UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERIA DE MINAS

GEOSTATISTICAL MODELING AND VALIDATION OF GEOLOGICAL LOGGINGS AND GEOLOGICAL INTERPRETATIONS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA DE MINAS

AMIR ADELI SARCHESHMEH

PROFESOR GUÍA:
XAVIER EMERY

MIEMBROS DE LA COMISIÓN
BRIAN TOWNLEY CALLEJAS
JOSÉ MUNIZAGA-ROSAS
PETER DOWD

SANTIAGO DE CHILE
2018

Resumen

La naturaleza visual del logueo geológico conduce a una clasificación cualitativa o semi-cuantitativa de los atributos petrofísicos de los testigos de sondajes, que está sujeta a errores. Debido al tiempo y dinero que se debe invertir para el relogueo, desarrollar una metodología objetiva y rápida para identificar muestras mal logueadas es una herramienta interesante para ayudar a los científicos e ingenieros antes del modelamiento geológico y geo-metalúrgico que se utilizará en la evaluación de recursos minerales y reservas mineras y en planificación minera. Respecto al modelamiento geológico, debido a la falta de cuantificación de la incertidumbre en las interpretaciones geológicas, el desarrollo de una metodología objetiva y rápida para identificar bloques mal interpretados conduce a una herramienta interesante para validar los modelos geológicos. Las interpretaciones geológicas son un insumo esencial para la evaluación de los recursos minerales y para la planificación minera y, como tal, afectan todas las etapas posteriores del proceso de minería.

En este contexto, esta tesis tiene como objetivo presentar metodologías geoestadísticas para validar el logueo geológico y las interpretaciones geológicas cuando variables cuantitativas están disponibles a partir de análisis geoquímicos o pruebas geo-metalúrgicas. Se pretende calcular, para cada testigo de sondaje o cada bloque del modelo interpretado, una medida de la consistencia entre la clase logueada o interpretada y las covariables cuantitativas. Dos modelos diferentes están diseñados para este propósito. Respecto a la validación del logueo geológico, se supone que estas covariables están subordinadas a los dominios geológicos definidos por las clases logueadas; la medida de coherencia luego se calcula utilizando como herramientas el modelo de coregionalización, el cokriging y la validación cruzada. Por el contrario, en lo que respecta a la validación de la interpretación geológica, se propone un modelo en el que los dominios geológicos se definen sobre la base de las covariables cuantitativas y se obtiene la medida de consistencia utilizando herramientas como transformación Gaussiana conjunta, modelo de coregionalización, simulación Gaussiana y clasificación por árboles de decisión.

Las herramientas y los modelos propuestos se aplican a un depósito de hierro. Los resultados muestran la capacidad de los modelos propuestos para identificar los datos para los cuales la categoría logueada no concuerda con las variables cuantitativas, así como para identificar los bloques para los cuales la categoría interpretada no concuerda con las variables cuantitativas. Vale la pena mencionar que las aplicaciones no están implementadas como caja negra, sino que permiten al profesional incluir criterios adicionales (por ejemplo, criterios geográficos) para detectar muestras o bloques sospechosos.

Abstract

The visual nature of geological core logging leads to a qualitative or a semi-quantitative classification of petrophysical attributes of drill cores, which is subject to errors. Because of the time and money needed to be invested for relogging, developing an objective and fast methodology for identifying mislogged samples is an interesting tool to assist both scientists and engineers before the geological and geo-metallurgical modeling that will be used in mineral resources and ore reserves evaluation and in mine planning. Regarding geological modeling, because of the lack of uncertainty quantification in geological interpretations, developing an objective and fast methodology for identifying misinterpreted blocks leads to an interesting tool in validating geological models. Geological interpretations are an essential input for mineral resources evaluation and for mine planning and, as such, affect all subsequent stages of the mining process.

In this context, this thesis aims to present geostatistical-based methodologies for validating geological logging and geological interpretations when quantitative variables are available from geochemical analyses or geo-metallurgical tests. It is intended to calculate, for each core sample or each block of the interpreted model, a measure of the consistency between the logged or interpreted class and the quantitative covariates. Two different models are designed for this purpose. Concerning geological logging validation, it is assumed that these covariates are subordinated to the geological domains defined by the logged classes; the consistency measure is then calculated by using coregionalization modeling, cokriging and leave-one-out cross-validation. In contrast, concerning geological interpretation validation, a model in which the geological domains are defined on the basis of the quantitative covariates is proposed and the consistency measure is obtained by using tools such as joint Gaussian transformation, coregionalization modeling, Gaussian simulation and decision-tree classification.

The proposed tools and models are applied to an iron ore deposit. The results show the ability of the proposed models to identify the data for which the logged category is not in agreement with the quantitative variables, as well as to identify the blocks for which the interpreted category is not in agreement with the quantitative variables. It is worthwhile to mention that the applications are not black-boxed and allow the practitioner to include additional criteria (e.g., geographical criteria) to detect suspicious samples or blocks.

Acknowledgments

It is a great pleasure to express my gratitude to my advisor Professor Xavier Emery for the continuous support in my Ph.D. study and research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I also acknowledge the members of my commission, Professors Brian Townley, José Munizaga-Rosas and Peter Dowd, for their insightful comments on my thesis draft.

A special thanks to my family; nothing could be possible without their support and love. Words cannot express how grateful I am to my mother, father and siblings for all the sacrifices they made on my behalf.

My thanks also go to the Mining Engineering Department and to the Advanced Mining Technology Center (AMTC) at University of Chile, as well as the CSIRO-Chile International Center of Excellence in Mining and Mineral Processing, which provided me funding and facilities to develop my Ph.D. research. My work was also partly supported by the Chilean Commission for Scientific and Technological Research (CONICYT), through the grants CONICYT / FONDECYT / REGULAR / 1130085, CONICYT / FONDECYT / REGULAR / 1170101 and CONICYT PIA Anillo ACT1407.

Finally, special thanks go to all people that helped me during this time in Chile in little and big things.

Contents

Chapter 1: Introduction

1.1. Problem setting

1.1.1. Validating and reclassifying geological logs

Drilling is the most expensive procedure in the exploration of mineral deposits and yields direct information about the geology and structure of the subsurface. Drilling is used for locating and defining economic mineralization and provides the ultimate test for all the theories and predictions that are generated in the preceding prospect generation and target generation phases of the exploration process. Information from drill holes can be extracted by different methods, such as assaying, down-the-hole geophysical logging or geological core logging (Knödel et al., 2007; Marjoribanks, 2010).

Geological core logging is the geological study, visual recording and classification of petrophysical attributes of drill cores, such as their lithology, alteration or mineralogical assemblage. The information gathered from geological core logging is the basis for constructing geological and geo-metallurgical models for mineral resources evaluation and classification, ore reserves definition and mine planning; in particular, it is used for partitioning heterogeneous deposits into geological or geo-metallurgical domains in which the regionalized properties of interest are homogeneously distributed (Sinclair and Blackwell, 2002; Moon et al., 2006; Haldar, 2013; Rossi and Deutsch, 2014).

However, due to the visual nature of logging, the classification of petrophysical attributes is qualitative and subject to errors, which may be explained by several factors, e.g. (Manchuk and Deutsch, 2012; Cáceres and Emery, 2013):

- Presence of complex rock textures caused by overprinting processes;
- Lack of chemical analyses during logging;
- Lack of experience of mining geologists;
- Non-unique logging criteria among geologists;
- Low core recoveries;
- High staff rotation;
- Inherent difficulties to estimate mineral percentages.

Inaccurate logs generate data that are inconsistent with geochemical analyses and geo-metallurgical tests, such as high copper grades in supposedly waste or leached zones, low iron grades in supposedly supergene hematite zones, or low acid consumption in supposedly calcareous rocks. According to the limited time and resources for relogging, the inconsistent logs are often seen as outliers or discarded in the geological or geo-metallurgical modeling

stage (Theys, 1999, Cáceres and Emery, 2013). This shows the importance of finding and developing methods for identifying suspicious geologically mislogged samples and checking just these suspicious samples.

1.1.2. Validating and reclassifying geological interpretations

A geological interpretation consists of a three-dimensional representation of an ore deposit constructed by resource geologists on the basis of their knowledge on the deposit, geological field observations, geophysical surveys and drill hole logs and assays. Geological interpretations that represent the spatial locations and extents of rock types or ore types are an essential input for mineral resources / ore reserves evaluation and for mine planning and, as such, affect all subsequent stages of the mining process (Duke and Hanna, 2001; Sinclair and Blackwell, 2002; Knödel et al., 2007; Marjoribanks, 2010; Rossi and Deutsch, 2014).

However, these models often correspond to a single interpretation of the ore deposit and often lack a quantification of the uncertainty in the actual rock types or ore type locations and extents. This motivates the need for quantitative methods to validate the model and to identify the areas of the deposit that have higher probabilities of being misinterpreted.

1.2. Key idea of the thesis and research objectives

For the first problem (validation of core logging), we will consider geological classes of a categorical regionalized variable, such as a rock type or a dominant alteration type, known with some imprecision, and a set of quantitative continuous covariates, such as metal grades, rock granulometry, rock density or metallurgical recoveries, known with precision from geochemical analyses and geo-metallurgical tests. Because of time and money needed to be invested for relogging, developing an objective and fast methodology for identifying mislogged samples based on the quantitative information is an interesting tool to assist scientists and engineers before the geological and geometallurgical modeling that will be used in resources/reserves evaluation and mine planning.

In this context, considering the regionalized nature of petrophysical attributes and their spatial dependence relationships with quantitative variables from geochemical analyses or metallurgical tests, a geostatistical approach based on leave-one-out cross-validation will be proposed for identifying possible mislogged samples. The proposal aims to calculate, for each sample, a measure of the consistency between the logged classes and the quantitative covariates. The rationale is to identify geologically mislogged samples and to reclassify these samples, so as to construct more reliable geological and geo-metallurgical models of mineral resources and ore reserves and to better quantify geological uncertainty.

For the second problem (validation of geological interpretations), we consider an interpreted geological model of a categorical regionalized variable and sampling data with geological classes of the same categorical regionalized variable and a set of quantitative continuous covariates. As for the first problem, the proposal will rely on the dependence relationships between the quantitative variables and the geological categories from geological core logging and interpretation. It includes the geostatistical modeling and simulation of the quantitative variables, followed by their classification into geological categories. Then, comparing the prior (without sampling information) and posterior (accounting for sampling information) probabilities of categories for each target location provides a means of identifying the locations and areas of the deposit that are most likely to be incorrectly interpreted.

1.3. Novel aspects

There are currently few geostatistical-based methodologies for validating geological logging or geological interpretations. In this context, the ultimate goal of this study is to propose geostatistical-based methodologies for validating geological logging and geological interpretations when quantitative variables are available from geochemical analyses or geo-metallurgical tests. In particular, the following aspects that will be explained in Chapters 3 to 5 are novel:

- Proposal of models to address the aforementioned problems (Chapter 3);

- Design of methodological and practical geostatistical-based tools, models and algorithms for validating geological logs, and for identifying data for which the logged category is not in agreement with the quantitative variables (Chapter 4);

- Design of methodological and practical geostatistical-based tools, models and algorithms for validating geological interpretations, and for identifying blocks or areas of the deposit for which the interpreted category is not in agreement with the quantitative variables (Chapter 5).

1.4. Hypotheses

The proposed methodologies will be based on the following four working assumptions (Cáceres and Emery, 2013):

- Chemical analyses, geo-metallurgical tests and geological features such as lithology, alteration or mineralogical assemblage should be consistent in spatial and statistical terms, insofar as they are related responses to the same geological processes.

- Chemical analyses and geo-metallurgical tests (quantitative measurements) are more accurate than geological logs and interpretations (semi-quantitative or qualitative data).

