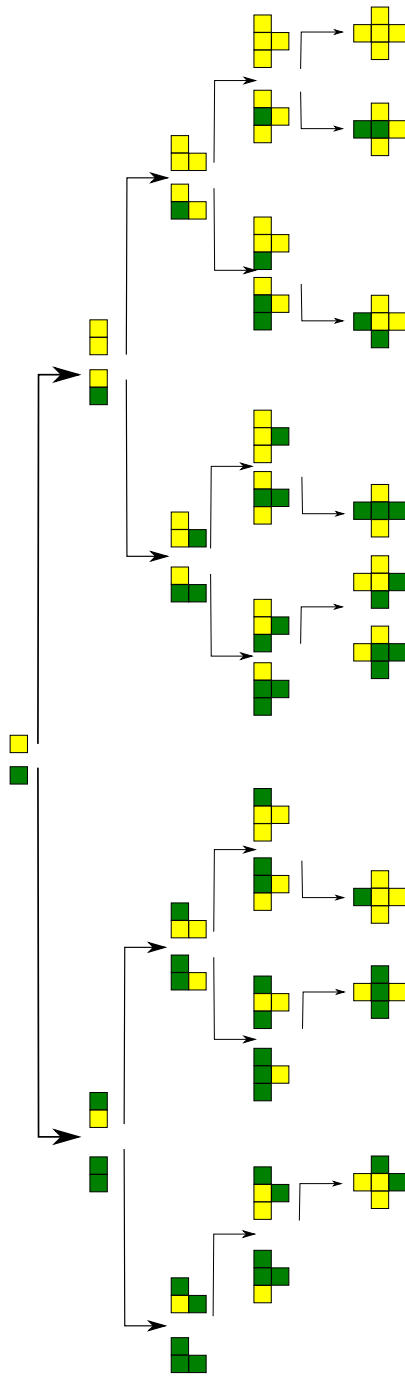
Figure A.7: Representation of multiple point events contained within each generated node of the search tree.

at the lowest-left corner of the training image. The root of the tree corresponds to the global pdf of categories green and yellow in the training image based on a single-point event. The two children of the root represent the two-point event composed by the center location u and the node u_1. Thus, e.g, the upper children stores 6 replicates where both the center location u and u_1 take the state yellow, and 7 replicates where the center location u takes the state green. The bottom children stores 3 replicates where both the center location u and u_1 take the state green, and 4 when the node u_1 change the state to yellow. A loop over the four nodes of the template allows updating and adding new nodes to the search-tree, the process finishes when all nodes of the training image have been visited. Figure A.7 illustrates the set multiple point events contained in each node of the tree.
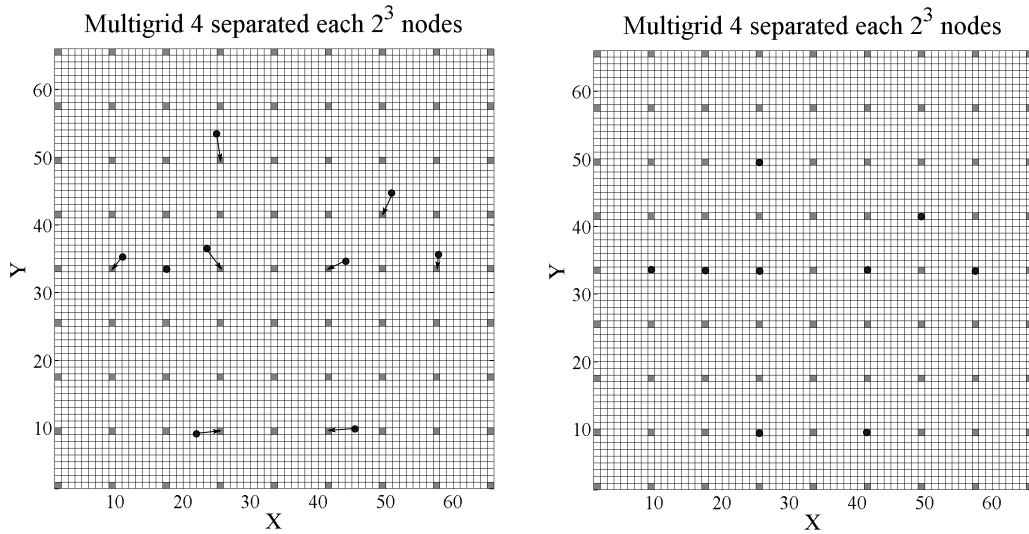
Figure A.8: Example of allocation of conditioning data to closest node of a fourth multigrid.