- A small proportion of the drill cores is likely to be mislogged.

- The quantitative variables are regionalized and can be interpreted as realizations of random fields that, either globally or within domains partitioning the deposit, can be transformed into stationary Gaussian random fields, i.e., random fields whose finite-dimensional distributions are multivariate normal and are invariant under a translation in space. This assumption is the basis of most current geostatistical approaches used for simulating quantitative variables and for quantifying geological uncertainty in mineral deposits (Verly, 1983; Chilès and Delfiner, 2012; Rossi and Deutsch, 2014).

Chapter 2: Literature review

In this chapter, some leading approaches that are used for drilling, core logging and data quality are reviewed. Then, some of the geostatistical concepts applied in this research are explained, and the state-of-the-art about the validation of geological logs and geological interpretations is recalled.

2.1. Drilling

2.1.1. Importance of drilling

Drilling is one of the most important and most expensive procedures in mineral exploration and yields direct information about the geology and structure of the ground below the surface. Drilling is used for locating and defining economic mineralization, and provides the ultimate test for all the theories and predictions that are generated in the preceding prospect generation and target generation phases of the exploration process (Marjoribanks, 2010; Knödel et al., 2007).

Even though a drilling program is planned in advance, which includes depth and location of drill holes, it should be concerned that such a drilling program is a dynamic system and the depth and positioning of subsequent drill holes should be updated by getting new information from already drilled boreholes (Marjoribanks, 2010).

2.1.2. Core recovery parameters

2.1.2.1. Percentage core recovery

This is the measured core recovery per drill run expressed as a percentage (De Beer et al., 1976). Recovery of a drill core is an important parameter for an effective project of mineral exploration. Good core recovery shows responsibility of the drill crews, while careless crew can ruin core by drilling too fast, over drilling a run beyond the core barrel capacity, using undesired type of barrel and faulty core lifter (Haldar, 2013).

Core recovery should be 100% except for drilling in highly fractured, altered, sheared or caved zones. If core recovery in mineralized zones is less than 85-90%, results got from cores are not trustful. Because mineralized and altered rock zones are frequently more fragile parts,

those are the first parts that will be lost during the coring, therefore such cores could not be representative of the drilled rock. Core recovery represents the precision and reliability of a deposit evaluation. In the case of having a borehole in a mineralized zone with the core recovery below 85-90%, it should be rejected and has to be drilled again with utmost care (Moon et al., 2006; Haldar, 2013).

2.1.2.2. Rock Quality Designation

Rock Quality Designation (RQD) is measured per drill run and is defined as the total length of the pieces of core greater than 10 cm divided by the length of the core run and expressed as a percentage. Measurement of the RQD yields a quantitative assessing for the quality of a rock mass between 0 and 100. A low percentage means a poor rock while a high percentage means a good quality rock (De Beer et al., 1976).

2.1.3. Core logging

The quality of drilling, in addition to the best possible core recovery, depends on some other items like proper core placement in preservation box with correct arrow marking and shifting to the core shade from drill-site, which needs extreme care before study of the core and sampling. Any mistake during this process such as misplacing or missing core during the drilling, collection, placement and shifting will add uncertainty of delimitation and estimation of resources and reserves. Initially, the core is spread in the core laboratory on the extra-long conical plastic tray. Then the arrow directions of pieces of core are checked and the edges of two successive core pieces have to perfectly match. Any discontinuity in matching two edges must be recorded (Haldar, 2013).

"Core logging" is the geological study, visual recording and classification of drill cores which is done by geologists in the core shack. Core logging includes some general description of the core, i.e., the main structural features (fracture spacing and orientation, shears, faults, folds) and a lithological description (rock type, alteration, mineralogy, visual estimates of metal values, color and texture) with other details such as core size (NX, BX, AX, BQ), run and core length, core recovery and the location of excessive core loss (when say >5%). These observations are made visually with the help of a hand-lens. The description should be systematic and as quantitative as possible; qualitative descriptions should be avoided. Lines are marked on mineralized core for splitting it into two identical halves (Moon et al., 2006; Haldar, 2013).

The more important part of the core logging is describing the geological characteristics of the rocks, which requires good knowledge in the rock formations, mineralization and alteration types of the deposit. The geological information is used for better understanding and predicting the mineralization of the studied area. There is always a degree of subjectivity

in the geological descriptions depending on the experience of the geologists and knowledge of the local geology. The geological information gathered from drill hole is as important as the assayed grades (Rossi and Deutsch, 2014).

Previously recording was just made on printed sheets but now, with advances in using computer and internet in the earth sciences, some special portable rugged laptops are designed and are extensively used for recording at site (Figure 2.1). These are designed with multiple spreadsheets for database with defined fields and features, so recording could be more accurate, more flexible and could be shared online. But still there are many exploration companies that continue working with printed sheets (Haldar, 2013).

Figure 2.1: A high-tech Internet interfaced data sharing drill core logger (Haldar, 2013)

2.1.4. Core logging systems

The importance of using the most appropriate system for geological core logging will be clearer when one concerns that the way data are recorded affects significantly the amount and type of the observed data. Although a huge number of different core logging forms are used in the industry (almost every organization and exploration group has its own standard procedure), there are only three basic methods for geological core recording. And all geological logging systems used in the industry correspond to one or a combination of these basic methods (Marjoribanks, 2010). These three geological core logging methodologies are described by Marjoribanks (2010) and are brought here.

2.1.4.1. Prose logging

In this method, an interval in depth is selected and then it is described in words. An example for this style of logging is shown in Figure 2.2. Long passages of prose are an ineffective and a laborious way of logging complex spatial relationships. Also there are not two geologists who would have the same descriptions from the same interval. It can be concluded that it is generally difficult and time consuming to extract precise and objective information from such logs. Therefore, it is not recommended to use this style, but it can be used as a complementary part in a "Comments" column to provide brief verbal commentaries in one of the other two styles of geological logging.

DEPTH	DESCRIPTION
123.45 – 136.9	*Sandstone with calcareous cement.Pale khaki colour with a medium,somewhat gritty grainsize.Bedding defined by thin shale partings at 123,34m 126.58 & 132.12 also pebbly band at128.4m. Bedding at generally 45° to LCA but becomes a bit flatter to end section. Numerous thin wispy qtz veins assoc. with pyrite speckling along really distinct margins(looks a lot like the last hole).You can see a weakish cleavage in places but mostly its not there..*

Figure 2.2: An example of the prose method for geological logging of core or cuttings (Marjoribanks, 2010)

2.1.4.2. Graphical scale logging

Graphical scale logging is the best recording system for primary drillings, which need to make use of a logging system that permits detailed observation and presents the information in a way that helps interpretation. In such logs, a down-the-hole strip map is prepared in a chosen scale and structures are added easily on the map same as how they appear in the core. As shown in an example of Graphical scale logging in Figure 2.3, different columns can be allocated for recording different features of the core. One of the advantages of this method is the possibility of indicating gradational contacts in a simple graphical manner.

Obviously, all the observations or measurements on the core cannot be shown graphically, so it is needed to be allocated some extra columns for recording digital data or for writing verbal description or some comments.

The important characteristic of these logs is that they gather many different types of geological observation on a form and all the spatial relationships can be seen at a glance. But

it should be concerned that preparing these logs is very time consuming and that is why they are just used in the first stage exploration drilling. As soon as understanding enough about the geology of a prospect (depending on the complexity of the geology and the quality of the data gathered before drilling, it could be after the first 1–2 holes or after the first 10–20 holes), a simplified, more focused and objective logging method is appropriate.

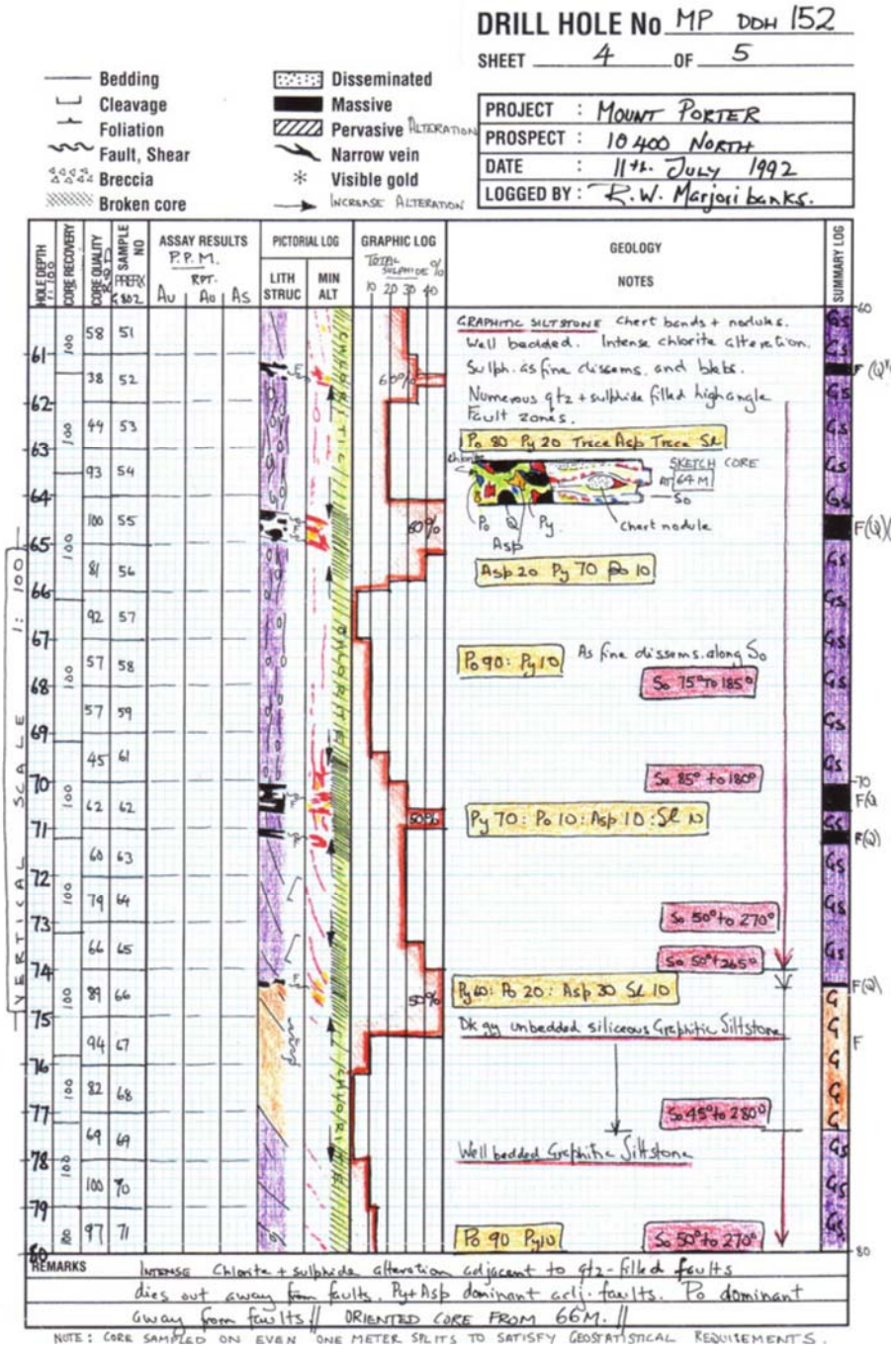
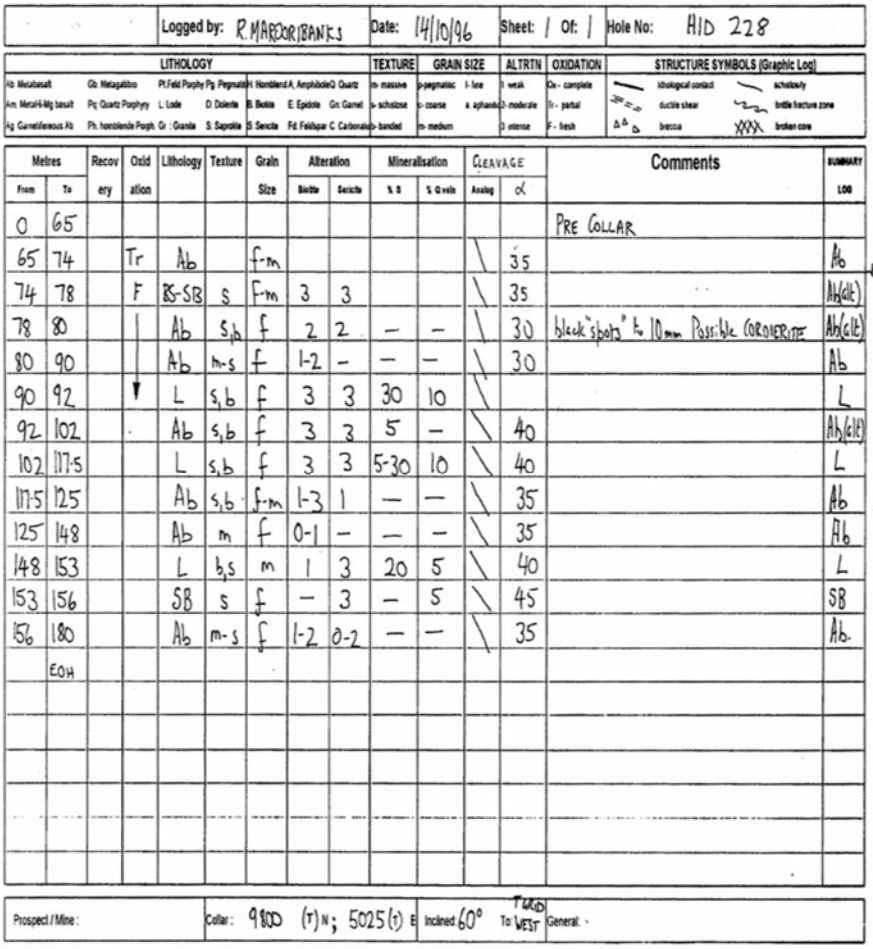


Figure 2.3: An example of the use of graphical scale logging of core (Marjoribanks, 2010).

2.1.4.3. Analytical spreadsheet logging

Analytical spreadsheet logging is ideal for the advanced drilling stages of an exploration program (resources evaluation and definition) where the main geological problems of the ore body have been solved and the logging is done just for routine recording of masses of reproducible data. In this method, observations of core are broken down into a number of objective pre-defined categories (columns) such as lithology, type and intensity of alteration, mineral content, number and type of veins, etc. Selected down-the-hole depth intervals (the rows) are then described under these categories.

For logging compactly and precisely, some numbers, abbreviations and symbols are used according to a pre-defined geocoding system. Using a geocoding system is suitable for direct entry of data into a computer and also supports a query system for faster access to the data. Figure 2.4 shows a simple example of the analytical spread sheet logging method. In spite of the advantages of this method, using this method has some problems like setting limits for the range of observations, or not providing a satisfactory way for recording the relationships between the different categories.



Logged by: R.MARJORIBANKS | Date: 14/10/96 | Sheet: 1 Of: 1 | Hole No: HID 228

Metres From	Metres To	Recovery	Oxidation	Lithology	Texture	Grain Size	Alteration Biotite	Alteration Sericite	Mineralisation % S	Mineralisation % Qvein	Cleavage Analog	Cleavage α	Comments	Summary Log
0	65												PRE COLLAR	
65	74		Tr	Ab		f-m					\	35		Ab
74	78		F	BS-SB	s	F-m	3	3			\	35		Ab(alt)
78	80		|	Ab	s,b	f	2	2	–	–	\	30	black spots" to 10mm Possible CORDIERITE	Ab(alt)
80	90		|	Ab	m-s	f	1-2	–	–	–	\	30		Ab
90	92		▼	L	s,b	f	3	3	30	10	\			L
92	102		.	Ab	s,b	f	3	3	5	–	\	40		Ab(alt)
102	117.5			L	s,b	f	3	3	5-30	10	\	40		L
117.5	125			Ab	s,b	f-m	1-3	1	–	–	\	35		Ab
125	148			Ab	m	f	0-1	–	–	–	\	35		Ab
148	153			L	b,s	m	1	3	20	5	\	40		L
153	156			SB	s	f	–	3	–	5	\	45		SB
156	180			Ab	m-s	f	1-2	0-2	–	–	\	35		Ab.
	EOH													

Prospect / Mine: | Collar: 9800 (T)N; 5025(T) E | Inclined 60° To: TURN WEST | General:

Figure 2.4: A simple example of the analytical spreadsheet logging method (Marjoribanks, 2010).

2.1.5. Sampling and assaying

Assaying diamond drill cores in the primary phases of exploration has two purposes. The first one is to provide more information on whether potentially mineable grades are present or not. The second is to give an understanding about place of economically significant elements in the system, which is necessary for targeting locations of new boreholes (Marjoribanks, 2010).

In the primary stage of exploratory drilling, the intervals for sampling should be selected by the geologist and be marked onto the core during the logging. As far as possible, the intervals should correspond to the mineralization boundaries that the geologist recognizes. Generally, in this stage, each sample is for answering a geologist's question about the core. Only where the core is relatively uniform, regular samples of predetermined length should be taken (Marjoribanks, 2010).

For sampling and analyzing the mineralized part of a core, the core is divided or split into two identical halves lengthwise with respect to the mineral distribution as observed during the logging. One half is grinded, reduced and sent to laboratory for chemical analysis while the other half is back to the core boxes as original record for future studies and audits. It is an essential reference for developing new concepts of both geological and grade continuity as the knowledge about the deposit evolves. Obviously, structural features should be recorded before splitting and a good option is photographing the wet core to produce a permanent photographic record. The second halves can be used as composite samples for metallurgical test works for recognizing the amenability, optimum grinding, liberation and recovery (Moon et al., 2006; Haldar, 2013; Sinclair and Blackwell, 2002).

2.1.6. Data quality

The mining industry collects more data than other natural resource industries, which provides a chance of better understanding local variations and obtaining robust local estimates. The abundance of data in the mining industry is in contrast with, for example, some petroleum and environmental modeling applications where the amount of data collected is limited, so the final results are more model-dependent (Rossi and Deutsch, 2014). The quality of the resources/reserves prediction is directly dependent on the quality of the data gathering and handling procedures (Erickson and Padgett, 2011). The concept of data quality is that data (samples) from a certain volume will be gathered and used to predict tonnages and grades of the elements of interest. Decisions are made based on geological knowledge and geostatistical analyses accompanied by other technical information. So the quality of the numerical basis for the analyses should be good to support sound decision-making. This is more important when one considers that just a very small fraction of the mineral deposit is sampled (Rossi and Deutsch, 2014).

Individual value measures are subject to variability from three different sources of real geological variations, sampling errors and measurement errors. It can be dealt with the effects of real geological variations through the use of geostatistical tools and methods, according to the concept of range of influence or range of correlation of a sample. This source of variability can be decreased limitedly by increasing the sample size (Sinclair and Blackwell, 2002).

There are several approaches for arranging sampling and data gathering to maximize the data quality and improve the resulting mineral resources predictions. Some of the more important factors mentioned by Sinclair and Blackwell (2002) are the following:

- ✓ knowing and reducing the magnitudes of errors in the database (assays, thickness, sample coordinates), for improving the quality of assays and the representativeness of samples;
- ✓ developing a sampling plan and procedure, appropriate to the task;
- ✓ making sure that analytical methods and procedures meet the adequate precision and accuracy;
- ✓ integrating geology into the data accumulation and evaluation phases so that data are obtained and used properly according to the material being sampled.

By minimizing errors in sampling and assaying, the nugget effect and the sill of the variogram of the measured variable decrease and the quality of the fitted spatial correlation model will be increased. Unfortunately, sampling methods, sample preparation procedures and assaying procedures are routinely accepted for different mineral deposits, without strictly checking specific characteristics of the deposit under study. Sampling parameters such as sample mass, particle size and number of particles per sample change from one mineralization to another, according to the characteristics of mineralization. There are some systematic tests to help in optimizing data-gathering procedures, which are not expensive measures and prevent from an increased error by added variance, but are too rarely used in practice (Sinclair and Blackwell, 2002).

Sampling methods and their associated parameters should not be accepted without testing (Sinclair and Vallée, 1993), as sampling methods suitable for one deposit type could be totally inappropriate for another one. The same quality control studies in sampling procedures should be applied to assaying practices and results (Vallée, 1992). Reports of assay results should be accompanied with information about sample preparation and analytical methods, containing detection limits (Sinclair and Blackwell, 2002).

Another important concept is that the samples should be representative of the volume (or material) being sampled, both in a spatial sense and at the location where the sample is being taken from. Representative sample refers to a sample that is statistically similar to any other

that one could have taken from the same volume. So it is considered that the sample values are a fair representation of the true value of the sampled volume of rock. Representation also implies that the samples have been taken in an approximately regular or quasi-regular sampling grid, so each sample is representative of a similar volume or area of the ore body. But practically, one will not face with such ideal conditions and some correction is required (Rossi and Deutsch, 2014).

2.1.6.1. Location of samples

Obviously, it is necessary to accurately know the locations of samples for using the resulting assays in mineral resources / ore reserves prediction, but sometimes one does not have accurate coordinates. So the sample sites and drill hole collars (headers) should be surveyed for accurate positioning in three dimensions. A significant reason for this problem is unintentional deviation in the borehole, which depends on the orientation of drill holes and the physical characteristics of the intersecting rocks. So drill holes should be surveyed at intermittent locations along their trace and then using a combination of this information with the collar coordinates can lead to the realistic three-dimensional coordinates of samples along the true drill-hole trace. Even small deviations from a planned orientation at the collar can lead to big movements at the end-of-hole from the planned position (Sinclair and Blackwell, 2002). Most of the commercial, down-the-hole surveying devices have the capability of measuring angle of plunge and angle of azimuth with accuracies better than 1 degree and 0.1 degree, respectively (Killeen and Elliot, 1997). This problem is shown in Figure 2.5, where small, progressive changes in the plunge of a drill hole (1–3 degrees over 50 m) result in shifting about 30 m at the end-of-hole from its planned position.