A.4   Multigrid Node Conditioning

The implementation of multigrids in snesim provides an organized scheme for linking and ensuring the correct reproduction of large and short multiple point events of the training image. The multigrids correspond to a series of nested regular grids spaced each 2^{m-1} nodes with respect to the simulation grid. Each multigrid is treated

independently based on the same user-define searching parameters. Once all nodes of the multigrid have been simulated these are added as conditioning data for the following multigrid. The conditioning data might not match the location of a grid node. To illustrate this point Figure A.8 presents an example of a fourth multigrid. From the 9 data (black dots on the left), just one is placed at the exact location of a multigrid node. Because of this, when the rescaled template scans the multigrid node will find just one data, affecting the reproduction of the multiple point events. To overcome this issue, the conditioning data is translated temporarily to the closest multigrid node, as shown on Figure A.8 to the right. After simulating a multigrid, the data is considered back at their original location leaving the multigrid nodes empty for the next multigrid.
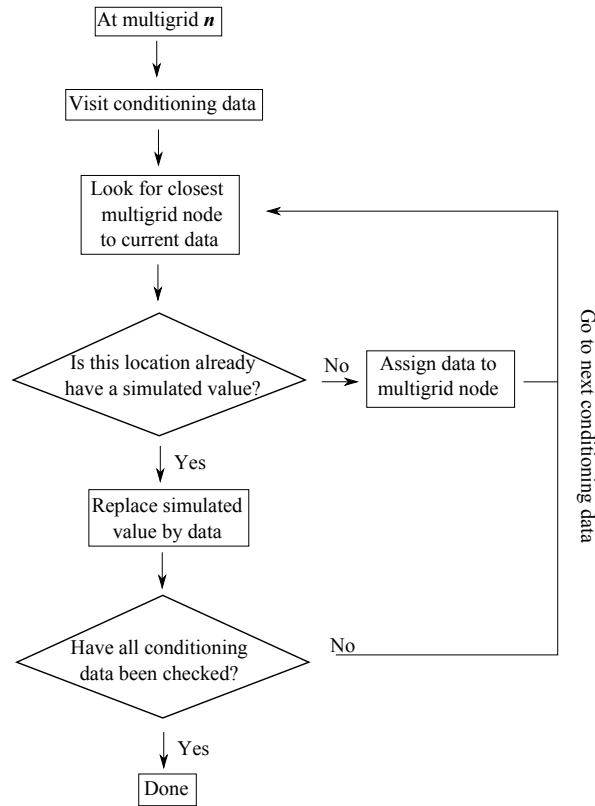
Figure A.9: Flowchart of multigrid nodes conditioning.

Figure A.9 shows the flowchart of the multigrid conditioning process. The loop is

209

performed over the conditioning data looking for the closest multigrid node location. In the case that the closest node has already a simulated value, it is replaced by the conditioning data. Once all data have been assigned, the simulation of the multigrid starts.

The relocation of the data might lead to some artifacts in the simulated results. The relocation process might require the data to be moved by a substantial distance. To avoid this, commonly the first two multigrids are implemented by using the original algorithm of , namely, no multigrid is needed and the template is scanned using the original configuration at each unsampled location. Then, the snesim mode is turned on for simulating the finest multigrids.

A.5   Pseudo-random Path

The MPS is a sequential simulation method that visits each node of the grid in a random fashion. In the case of the snesim algorithm, an informative element is incorporated to the random path generation. The most informative nodes are visited first without significantly altering the required randomness of the simulations.

Figure A.10 illustrates the flowchart for building the pseudo-random path. The process starts generating the random path by assigning a random value from a uniform distribution to each node of the grid. All the nodes are then visited to count and store the number of conditioning data within the search ellipsoid. After all nodes have been visited, the number of conditioning data in the neighborhood are added to the random values of the path. Finally, the pseudo-random path is obtained by sorting the updated random sequence in a descending order. This ensures visiting the best informed nodes first in the simulation sequence. The pseudo-random path does not lose the random character since for every realization an original random path is constructed first, resulting in a different random path.
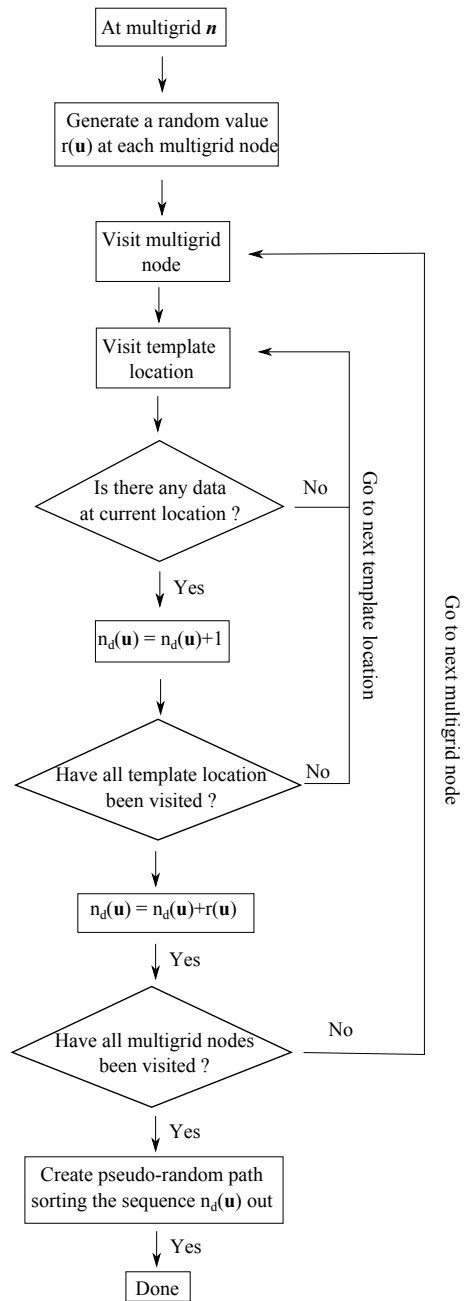
Figure A.10: Pseudo-random path building process.

A.6 Conditioning to Hard Data

The conditioning of hard data to the template depends upon two conditions, the multiple point event is in the search-tree and the number of replicates is greater or equal than *numcd*. Suppose n conditioning data are retained within the template, giving arise to the data event $D_n(u)$. The computation of the conditional probability for the data event starts from the premise that its configuration is represented by a node in the search-tree; however, this might not necessarily occur. In this case, a practical solution consists on dropping off the farthest data with respect to the center location and evaluate anew the updated data event $D_{n-1}(u)$. The data dropping process continues until the data event exists in the search-tree and its number of replicates is greater or equal to the user defined minimum number of replicates *numcd*. Figure A.11 shows a complete flowchart of the process.
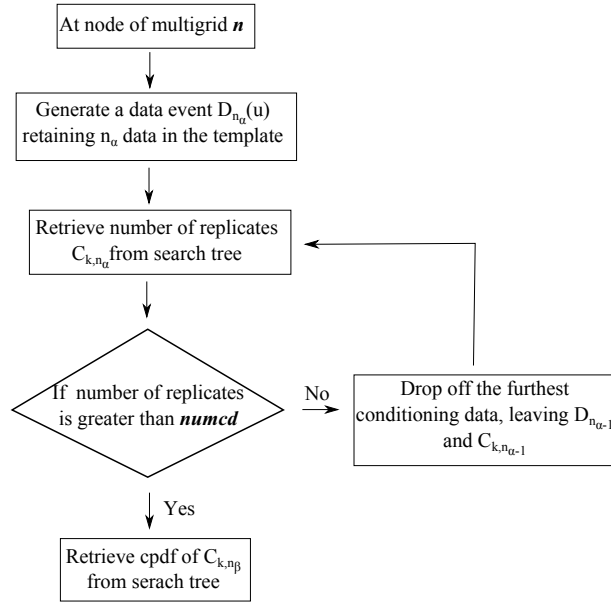
Figure A.11: Multiple point event construction by allocation conditioning data and previous simulated nodes to template.