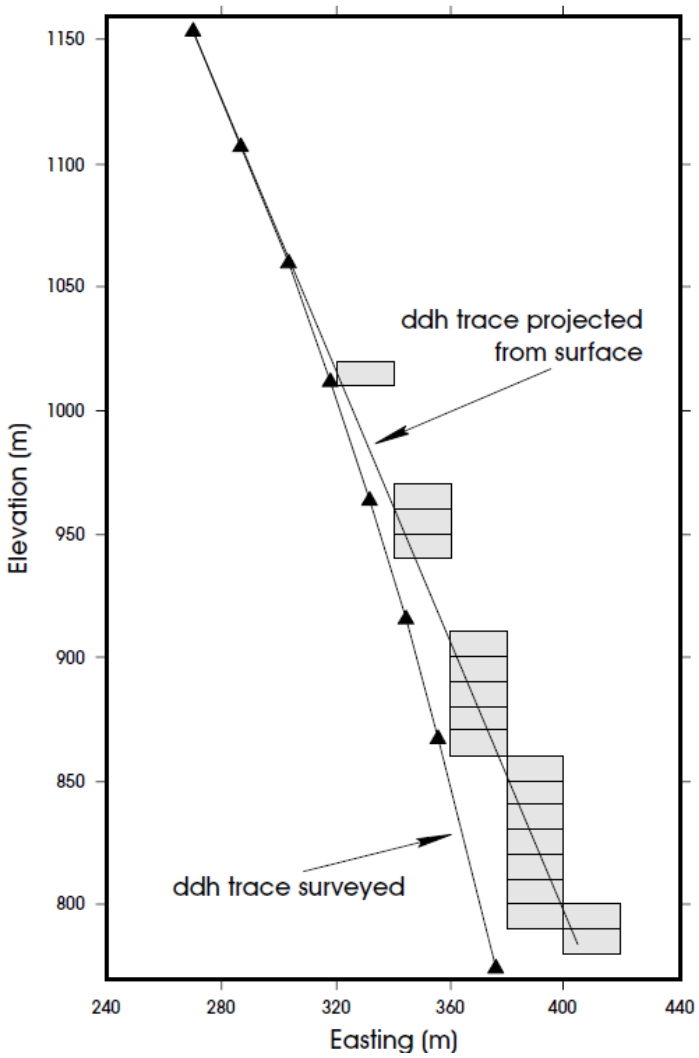


Figure 2.5: An example of a correct drill-hole trace based on surveying, comparing with the expected (projected) trace if the hole did not deviate. Triangles are survey positions along the drill hole. Dotted pattern indicates eighteen 10 ×20 m blocks intersected by expected collar orientation, but not intersected by the surveyed drill-hole position (Sinclair and Blackwell, 2002).

2.1.6.2. Sampling patterns

The first sampling patterns (grid orientation, spacing and sample sizes) are generally chosen for delimiting long range geological continuity and providing a general view of the grade distribution. Primary grids are commonly more or less square or rectangular with relatively wide-spaced grid sampling, but sampling patterns evolve by progresses in the deposit evaluation process through different stages. So the next stages should be supplemented by some local, closely spaced samples in all major directions for achieving the primary goal of a quantitative, 3D model for short-range, as well as long-range and grade continuity (Sinclair and Blackwell, 2002).

By progressing in the evaluation stages, additional factors like stricter sampling parameters (for ensuring about high quality of data), denser sampling networks and bulk-type samples are concerned for geological purposes and local estimation; especially by approaching to the mining feasibility studies. For quantifying short range continuity and possible anisotropy, closely spaced samples (e.g., assay mapping) are essential along all the principal dimensions of a deposit (Sinclair and Blackwell, 2002).

2.1.6.3. Sampling database construction

A computerized database is necessary for resources and reserves prediction and it presents another potential source of errors. If the primarily recording of the data was manual, there could be errors in the transcription and sometimes a lack of record-keeping. Also there could be some errors in digitizing the data from transcriptions. Geological databases should be compared to the information originally mapped and inconsistencies should be considered; inconsistencies can be along of errors or a decision to re-code certain drill hole intervals (Rossi and Deutsch, 2014).

Quality control programs implemented on the data should also provide an estimate of error rates in the database. There should be some safeguards and checks for quality controlling, for example percentage of grade in rock is allowed to be between 0 % and 100 %, and also there is limitation in the maximum percentage of grade that can exist in the rock according to its mineralogy. Manual checks of original assay certificates and other information like grades, down-the-hole surveys and the surveyed drill hole collar locations should be done routinely as part of the quality control of the database. It is standard for auditors to check data line by line and manually verify about 10 % of the total information in the database. Generally, error rate of less than 1 % in the checked information is considered acceptable, but error rate more than 2 or 3 % makes a line-by-line check of the entire database (Rossi and Deutsch, 2014).

2.1.7. Quality Assurance and Quality Control

There are some key principles to ensure that good scientific information is collected, which include the following steps (Mitchell, 2006):

- Step 1: Scoping and Design: A competent program design needs a good scientific knowledge about the issues, the study objectives and appropriate methods.
- Step 2: Sampling: Collecting samples requires expertise and skill.
- Step 3: Sample Analysis: Chemical analyses must be performed by valid laboratories and results must meet scientific tests to could be used trustfully.

- Step 4: Data Validation: High quality and reliability of the data must be ensured, this can be done by confirming field and laboratory methods, checking results and quality control of data, etc.
- Step 5: Data Storage: All data must be stored reliably and accessible easily. It is undesirable to have many independent databases with separate validation procedures.
- Step 6: Reporting: Qualified persons should convert the data to accurate, reliable and scientifically defensible information.

To ensure about quality of the data, some quality assurance (QA) plans should be included. QA consists of the overall policy established to achieve the orientation and objectives of an organization regarding quality (Vallée, 1998). QA is process-oriented and focuses on defect prevention, it makes sure that one is doing the right things in the right way. For example, sampling program design is a fundamental step in quality assurance. An important part of QA is quality control (QC). Quality control refers to the technical activities used to reduce errors. These activities measure the performance against defined standards, then confirm if the data meet the expected quality or not (Vallée, 1998). Actually, quality control is product-oriented and focuses on defect identification, it makes sure the results of what one has done are what one expects (USEPA, 1996). The four key steps of QC are: setting standards, appraising conformance, acting when necessary, and planning for improvements (Vallée, 1998). Table 2.1 shows general differences between QA and QC from the Statistics Canada website (https://www.statcan.gc.ca).

Table 2.1: Comparison of quality assurance and quality control, from Statistics Canada website.

Quality assurance	Quality control
• Anticipates problems before being occurred • Uses all the available information to generate improvements • Is not tied to a specific quality standard • Is applicable mostly at the planning stage • Is all-encompassing in its activities	• Responds to the observed problems • Uses ongoing measurements to make decisions on the processes or results • Requires a pre-specified quality standard for comparability • Is applicable mostly at the processing stage • Is a set procedure that is a subset of quality assurance

There are no universally accepted procedures for QA/QC, although certain basic steps are always recommended. For being confident, any mineral resources model needs to take into account overall process of QA/QC for field practices, sampling, sample preparation,

assaying, data management, correctness of the laboratory reports and transfer of the information to the database(s). The quality of the mineral resources prediction depends firstly on the available data and the geological complexity of the deposit. It is also strongly dependent on the technical skills and experience of the mine staff, the level of attention to details, the way of solving encountered problems, the open disclosure of basic assumptions accompanied with their justifications, and the quality of the documentation for each step. Justification and documentation of every important decision is part of the quality control of the work, because it forces detailed internal reviews and facilitates third-party reviews and may have a significant impact on the overall perception of the model quality (Rossi and Deutsch, 2014).

The QA/QC measures are compulsory in modern exploration programs for providing confidence in the quality of assayed data to be used for the estimation. Some of the control measures are (Haldar, 2013):

- Normally during drill core sampling, one half of the core is sent for analysis without confirming the equal representation of the other half. But in advance, some lines could be marked on mineralized core for splitting it into two halves of identical mirror images according to the mineral distribution.

- With respect to the inherent human and process error during analyzing a sample in a laboratory, some duplicate samples are inserted between the samples and analyzed at the same laboratory without disclosing the identity of the samples.

- The standard (certified reference material of known value) and blank (certified reference material of zero value) samples are sent to the laboratory with other samples for quality control purposes. Generally, these samples are put at the start, end and every 10th or 20th position in the sample string. But for quality assurance, the sequence of inserting blank and standard samples should be changed. In the case of major differences, the samples are sent back to the laboratory for repeating the analysis.

- If the samples from the same deposit are analyzed at different laboratories, according to different laboratory personnel and analytical procedures, some bias could exist between the results. A set of same sample should be analyzed at all the involved laboratories, also in a Referee Laboratory of international reputation.

- The “data error” between paired set should be tested by various statistical tests. The erratic sample pair must be identified and isolated, also possible sources of errors should be investigated and recognized. Comparing with the geological conditions, samples can be verified or rejected. The filtered data set is suitable for QA/QC analysis.

- QA/QC programs include assay data relating to various sampling methods, duplicate, standard, blank, both half-core and inter-laboratory analysis from the above mentioned activities. The data sets are statistically compared using different methods and tests. There should be a remarkable degree of correlation at high confidence level between the original and respective new assay values received from the laboratory. If the check assay results are in the acceptable range of the standard deviation or within less than 5% variation from the mean value at 95% level of significance, then the assay results are incorporated to the main assay database. This process continues with incoming additional assay inputs until ending the exploration. Unacceptable results during the process are rejected and could not be included in the database. After meeting all the QA/QC protocols, the total sample database can only be used for resources/reserves prediction and grade parameters.

2.2. Geostatistical concepts

According to the Geostatistical Glossary and Multilingual Dictionary (Olea, 1991), geostatistics is "the study of phenomena that fluctuate in space and/or time". Geostatistics is a branch of statistics defined as the application of probabilistic models for analyzing spatial datasets. It provides a collection of deterministic and statistical tools and methods for understanding and modeling spatial variability (Deutsch and Journel, 1998). Spatial – also called regionalized - variables are not completely random, as they usually show some form of structure based on the fact that locations close in space tend to assume close values (Chilès and Delfiner, 2012).

The uncertainty of a value at an unsampled location can be modeled through a probability distribution that is location-dependent. The paradigm of predictive statistics is to characterize the probability distribution of any unsampled value z(x) as a random variable Z(x), where x denotes a generic spatial location. The collection of such random variables when x belongs to the 3D space (or to a region of interest) is called a random field. In such a setting, the data values are samples from one specific realization of this random field at given locations scattered in space (Deutsch and Journel, 1998; Leuangthong et al., 2008).

2.2.1. Regionalized variable

With the assumption that only one property has been measured at different spatial locations and that the time at which the measurements have been made has not been recorded, one has n observations $z(x_\alpha)$ taken at locations x_α, with $\alpha = 1, \ldots, n$. The sampled locations in a region D can be considered as a part of a larger collection of locations. Limitations like cost and effort caused not taking more samples than the ones collected. For example, if the locations

have a (quasi-)point support, infinitely many observations are possible to be collected in the region. This possibility of taking infinitely many observations of the same type is introduced by dropping the index α to denote a generic location and by defining the *regionalized variable* as $z = \{z(x): x \in \mathcal{D}\}$. The primary dataset is now considered as a collection of a few values of the regionalized variable (Wackernagel, 2003).

A regionalized variable is a deterministic function that represents, at every location in (a region of) the geographical space and/or in time, the value of a physical property associated with a natural phenomenon (also called regionalized phenomenon). A regionalized variable usually is characterized by at least three aspects (Wackernagel, 2003):

- its nature: it can be measured on a continuous quantitative scale (continuous variable), on a discrete quantitative scale (discrete variable), or can result from the codification of the property into categories or classes (nominal or categorical variable);
- the region or domain $\mathcal{D}$ in which the variable is defined;
- the geometrical support on which the variable is measured (core of a borehole, blast hole detritus, selective mining unit, etc.).

The grades of elements of interest (products, by-products and contaminants) such as copper, arsenic, molybdenum, gold and silver, the rock density (specific gravity) and the rock type are examples of regionalized variables.

2.2.1.1. Random variable and regionalized value

Each measured value in the domain $\mathcal{D}$ is called a regionalized value. A new viewpoint is to consider a regionalized value as the outcome of some underlying random mechanism, which is called a random variable, so that a sampled value $z(x_\alpha)$ shows one draw from the parent random variable, denoted as $Z(x_\alpha)$. In the following, the random variables are shown by a capital letter (Z), while their realizations (outcome values) are shown with the corresponding lower-case letter (z) (Deutsch and Journel, 1998; Wackernagel, 2003). A random variable can take a variety of outcome values according to a given probability distribution. Since the mechanism responsible for making a value $z(x_\alpha)$ may be different at each point x_α, the random variable $Z(x_\alpha)$ could have a distribution that depend on the associated location x_α.