The *numcd* influences the accuracy of the multiple point event reproduction, specially those in the coarsest multigrid since the number of replicates is considerably

lower and the variety of patters is substantially greater. A greater *numcd* does not allow finding the enough number of replicates, triggering to drop away too many nodes affecting the reproduction of large-scale features. The solution might be to set up a smaller *numcd* and find more replicates; however, some sensitivity analysis Liu (2006) has showed that the lack of large-scale reproduction persists. The reason is that the data events are not contained even once in the search-tree.

A.7 Compute cpdf of Data Event

The search-tree allows an efficient computation of the conditional probabilities by inquiring the information contained in the nodes. An array in each node of the tree stores the number of replicates founded in the training image. Thus, the conditional probability may be efficiently computed consulting the replicates of the array. For illustrating the process, the same example of Figure A.6 is used. The root of the tree provides the array for computing the global pdf. The proportions are $15/25 = 0.6$ and $10/25 = 0.4$ for yellow and green state respectively. In the same manner, the conditional probabilities for its two children are computed as follows:

Upper children:

$$Prob(S(u) = yellow | S(u_1) = yellow) = 6/13 = 0.46$$

$$Prob(S(u) = green | S(u_1) = yellow) = 7/13 = 0.54$$

Bottom children:

$$Prob(S(u) = yellow | S(u_1) = green) = 4/7 = 0.57$$

$$Prob(S(u) = green | S(u_1) = green) = 3/7 = 0.43$$

A.8 Correction of Marginal Distribution: Servosystem

The relation between the marginal distributions (rock type proportions) of realizations and the training images is not straightforward. Without taking specific steps,

the the differences might be important. Ideally, the proportions of the training image should be as close as possible to the desired proportions for the realizations; however, generating training images for such specific constraints is not an easy task. To gain control over the reproduction of the marginal distribution Strebelle (2002) proposed a correction scheme known as a servosystem to reach a pre-established set of target proportions. The procedure updates the estimated conditional probabilities before simulating the category for the locaiton.

The servosystem correction updates the conditional probability as the simulation progress from one node to the next. The idea is to modify the conditional probability based on the difference between the target and current probability. The current probability $P^c(u)$ at node location u corresponds to the proportions of all attributes previously simulated nodes. Thus, the estimated conditional probability is updated as follows:

$$P^*(A|B) = P(A|B) + \mu(P(A) - P^c(u)) \tag{A.2}$$

where $P(\mathrm{A})$ is the target pdf and μ is the servosystem correction parameter defined as $\mu = \dfrac{\lambda}{1-\lambda}$. As λ is closer to one, the more significant the difference between the target and current pdf, the stronger is the correction of the original conditional probability $P(A|B)$. After updating the conditional probability, the Monte Carlo method samples the categorical attribute from the distribution. The simulated node is added to the conditioning data and the algorithm moves to the next multigrid node. This parameter has to be properly set up since too much weight on the reproduction of the proportions might affect considerably the reproduction of patterns in the simulations.

Appendix B

Programs for Multiple Data-Driven Training Image MPS Modeling

This section covers the usage of the programs suited for implementing the multiple data-driven MPS simulation. Three programs were coded: the first details the steps involved in the generation of an implicit geologic model based on the use of distance functions measures. The second corresponds to a modification of the original snesim algorithm proposed by Strebelle (2002) to process and integrate multiple training images. The last program allows calibrating the geologic variability from the training images with respect to the drillholes.

B.1   Distance Function Implicit Modeling: *dfmod*

This program provides a quick and efficient manner of obtaining geologic models for multiple domains. Even though these models are later used as training images in the context of the current research, its application is rather general becoming in a suitable competitor for well-established implicit modeling approaches. The parameter file is presented in Figure B.1 and the details are described below:

```
 1                     Parameters for DFMOD
 2                     ********************
 3
 4  START OF PARAMETERS:
 5  cleaned_samples.prn            -   file with data
 6  0   1    2    3    4           -   columns for DH,X,Y,Z,var
 7  -1.0      1.0e21               -   trimming limits
 8  5                             -   number of categories
 9  2 3 4 5 6                      -   codes
10  1                             -   coumn for host rock
11  0.26  0.42   0.07   0.15  0.10  -   (target) global pdf
12  0      0.99                   -   target pdf repro. (0=no, 1=yes), parameter
13  dfmod.out                     -   file for distance function output
14  rtmod.out                     -   file for rock type output
15  0                             -   option: 0=grid, 1=cross, 2=jackknife
16  xvk.dat                       -   file with jackknife data
17  1    2    0     3     0        -   columns for X,Y,Z,vr and sec var
18  51       470170     25        -   nx,xmn,xsiz
19  48       6936500    25        -   ny,ymn,ysiz
20  89          3115    15        -   nz,zmn,zsiz
21  3     3       3               -   x,y and z block discretization
22  4    128                      -   min, max data for kriging
23  0                             -   max per octant (0-> not used)
24  1000   1000    1000           -   maximum search radii
25   0.0     0.0    0.0           -   angles for search ellipsoid
26  1      2.302                  -   0=SK,1=OK
27  1      0.1              -nst, nugget effect
28  1      0.9       0.0  0.0   0.0  -it,cc,ang1,ang2,ang3
29         280    280  820.0  -a hmax, a hmin, a vert
```

Figure B.1: Parameter file of *dfmod* program.

- **Line 5:** this corresponds to the name of the file with the conditioning data

- **Line 6:** here, the columns of the previous file are associated to samples coordinates and the geologic variable used for the implicit modeling

- **Line 7:** trimming limits for selected variable

- **Line 8:** number of categories or geologic domains

- **Line 9:** code of domains

- **Line 10:** column associated to the host rock. This parameter is used for avoiding extrapolation issues

- **Line 11:** target global proportions for geologic domains

- **Line 12:** 0 if servosystem is not used, 1 on the contrary. The second parameter corresponds to the degree of correction desired.

- **Line 13:** name of the output file that contains the distance function values for each domain

- **Line 14:** name of the output file for the implicit geologic model

216

- **Line 15:** option for performing the distance function modeling. (0) based on a gridded mode, (1) on cross validation or (2) applying jackknife.

- **Line 16:** name of the file for jackknife

- **Line 17:** columns that describe the content of the jackknife file

- **Line 18 to 20:** dimension of the grid model.

- **Line 21:** discretization of the block size for kriging estimation.

- **Line 22:** minimum and maximum number of samples for kriging

- **Line 23:** number of samples used by octant. (0) if octants are not applied

- **Line 24:** dimension of searching neighborhood

- **Line 25:** angles for search ellipsoid

- **Line 26:** type of kriging used, (0) for simple and (1) for ordinary kriging

- **Line 27:** number of structures for the variogram model and the value of nugget effect

- **Line 28:** type of model, value of the sill and angles of anisotropy

- **Line 29:** ranges of the variogram model

B.2 Multiple Training Image MPS: *snemimmti*

This program allows integrating multiple training images using the linear opinion pool scheme in the context of MPS simulation. Same parameters as snesim algorithm applies, the novelty relies on the use of multiple training images. Figure B.2 presents the parameter file.