The cumulative distribution function (for short, cdf) of a random variable $Z(x_0)$ at a given location x_0 in the domain $\mathcal{D}$ is defined as:

$$F(x_0; z) = \text{Prob}\ \{Z(x_0) \leq z\},\ z \in R \qquad (2.1)$$

When the cdf is made specific to a particular information set of "*n*" neighboring data values $Z(x_\alpha) = z(x_\alpha)$, with $\alpha = 1, \ldots, n$, the notation "conditional to *n*" will be used. The conditional cumulative distribution function (ccdf) is defined as:

$$F(x_0; z|(n)) = \text{Prob}\ \{Z(x_0) \leq z \mid (n)\},\ z \in R \qquad (2.2)$$

Both of the above expressions (2.1) and (2.2) model the uncertainty about the value $z(x_0)$ of the regionalized variable at location x_0, the first one prior to using the information set (*n*); the second one models the posterior uncertainty once the information set (*n*) has been accounted for. The goal of any predictive algorithm is to go from prior models of uncertainty to posterior models, which incorporate the available sampling information on the regionalized variable. Note that, in general, the ccdf is a function of the location x_0, as well as the size (*n*), geometric configuration and values of the sampling data (Deutsch and Journel, 1998).

2.2.1.2. Random field

Considering the regionalized values at all points in a domain, the associated function z(x) for $x \in Ɗ$ is a regionalized variable. The set of values $\{z(x): x \in Ɗ\}$ can be viewed as one draw from an infinite set of random variables (one random variable at each location of the domain). The family of random variables $Z = \{Z(x): x \in Ɗ\}$ is called a random field (Wackernagel, 2003; Chilès and Delfiner, 2012).

Figure 2.6 shows how the random field model has been set up by considering data from to different points of view. One aspect says that the data values depend on their location in the domain, so they are *regionalized*. Another aspect says that the regionalized sample values $z(x_\alpha)$ generally cannot be modeled with a simple deterministic function z(x). According to the values of the samples, it seems that the behavior of z(x) is too complex to be modeled with deterministic methods, so a probabilistic approach is chosen, i.e. the mechanism is considered as *random*. Joining together these two aspects of *regionalization* and *randomness* leads to the concept of random field. A regionalized value $z(x_0)$ at a specific location x_0 is a realization of a random variable $Z(x_0)$, which is itself a member of an infinite family of random variables, the random field $\{Z(x): x \in Ɗ\}$. The location x_0 is an arbitrary point of the region which may or may not have been sampled (Wackernagel, 2003).

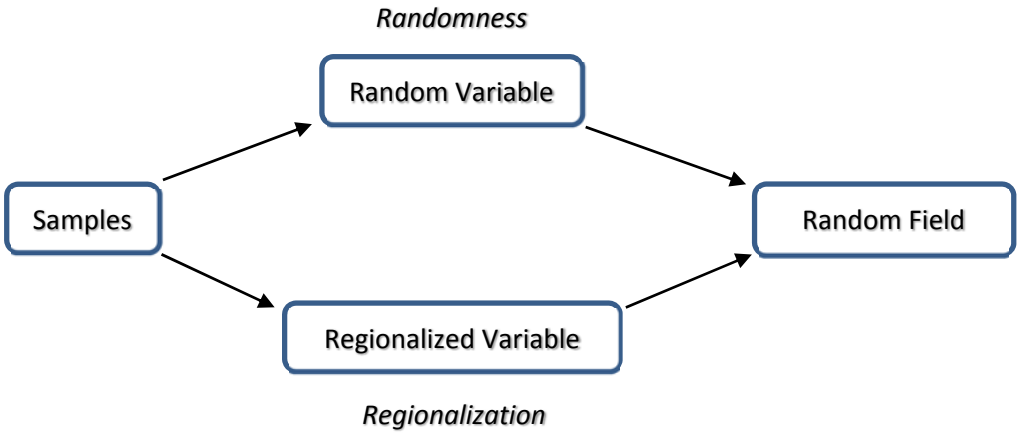


Figure 2.6: The random field model (Wackernagel, 2003).

As it is shown in expression (2.1), the random variable $Z(x_0)$ is characterized by its cdf. The random field Z is characterized by all the sets of *k* random variables located at *k* different points as (Deutsch and Journel, 1998):

$$F(x_1, \ldots, x_k; z_1, \ldots, z_k) = \text{Prob}\ \{Z(x_1) \leq z_1, \ldots, Z(x_k) \leq z_k\} \qquad (2.3)$$

with $k \in N^*$, $z_1, \ldots, z_k \in R$, $x_1, \ldots, x_k \in D$.

Based on expression (2.3), one has an extraordinarily general model that is able to describe many processes in nature or technology. Generally, the inference of any statistics, whether a univariate cdf such as (2.1) or any of its moments (mean, variance), or a multivariate cdf such as (2.3) or any of its moments (covariance, variogram), requires some repetitive sampling. In practice, at best, one just has one sample at a specific location x, in which case z(x) is known (ignoring sampling errors), so it is impossible to infer all uni- and multivariate distribution functions for any set of points without further assumptions. Some simplification is needed and it can be provided by the idea of stationary (Deutsch and Journel, 1998; Wackernagel, 2003; Chilès and Delfiner, 2012).

2.2.1.3. Strict stationarity

The lack of some repetitive samples at a given location x causes to incorporate these samples from somewhere else in space and/or time. For example, the cdf F(x; z) can be inferred from the samples gathered at other locations within the same domain, or at the same location x but at different times (Deutsch and Journel, 1998).

Referring to Wackernagel (2003), stationary means that the characteristics of a random field stay the same when shifting a given set of *k* points from one part of the domain to another.

This is called translation invariance. To be more specific, the random field $\{Z(x): x \in D\}$ is said to be strictly stationary within the domain D if its multivariate cdf (2.3) is invariant for any set of k points $x_1, \ldots, x_k$ (where k is an arbitrary positive integer) and for any translation vector h, that is:

$$F(x_1, \ldots, x_k; z_1, \ldots, z_k) = F(x_1+h, \ldots, x_k+h; z_1, \ldots, z_k) \quad (2.4)$$

According to the concept of a strictly stationary random field, the distribution functions are everywhere and always the same. But logically there should be some limits for this concept, which leads to defining several types and degrees of stationary. These lie in the wide range between the concept of a non-stationary random field, whose characteristics change at any time and at any location, and the concept of a strictly stationary random field (Wackernagel, 2003).

Strictly speaking, stationarity is a property of the random field and not of the regionalized variable, so it cannot be checked from sampling data. However, it should be mentioned that a proper or judicious decision of stationarity is often critical for the representativeness and reliability of the geostatistical tools and datasets used. For instance, mixing data across different geological domains could hide important geological differences; on the other hand, dividing data into too many subdomains could lead to unreliable statistics and results based on too few data per domain and an overall confusion (Deutsch and Journel, 1998).

2.2.1.4. Second-order and intrinsic stationarity

Based on expression (2.4), for the random field to be strictly stationary, it is required that its multivariate cdf are translation invariant for any set of k points $\{x_1, \ldots, x_k\}$. A lighter strategy is to consider only pairs of points $\{x_1, x_2\}$ in the domain and try to characterize just the first two moments (mean value and covariance function), instead of a full distribution. It should be considered that such a strategy is ideal for the multivariate Gaussian distribution where, by characterizing the first two moments, the multivariate distribution can be characterized entirely (Wackernagel, 2003).

There are two possibilities for stationarity of the first two moments, one is *second-order stationary* which assumes the existence and stationarity (translation invariance) of the first two moments of the random field. And the second one is *intrinsic stationarity* which assumes the stationarity of the first two moments of any increment (difference of a pair of values at two points) of the random field (Wackernagel, 2003).

2.2.2. Modeling spatial continuity: univariate case

2.2.2.1. Covariance of a second-order stationary random field

The covariance function C(h) (or covariance, for short) is defined based on the hypothesis of second-order stationarity (or weak stationarity), which assumes the existence and translation invariance of the first two moments (mean and covariance) of the random field:

$$\begin{aligned} &E[Z(x+h)] = E[Z(x)] = m && \text{for all } x, x+h \in Ɗ \\ &E[Z(x) - m]\,[Z(x+h) - m] = E[Z(x)\,Z(x+h)] - m^2 = C(h) && \text{for all } x, x+h \in Ɗ \end{aligned} \quad (2.5)$$

The mean value m is constant in space and the covariance function C(h) only depends on the separation vector h. A covariance function must be a positive semi-definite function. It is bounded and its absolute value does not exceed the variance (Wackernagel, 2003; Chilès and Delfiner, 2012):

$$|C(h)| \leq C(0) = \text{var}(Z(x)) \quad (2.6)$$

It is also an even function: C(-h) = C(+h). It should be noticed that h stands for a vector, so the covariance function depends both on its length (the distance between x and x+h) and on its direction. When the covariance depends only on length |h| and not on the direction, it is said to be isotropic (Wackernagel, 2003; Chilès and Delfiner, 2012).

Due to its finite variance, a stationary random field tends to fluctuate around its mean. There are some regionalized phenomena that they do not exhibit this behavior: for example, when consider increasingly large domains, the sample mean may not stabilize, and the sample variance may always increase. This is a motivation for the next model (Chilès and Delfiner, 2012).

2.2.2.2. Correlogram

A related function is the correlogram function, which is the covariance function divided by the variance and shows the correlation coefficient between Z(x) and Z(x+h). The correlogram shows how this correlation evolves with the separation vector h:

$$\rho(h) = \frac{C(h)}{C(0)} \quad (2.7)$$

which is bounded by $-1 \leq \rho(h) \leq 1$.

2.2.2.3. Variogram of an intrinsically stationary random field

The theoretical variogram $\gamma(h)$ is defined by the so-called intrinsic hypothesis, which is a milder hypothesis. This hypothesis is a statement about the type of stationarity characterizing the random field and is formed by two assumptions about the increments $(Z(x+h)-Z(x))$:

- The drift, which is the mean $m(h)$ of the increments, is invariant for any translation of a given vector h in the domain (linear drift): $m(h) = E[Z(x + h) – Z(x)] = \langle a, h \rangle$, where $\langle , \rangle$ denotes the scalar product.

- The variance of the increments has a finite value $2\gamma(h)$ depending on the length and the orientation of vector h, but not on the position of x in the domain:

  $$\text{var}[Z(x + h) – Z(x)] = 2\gamma(h)$$

To prove that the drift is linear, one considers the obvious relation of:

$$Z(x+h+h') - Z(x) = [Z(x+h) - Z(x)] + [Z(x+h+h') - Z(x+h)]$$

Using the mathematical expectation for this relation gives $m(h+h') = m(h) + m(h')$, which shows that $m(h)$ is a linear function of the vector h, namely $m(h) = \langle a, h \rangle$ for some gradient vector a. Quite often, one considers that the intrinsic random field has no drift, or $m(h) \equiv 0$. The opposite case can be included in the universal kriging model or in the formalism of intrinsic random fields of higher orders (Chilès and Delfiner, 2012).

An intrinsically stationary random field does not need to have a constant mean or a constant variance. The variogram shows how the dissimilarity between $Z(x)$ and $Z(x+h)$ evolves with the separation vector h. It is an even, nonnegative function valued 0 at the origin:

$$\gamma(h) = \gamma(-h), \qquad \gamma(h) \geq 0, \qquad \gamma(0) = 0$$

Also, $-\gamma(h)$ must be a conditionally positive definite function (Chilès and Delfiner, 2012).