- **Line 5:** name of the file that contains the conditioning data

- **Line 6:** columns of the previous file for samples coordinates and the geologic domain variable

- **Line 7:** number of categories or geologic domains

- **Line 8:** code of domains

```
 1                    Parameters for SNESIM
 2                    *********************
 3
 4   START OF PARAMETERS:
 5   cond_data.prn                  - file with original data
 6   1   2   3   4                  - columns for x, y, z, variable
 7   4                              - number of categories
 8   1 3 4 5                        - category codes
 9   0.06   0.06   0.15   0.73      - (target) global pdf
10   0                              - use (target) vertical proportions (0=no, 1=yes)
11   vertprop.dat                   - file with target vertical proportions
12   0     0.9                      - target pdf repro. (0=no, 1=yes), parameter
13   snesimti3D_10_90.out           - file for simulation output
14   10                             - number of realizations to generate
15   70         16045          15   - nx,xmn,xsiz
16   140        49915          15   - ny,ymn,ysiz
17   60          4110           5   - nz,zmn,zsiz
18   525479                         - random number seed
19   templatexyz.dat                - file for primary data template
20   32                             - max number of conditioning primary data
21   0                              - max number of data per octant (0=not used)
22   1                              - min number of data events
23   2                              - number of training images
24   0.10 0.90                      - Weights for Training Images
25   4   2                          - number of multi-grids with search trees for  training image 1
26   rtmod.out                      - file for training image 1
27   70 140 60                      - training image 1 dimensions: nxtr, nytr, nztr
28   1                              - column for variable of training image 1
29   750     750     750            - maximum search radii (hmax,hmin,vert) for  training image 1
30   0.0     0.0     0.0            - angles for search ellipsoid of training image 1
31   4   2                          - number of multi-grids with search trees for  training image 2
32   sisim.out                      - file for training image 2
33   70 140 60                      - training image 2 dimensions: nxtr, nytr, nztr
34   1                              - column for variable of training image 2
35   750     750     750            - maximum search radii (hmax,hmin,vert) for  training image 2
36   0.0     0.0     0.0            - angles for search ellipsoid of training image 2
37   70         16045          15   - nx,xmn,xsiz
38   140        49915          15   - ny,ymn,ysiz
39   60          4110           5   - nz,zmn,zsiz
40   1                              - consider local proportion model, 1=yes
41   0.0                            - Tau weight
42   lvm3D.out                      - file
43   1 2 3 4                        - columns
44   0                              - MPS RP Annealing post processing (0=not used)
45   69069                          - random number seed (MPS RP - post processing)
46   50000     50                   - reporting, number of loops
47   80        10                   - two point hist: nlag, cwt
48   0.0 0.0 0.0   100   100   10.0 - ang1,ang2,ang3,rad1,rad2,rad3
49   2                              - size of multipoint histogram cube (2-5)
50   0                              - use local proportion model (0=not used)
```

Figure B.2: Parameter file of *snesimmti* program.

- **Line 9:** target global proportions for geologic domains

- **Line 10:** 0 if vertical target proportions are used

- **Line 11:** name of the file if vertical proportions are used

- **Line 12:** 0 if servosystem is not used, 1 on the contrary. The second parameter corresponds to the degree of correction desired

- **Line 13:** name of the output file

- **Line 14:** number of realizations to simulate

- **Line 15 to 17:** dimension of the grid model

- **Line 18:** seed value for random number generator

- **Line 19:** name of the template file. This file requires the locations that conform the configuration of the template and has to be set up previously to the simulation.

- **Line 20:** maximum number of conditioning data within the template to compute the conditional probability from the training images

- **Line 21:** maximum number of conditioning data used if a search by octant is applied

- **Line 22:** minimum number of replicates of data event to compute the conditional probability. If there are not found enough replicates, the furthest sample is dropped off and the a new data event is considered

- **Line 23:** number of training images to use. In this research a maximum of two are considered; however, the number of image may be extended as required by the user

- **Line 24:** weights assigned to each training image used when integrating their conditional probabilities by the LOP scheme (linear function)

- **Line 25:** number of multigrids considered. The first parameter refers to the use of multigrid without search-trees. The second parameters corresponds to the multigrids simulated using search-trees

- **Line 26:** name of the input file that contains the first training image.

- **Line 27:** dimension of the training image

- **Line 28:** column associated to geologic domains in the training image file.

- **Line 29:** maximum size of the search ellipsoid

- **Line 30:** angles of anisotropy of the search ellipsoid

- **Line 31 to 36:** the lines from 25 to 30 are repeated for the second training image

- **Line 37 to 39:** correspond to the dimension of the grid for the local proportion mode

- **Line 40:** option that activates the use of the local proportion model

- **Line 41:** parameter associated to the Tau model. This is used for the integration of the local proportion model with the conditional probability computed from the training images.

- **Line 42:** name of the file that contains the local proportion model

- **Line 43:** columns of the previous file for the coordinates and the geologic domain

- **Line 44:** option that actives the simulated annealing post-processing method

- **Line 45:** seed value for random number generator

- **Line 46:** the first parameter corresponds to how often the program will write the value of the objective function to standard output. The second parameters corresponds to the number of times that the program looks over all grid nodes

- **Line 47:** the first number is the number of lags for the computation of two-point transition probabilities, the second corresponds to a factor that increases the importance of two-point transition provabilities associated to conditioning data

- **Line 48:** angles and ranges of anisotropy for the computation of two-point transition probabilities

- **Line 49:** this corresponds to the size of the multiple histogram cube. The size fluctuates between 2 to 5 nodes

- **Line 50:** consider the option of using the local proportion model for the post-processing simulated annealing

B.3 Calibration of Geologic Variability: *caliti*

The following program calibrates the geologic variability between the integrated training images and the drillholes. The calibration is realized based on the measure of entropy. To make comparable this measure between objects dimensionality dif-

220

ferent, a sampling process over the training image is proposed. Parameters related to this process are presented in Figure B.3.

```
 1                      Parameters for CALITI
 2                      *********************
 3
 4  START OF PARAMETERS:
 5  4                              - number of categories
 6  1   2   3   5                  - category codes
 7  2                              - 1-D template length
 8  drillholes.prn                 - file with data
 9  5   1   2   3   4              - columns for DH,X,Y,Z,var
10  100       586356.25    12.5    - nx,xmn,xsiz
11  95        145356.25    12.5    - ny,ymn,ysiz
12  20              877     5.0    - nz,zmn,zsiz
13  20 20                          - sampling mesh for training image
14  0                              - export sampling drillholes 0:No/1:Yes
15  sampling.out                   - file with sampling drillholes
16  entropy.out                    - file with sampling drillholes
17  6                              - number of training images
18  caliti_0_100.out               - file for training image
19  1                              - column for training variable
20  caliti_20_80.out               - file for training image
21  1                              - column for training variable
22  caliti_40_60.out               - file for training image
23  1                              - column for training variable
24  caliti_60_40.out               - file for training image
25  1                              - column for training variable
26  caliti_80_20.out               - file for training image
27  1                              - column for training variable
28  caliti_100_0.out               - file for training image
29  1                              - column for training variable
```

Figure B.3: Parameter file of *caliti* program.

- **Line 5:** number of categories (geologic domains) used. The categories considered must be the same for the training images and the drillholes

- **Line 6:** code of domains

- **Line 7:** size of the one-dimensional template. Recall that the number of possible configurations is given by K^N, where K is the number of location of the template and N the number of categories. As large the template, greater the computation of the entropy

- **Line 8:** name of the input file with drillhole information

- **Line 9:** columns in the previous file for drillhole ID, coordinates and geologic domain

- **Line 10 to 12:** dimension of the grid model. This has to consider the training image grid dimension

- **Line 13:** dimension of the regular sampling mesh. This corresponds to the mesh that will be used for sampling the training image

- **Line 14:** option for exporting the sampled replicated. This file may be large if the sampling mesh is too fine

- **Line 15:** name of the output file that contains the sampled replicates

- **Line 16:** name of the output file with the entropy of the drillholes and the resultant entropy after sampling the training images

- **Line 17:** number of training images to be sampled according to the mesh of Line 13 and the template designed in Line 7

- **Line 18 to 29:** on each line the name of the file of training images is introduced. The second line corresponds to the column of the file related to the geologic domains