In case of second-order stationarity, the variogram function exists and can be deduced from the covariance function by $\gamma(h) = C(0) – C(h)$, but in general the reverse is not true, because the variogram is not necessarily bounded under the intrinsic hypothesis. So the hypothesis of intrinsic stationarity is more general than the second-order stationarity and unbounded variogram models do not have an associated covariance function (Wackernagel, 2003; Chilès and Delfiner, 2012).

It is quite common that the preliminary steps of variogram calculation, interpretation, and modeling be performed hastily or even skipped altogether, while they have a critical role in

geostatistical studies: (1) it is a tool to investigate and quantify the spatial variability of the phenomenon under study, and (2) most geostatistical prediction or simulation algorithms require an analytical variogram model. So the mentioned practice about performing hastily variogram analysis should be reversed and much more attention should be devoted to establishing a robust model of spatial variability (variogram) before proceeding with building numerical geological models. The variogram has an extremely important role to play in the appearance and behavior of the resulting 3D models (Gringarten and Deutsch, 2001).

2.2.2.4. Basic variogram models

In the following, a few basic variogram models are presented, all of which are defined for isotropic (rotation invariant) random fields.

2.2.2.4.1. Nugget effect model

The variogram model γ(h) that models the absence of spatial correlation is the nugget effect model (Figure 2.7):

$$\gamma_{nug}(\mathrm{h}) = \begin{cases} 0 & if\ \ h = 0 \\ b & otherwise \end{cases} \quad \text{and} \quad \lim_{|h|\to 0} \gamma(h) = b \qquad (2.8)$$

where b is a positive value. In this model, two different data have independent (uncorrelated) values, irrespective of the distance separating the data locations.

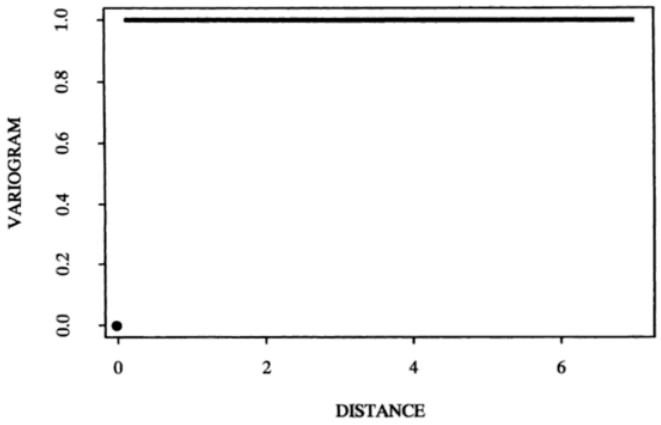


Figure 2.7: A nugget-effect variogram: its value is zero at the origin and b = 1 elsewhere (Wackernagel, 2003).

Some causes of the nugget effect (Carrasco, 2010) are:

- a "micro-structure": there actually exists a spatial structure with a very short range in comparison with the scale of observation;
- small measurement support: the nugget effect variance is inversely proportional to the volumetric support of the sample;
- measurement and position errors: sampling and/or assaying errors can create an artificial nugget effect or the so called 'human nugget effect'.

2.2.2.4.2. Spherical model

A commonly used model is the spherical model, shown in Figure 2.8a

$$\gamma_{sph}(\mathrm{h}) = \begin{cases} b\left\{\frac{3}{2}\frac{|h|}{a} - \frac{1}{2}\left(\frac{|h|}{a}\right)^3\right\} & if\ \ |h| \le 0 \\ b & otherwise \end{cases} \qquad (2.9)$$

where the parameter a shows the range of the spherical variogram model and the parameter b shows the sill (variance). The nugget effect model can be considered as a specific case of a spherical model with an infinitely small range. But is should be considered there is an important difference between these two models: $\gamma_{nug}(\mathrm{h})$ shows a discontinuous phenomenon whose values change suddenly by changing the location, while $\gamma_{sph}(\mathrm{h})$ shows a continuous phenomenon (Wackernagel, 2003).

2.2.2.4.3. Exponential model

The exponential variogram model (Figure 2.8b) increases exponentially with increasing distance and is similar to the spherical, except that it rises more steeply and reaches the sill asymptotically.

$$\gamma_{exp}(\mathrm{h}) = b\left\{1 - \exp\left(-\frac{3|h|}{a}\right)\right\} \qquad (2.10)$$

where the parameter a shows the "practical" range of the exponential variogram model, which corresponds to the distance for which the variogram reaches 95% of its sill b. As the spherical model, the exponential model is linear near the origin (Wackernagel, 2003).

2.2.2.4.4. Gaussian model

$$\gamma_{gauss}\,(\mathrm{h}) = b\,\{1 - \exp(-\frac{3|h|^2}{a^2})\} \qquad (2.11)$$

where the parameters of a and b show the practical range and sill of the Gaussian variogram model. The Gaussian model (Figure 2.8c), with its parabolic behavior at the origin, implies more short scale continuity and smoothness. It is suitable for slowly-varying variables, since the increase in variance is very gradual with distance (Rossi and Deutsch, 2014).

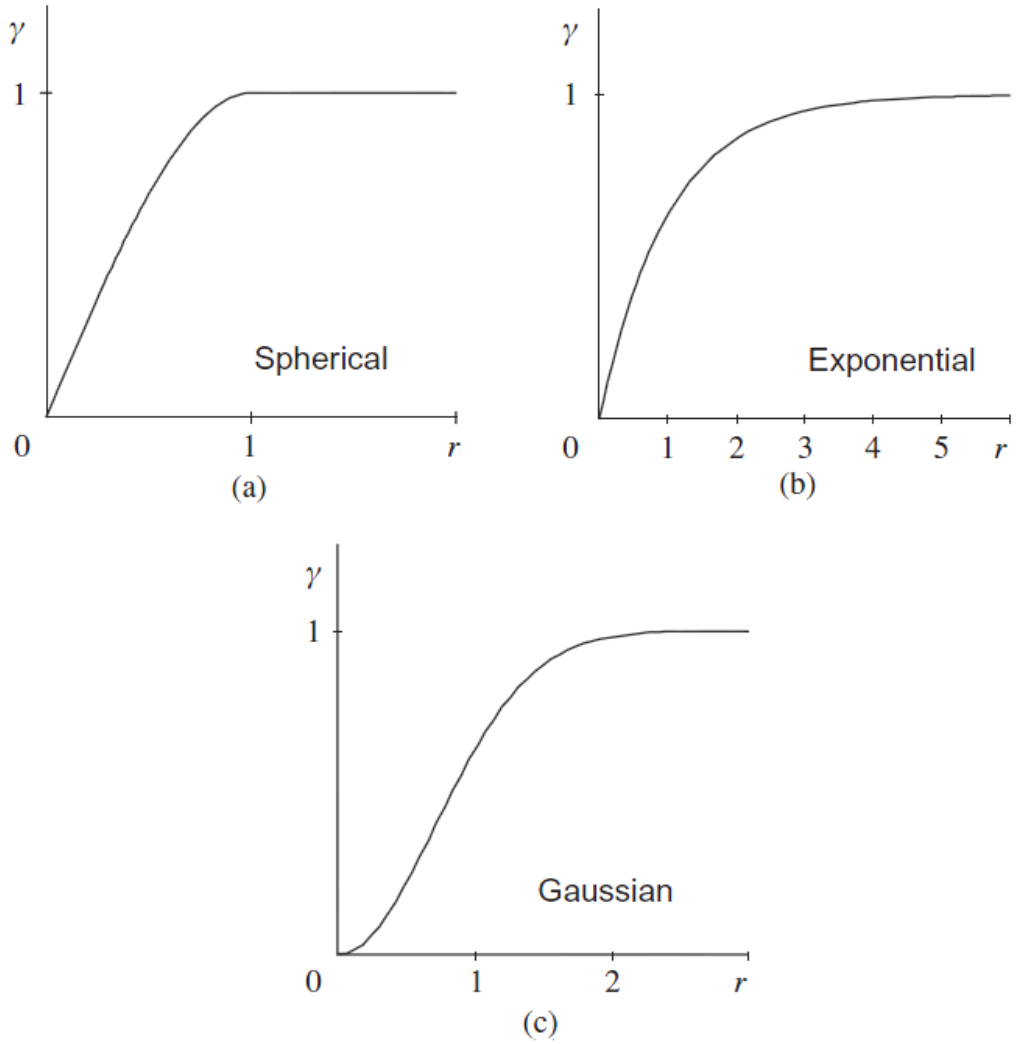


Figure 2.8: Variogram models with unit sill: (a) spherical; (b) exponential; (c) Gaussian (Chilès and Delfiner, 2012).

2.2.2.5. Anisotropy

In the previous section, it was assumed that all the discussed variograms and generally the spatial correlation structures are the same in all directions, or isotropic. So the covariance function, correlogram and variogram depend only on the length of the lag vector of h, and not on its direction. Experimental covariance, correlogram or variogram can be computed by pooling data pairs separated by the appropriate distances and regardless of the direction; such an experimental covariance, correlogram or variogram is referred to as omnidirectional.

However, in many cases experimental calculations reveal a different behavior in different directions, which is called anisotropy. Anisotropy indicates that the regionalized variable possesses some preferential directions of spatial continuity. In practice, anisotropies can be detected by the variogram map, i.e., the map of the experimental variogram as a function of the separation vector (distance and orientation). Anisotropy directions can be pre-determined based on geological knowledge. As previously explained, variogram models are defined for the isotropic case, so some transformations of the coordinates are needed to obtain anisotropic random fields from the isotropic models (Wackernagel, 2003; Rossi and Deutsch, 2014).

2.2.2.5.1. Geometric anisotropy

The variogram map as a function of a vector h can be drawn for showing the behavior of the experimental variogram. If the iso-value lines are circular around the origin, the variogram just depends on the length of the vector h and the phenomenon is isotropic. In other cases, the iso-value lines can be approximated by concentric ellipses (2D) or ellipsoids (3D) along a set of perpendicular main axes of anisotropy. This type of anisotropy where the directional variograms present the same level of variance (sill) in all directions, but the ranges are different, is called a geometric anisotropy (Figure 2.9) and relates the anisotropic variogram to a corresponding isotropic variogram by a geometric transformation (rotation-reduction of the coordinates):

$$\gamma_{anisotropic}(\mathrm{h}) = \gamma_{isotropic}(\mathrm{r}) \text{ with } r = \sqrt{\left(\frac{\mathrm{h}_x}{\mathrm{a}_x}\right)^2 + \left(\frac{\mathrm{h}_y}{\mathrm{a}_y}\right)^2 + \left(\frac{\mathrm{h}_z}{\mathrm{a}_z}\right)^2} \tag{2.12}$$

with (x,y,z) the main anisotropy axes and $(\mathrm{a}_x, \mathrm{a}_y, \mathrm{a}_z)$ the correlation ranges along these axes.

This transformation extends in a simple way a given isotropic variogram to a whole class of anisotropic variograms. The modeling only needs specifying the sill, the main directions (orthogonal) and the corresponding correlation ranges (Wackernagel, 2003; Rossi and Deutsch, 2014).

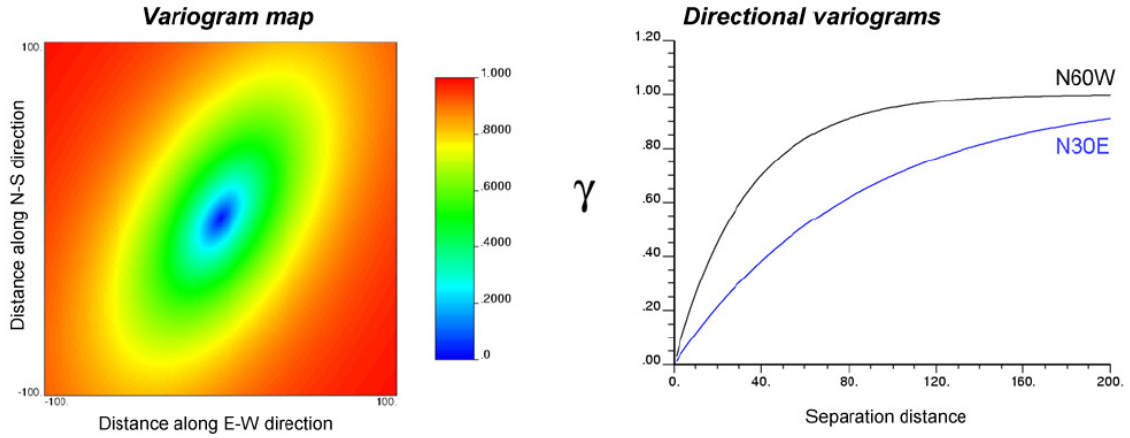


Figure 2.9: An example of a variogram map and related directional variograms with geometric anisotropy (Emery, 2017).

2.2.2.5.2. Zonal anisotropy

Another type of anisotropy is zonal anisotropy and it happens when the variograms calculated in different directions suggest a different value for the sill (Figure 2.10). Generally, it cannot be modeled using a simple coordinate transformation; in this case, one option is to consider an additional structure in the specific direction where the zonal component appears. This option is a special case of geometric anisotropy where the sill is reached asymptotically at large distances. So the zonal anisotropy changes to a geometric anisotropy case with a very large correlation range in one or more of the main anisotropy axes (Rossi and Deutsch, 2014).

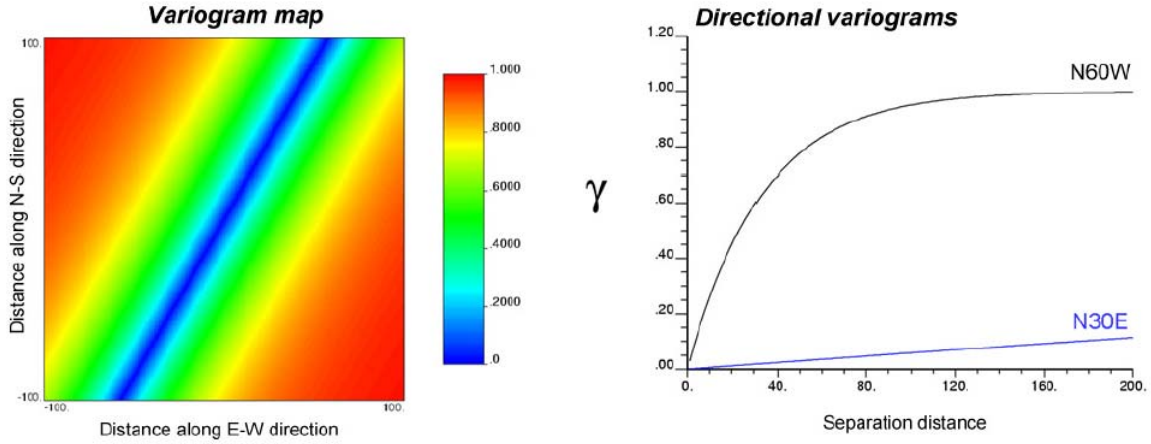


Figure 2.10: An example for a variogram map and related directional variograms with zonal anisotropy (Emery, 2017).

2.2.2.6. Fitting a variogram model

The experimental variogram (also known as sample variogram) measures the mean squared deviation between two data values, as a function of the separation vector between the data locations:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|}\sum_{N(h)}[z(x_\alpha) - z(x_\beta)]^2 \quad (2.13)$$

where N(h) is the number of pairs of data locations $\{x_\alpha, x_\beta\}$ a vector h apart (with, in practice, some calculation tolerances). Also a sample covariance can be calculated as:

$$\hat{C}(h) = \frac{1}{|N(h)|}\sum_{(\alpha,\beta)\epsilon N(h)}[z(x_\alpha) - \bar{z}][z(x_\beta) - \bar{z}] \quad (2.14)$$

where $\bar{z}$ is an estimate of the mean value, e.g., the average of all the data values.

Variogram or covariance inference provides a set of experimental values for a finite number of lags and directions. Continuous functions (theoretical variogram or covariance) should be fitted to these experimental values to deduce variogram or covariance values for any possible lag h required by prediction or simulation algorithms (Goovaerts, 1997).

2.2.2.6.1. Manual, semi-automatic or automatic fitting

Fitting a model can be done with manual fitting, semi-automatic and automatic methods (Goovaerts, 1997; Chilès and Delfiner, 2012). But it should not be trusted to purely automatic fitting procedures because they cannot take into account ancillary information which is critical when sparse or preferential sampling makes the experimental variogram non-robust. Semi-automatic fitting methods are suggested when fitting is considered as an interactive work and the user has the final word. So the user pays attention to many factors during the fitting, such as the representativeness of the experimental variogram (number of data pairs used at each lag and direction), available information on the regionalized variable (sampling errors, geological information about the main directions of continuity) or working scale (Goovaerts, 1997; Chilès and Delfiner, 2012).

Semi-automatic procedures may facilitate the determination of model parameters when the form of the experimental variogram is clear. In these methods, the most important decisions regarding the number, type and anisotropy of basic variogram models (nested structures) must be taken by the user. Depending on the defined procedure, some methods need the sill or the range for each basic variogram model. With the help of a good interactive graphical program, the user could do variogram fitting better than sophisticated fully automatic fitting procedures (Goovaerts, 1997).

Weighted least squares are commonly used in software to provide an automatic or semi-automatic fit. In these methods the sum of squared deviations between the modeled points and the corresponding experimental variogram values is measured and minimized (Cressie, 1985; Isaaks, 1999; Rossi and Deutsch, 2014).

It is a good idea to perform leave-one-out cross-validation or split-sample jackknife to compare alternative variogram models. The comparison is done based on the results of the final objective, which most often is some kind of prediction. Each of two methods, leave-one-out cross-validation or split-sample jack-knife, could be performed on different sets of estimates resulted from alternative variogram models; the better variogram model would be the one that yields a lower average prediction error (Rossi and Deutsch, 2014).

2.2.2.6.2. Linear model of regionalization

A regionalized phenomenon can be considered as being the sum of several uncorrelated sub-phenomena with different spatial scales. So a model can be set up, which splits the random field representing the phenomenon into several uncorrelated random fields (Wackernagel, 2003).

For instance, a second-order stationary random field Z can be built by adding independent zero-mean second-order stationary random fields Z_u and a constant m representing the expectation of Z:

$$Z(x) = Z_1(x) + \ldots + Z_u(x) + \ldots + Z_s(x) + m, \quad (2.15)$$

where $\text{cov}(Z_u(x), Z_v(x+h)) = 0$ for $u \neq v$. With a simple computation, it can be proved that the corresponding covariance function model is a nested model of the form (Goovaerts, 1997; Wackernagel, 2003):

$$C(h) = C_1(h) + \ldots + C_u(h) + \ldots + C_s(h). \quad (2.16)$$

In the same way, an intrinsic random field Z(x) associated with a nested variogram can be thought as being the sum of S components acting at different scales, the increments of which are zero on average and uncorrelated:

$$Z(x) = Z_1(x) + \ldots + Z_u(x) + \ldots + Z_s(x), \quad (2.17)$$

With a simple computation, it can be shown that the corresponding variogram model is of the form (Goovaerts, 1997; Wackernagel, 2003):

$$\gamma(h) = \gamma_1(h) + \ldots + \gamma_u(h) + \ldots + \gamma_s(h). \quad (2.18)$$

Often, several sills can be recognized on the variogram, which are related to the morphology of the regionalized variable. For example, Serra (1968) investigated spatial variations in an iron deposit and found up to seven sills, each with a geological interpretation, in the multiple transitions between the micrometric and the kilometric scales. For identifying small-scale factors, sufficiently fine sampling grids are required, while for identifying large-scale factors on the variograms, sufficiently large diameter of the sampling domain is needed (Wackernagel, 2003).

By numbering different sills observed on the experimental variogram with an index u = 1, …, S, the nested variogram (2.18) with S elementary variograms can be rewritten as:

$$\gamma(h) = \sum_{u=1}^{S} \gamma_u(h) = \sum_{u=1}^{S} b_u g_u(h) \quad (2.19)$$

where the $g_u(h)$ are basic normalized variograms with given ranges, i.e. elementary variogram models with a sill normalized to one. The positive coefficients b_u express the actual values of the sills of the basic variograms (Goovaerts, 1997; Wackernagel, 2003).

2.2.3. Modeling spatial continuity: multivariate case

2.2.3.1. Cross-covariance function

The direct and cross-covariance functions are defined in the context of a joint second-order stationary hypothesis for N random fields Z_i, with i = 1,…, N, when for any x, x+h ∈ Ɗ and all pairs i, j = 1, …, N, one has:

$$\begin{cases} E[Z_i(x)] = m_i \\ E\left[(Z_i(x) - m_i).\left(Z_j(x+h) - m_j\right)\right] = C_{ij}(h) \end{cases} \quad (2.20)$$

The mean of each random field Z_i at any x ∈ Ɗ is equal to a constant (m_i) and the covariance between a pair of random fields just depends on the translation vector h, so it is invariant for any translation of the point pair in the domain (Wackernagel, 2003).

The cross-covariance function C_{ij} for i ≠ j generally is not an even or an odd function. The following equalities and inequalities hold for cross-covariance functions:

$$\begin{cases} C_{ij}(h) \neq C_{ij}(\mathrm{h}) \\ C_{ij}(-h) \neq C_{ij}(\mathrm{h}) \\ C_{ij}(h) = C_{ji}(-\mathrm{h}) \end{cases} \quad (2.21)$$

The cross-covariance function between two random fields may take negative values, even at the origin, which indicates that the random fields are negatively correlated. Also it should be considered that the maximum value of a cross-covariance function may occur for a non-zero lag separation vector (delay effect, frequently observed with random fields that depend on time and evolve asynchronously).

2.2.3.2. Cross-variogram

The direct and cross-variograms of a set of N random fields Z_i, with i = 1,…, N, are defined in the context of a joint intrinsic hypothesis, when for any x, x+h ∈ Ɗ and all pairs i, j = 1, …, N, one has:

$$\begin{cases} E[\mathrm{Z}_i(\mathrm{x}+\mathrm{h}) - \mathrm{Z}_i(\mathrm{x})] = 0 \\ cov\left[\left(\mathrm{Z}_i(\mathrm{x}+\mathrm{h}) - \mathrm{Z}_i(\mathrm{x})\right), \left(\mathrm{Z}_j(\mathrm{x}+\mathrm{h}) - \mathrm{Z}_j(\mathrm{x})\right)\right] = 2\gamma_{ij}(h) \end{cases} \quad (2.22)$$

The cross-variogram can also be defined as in the following equation (2.23), which shows that it is an even function (Wackernagel, 2003):

$$\gamma_{ij}(\mathrm{h}) = \frac{1}{2} E[(Z_i(\mathrm{x}+\mathrm{h})-Z_i(\mathrm{x})).(Z_j(\mathrm{x}+\mathrm{h})-Z_j(\mathrm{x}))] \quad (2.23)$$

The cross-variogram is zero at the origin ($\gamma_{ij}(0) = 0$), is symmetric with respect to the indices ($\gamma_{ij}(\mathrm{h}) = \gamma_{ji}(\mathrm{h})$) and may be negative, which occurs with negatively correlated random fields.

It is useful to know about the relation between the cross-variograms and cross-covariance functions in the joint second-order stationary framework. This relation is shown in equation (2.24). It can be seen that the cross-variogram takes the average between the corresponding cross-covariance function for –h and +h:

$$\gamma_{ij}(\mathrm{h}) = C_{ij}(0) - \frac{1}{2}(C_{ij}(\text{-h}) + C_{ij}(\text{+h})) \quad (2.24)$$

By decomposing the cross-covariance function into an even and an odd function (shown as equation (2.25)), it can be seen that the cross-variogram just recovers the even term of the cross-covariance function (Wackernagel, 2003):

$$C_{ij}(h) = \underbrace{\frac{1}{2}\left(C_{ij}(+h) + C_{ij}(-h)\right)}_{\text{even term}} + \underbrace{\frac{1}{2}\left(C_{ij}(+h) - C_{ij}(-h)\right)}_{\text{odd term}} \quad (2.25)$$

2.2.3.3. Pseudo cross-variogram

Myers (1991) and Cressie (1993) proposed the pseudo cross-variogram as an alternative for the traditional cross-variogram, by calculating the variance of cross increments:

$$\text{var}[Z_i(x+h) - Z_j(x)] = 2\pi_{ij}(h) \quad (2.26)$$

By assuming the expectation of the cross increments is zero, $E[Z_i(x+h) - Z_j(x)] = 0$, the pseudo cross-variogram can be calculated from

$$\pi_{ij}(h) = \frac{1}{2} E[\left(Z_i(x+h) - Z_j(x)\right)^2] \quad (2.27)$$

The advantage of this function is not being even. But it has some drawbacks, in particular the assumption of stationary cross increments does not look realistic, because it may not make sense to take the difference between two variables measured in different units (the difference between the two variables does not have a physical meaning) and it is difficult to interpret. According to the mentioned drawbacks, the pseudo cross-variogram is hardly ever used in applications (Wackernagel, 2003).

2.2.3.4. Fitting a variogram model

The experimental cross-variogram between two variables for a given lag separation vector h can be estimated by using all pairs of data points where both variables are known, using the following unbiased estimator:

$$\hat{\gamma}_{ij}(h) = \frac{1}{2|N(h)|}\sum_{N(h)}\left[z_i(x_\alpha) - z_i(x_\beta)\right]\left[z_j(x_\alpha) - z_j(x_\beta)\right] \quad (2.28)$$

where N(h) is the number of pairs of data locations a vector h apart (with some calculation tolerances). Data locations where only one of the variables is present are simply ignored in calculating the cross-variogram (Chilès and Delfiner, 2012).

2.2.3.4.1. Sampling designs

The available measurements for different variables Z_i in a given domain may be located at the same sample points or at different points for each variable, as shown on Figure 2.11. The following situations can be distinguished (Wackernagel, 2003):

- *entirely heterotopic data*: the variables are measured on different sets of sample points and there is not any sample location in common;
- *partially heterotopic data*: some variables share some sample locations;
- *isotopy*: all the variables are known at all sampling points.

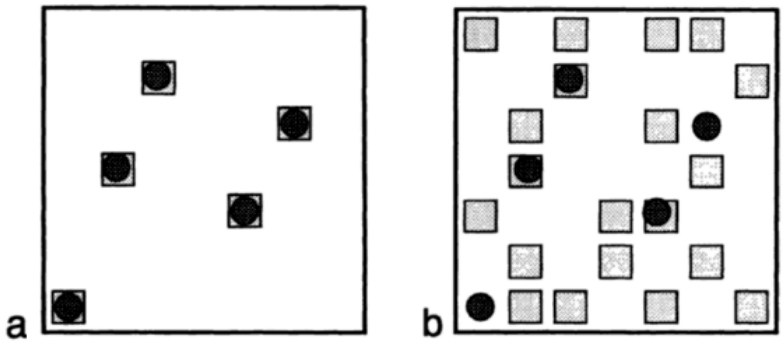


Figure 2.11: Isotropic data, sample sites are shared (a), partially heterotopic data, sample sites may be different (b) (Wackernagel, 2003).

It should be considered that the experimental cross-variogram cannot be calculated in the case of an entirely heterotopic sampling, or when most of the sampling data are heterotopic. So in this case, one has to use experimental cross-covariances or pseudo cross-variograms (Chilès and Delfiner, 2012).

2.2.3.4.2. Linear model of coregionalization

As it was explained about the necessity of fitting a variogram model in the univariate case, theoretical cross-variograms should be fitted on the experimental cross-variograms too. The major difference here is about the complexity of the process for the multivariate case. There exist many modeling strategies to fit the direct and cross-variograms. In particular, the linear model of coregionalization (Wackernagel, 2003; Chilès and Delfiner, 2012) is widely used in practice, for its versatility and simplicity of use.

This model extends the concept of nested structures to the multivariate context. Generally, the problem of fitting a suitable model includes three main aspects: (i) choosing the appropriate nested structures, (ii) estimating the parameters of the model, and (iii) meeting the requirements for the consistency of the multivariate model. In this respect, a sufficient condition is that, for each nested structure, the sill matrix is symmetric positive semi-definite, i.e., that its eigenvalues are non-negative (Journel and Huijbregts, 1978; Wackernagel, 2003).

It is implicitly assumed that all the regionalized variables being studied are generated by a same set of physical processes acting additively at different spatial scales. So, in the linear model of coregionalization, the form of all the covariances (variograms) and cross-covariances (cross-variograms) is that of a nested model composed of the same elementary covariance (variogram) functions, just weighted by specific coefficients. The parameters of the nested model should be estimated by fitting simultaneously the covariances (variograms) and cross-covariances (cross-variograms).

Under this model, a second-order stationary vector random field Z = {Z_i: i = 1, …, N} can be decomposed into a set of S spatially uncorrelated components $\{Z_u : u = 1, \dots, S\}$ and a mean value, such that:

$$Z_i(\mathrm{x}) = \sum_{u=1}^{S} Z_u^i(x) + m_i \tag{2.29}$$

where one has the following equations for all values of the indices i, j, u and v,

$$\begin{cases} E[\mathrm{Z}_i(\mathrm{x})] = m_i & (2.30) \\ E\left[Z_u^i(x)\right] = 0 & (2.31) \\ cov\left(Z_u^i(x)\right), \left(Z_u^j(x+h)\right) = E\left[Z_u^i(x)\, Z_u^j(x+h)\right] = C_{ij}^u(h) & (2.32) \\ cov\left(Z_u^i(x)\right), \left(Z_v^j(x+h)\right) = 0 \qquad when \qquad u \neq v & (2.33) \end{cases}$$

The cross-covariance functions $C_{ij}^u(h)$ associated with the spatial components can be written as the products of real coefficients b_{ij}^u and real-valued correlation functions $\rho_u(\mathrm{h})$:

$$C_{ij}(\mathrm{h}) = \sum_{u=1}^{S} C_{ij}^u(h) = \sum_{u=1}^{S} b_{ij}^u\, \rho_u(h) \tag{2.34}$$

A set of N×N coregionalization matrices B_u can be set up, so one will have a multivariate nested covariance function model C(h) with symmetric, positive semi-definite matrices B_u (known as coregionalization matrices), for u = 1… S:

$$C(h) = \sum_{u=1}^{S} B_u \, \rho_u(h)$$

Similarly, the multivariate nested variogram model associated with a intrinsically stationary random fields is

$$\Gamma(h) = \begin{pmatrix} \gamma_{11}(h) & \cdots & \gamma_{1N}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{1N}(h) & \cdots & \gamma_{NN}(h) \end{pmatrix} = \sum_{u=1}^{S} B_u \, g_u(h) \qquad (2.35)$$

where the $g_u(h)$ are basic variogram models with given ranges and with unit sills and the B_u are symmetric, positive semi-definite symmetric sill matrices i.e., with non-negative eigenvalues (Wackernagel, 2003).

2.2.3.4.3. Manual or semi-automatic fitting

By increasing the number of variables to three or more, checking the positive semi-definiteness condition is difficult (one has to calculate the eigenvalues of each coregionalization matrix in order to verify that none of them is negative). There exist automatic sill fitting algorithms to ensure the positivity of the coregionalization matrices, for which the user has to propose the set of basic variogram models, in particular, their number (S), types and ranges (Goulard, 1989; Goulard and Voltz, 1992; Emery, 2010a; Desassis and Renard, 2011). The only difference with the univariate case (linear model of regionalization) is that the scalar sill of each elementary structure is replaced by a positive semi-definite matrix of sills.

2.2.3.5. Coregionalization analysis

In the second-order stationary model, let us consider the spectral decomposition of the variance-covariance matrix of the vector random field $Z = \{Z_i: i = 1, \ldots, N\}$ at h=0:

$$C(0) = ADA^T \qquad (2.36)$$

where *A* is an orthogonal matrix (eigenvectors) and *D* is a diagonal matrix with non-negative entries (eigenvalues). Also one can define a vector of N factors (principal components), *Y*, as (Wackernagel, 2003):

$$Y(x) = A^T Z(x) \qquad (2.37)$$

Such principal components have no cross-correlation at lag h=0. However, they may be cross-correlated at other lag vectors, unless specific conditions (in particular, when the linear model of coregionalization contains only one nested structure, i.e., S = 1).

Coregionalization analysis (Goovaerts, 1992; Wackernagel, 2003) consists in decomposing the vector random field Z into S sets of spatially uncorrelated vector random fields Z_1, …, Z_S, whose respective covariance matrices at lag vector h are $B_1\ \rho_1(h), \ldots, B_S\ \rho_S(h)$ as defined in equation (2.29):

$$Z(x) = \sum_{u=1}^{S} Z_u\,(x) + m \qquad (2.38)$$

Then each of these vector random fields is factorized through principal component analysis, by putting:

$$\forall u \in \{1, \ldots, S\}, Y_u(x) = A_u^T Z_u(x) \qquad (2.39)$$

where A_u is the orthogonal matrix of eigenvectors of B_u. Because the vector random fields Z_1, …, Z_S are spatially uncorrelated and because their respective coregionalization models contain a single nested structure, the components of Y_1, …, Y_S have no spatial cross correlation. So the components of the initial vector random field Z are decomposed into N×S spatially orthogonal factors (the components of Y_1, …, Y_S) (Emery and Peláez, 2012). These factors can be interpolated in space, through cokriging or simulation (Wackernagel, 2003; Larocque et al., 2006).

2.2.4. Spatial prediction

2.2.4.1. Kriging

The basis for the kriging framework is to predict a random field Z at a target location x by weighting the known values $Z(x_\alpha)$ at surrounding locations $\{x_\alpha: \alpha = 1, \ldots, n\}$, so that the expectation of the prediction error is zero and the prediction error variance is minimized. It is a probabilistic approach based on the random field model, in which the weighting of the data is determined according to the distances between the data and the target location, the redundancies between the data and the spatial continuity of the random field given by its variogram model. There are many flavors of kriging, but the basic forms differ mostly on the assumptions they make regarding the stationarity assumption and the knowledge of the mean value. This is expressed as conditions on the set of weights. The more common types of kriging are the following (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1989; Chilès and Delfiner, 2012; Rossi and Deutsch, 2014):

- *Simple kriging (SK)*, in which the mean m is known and often considered as a constant (second-order stationarity assumption), so that it can be inferred from the available samples in the entire domain. The predictor of the random field Z at location x_0 is:

$$Z^*(x_0) = m + \sum_{\alpha=1}^{n} \lambda_\alpha (Z(x_\alpha) - m) \quad (2.40)$$

where λ_α are weights attached to the residuals $Z(x_\alpha) - m$ and are obtained by solving the following system of linear equations (Wackernagel, 2003; Chilès and Delfiner, 2012):

$$\sum_{\beta=1}^{n} \lambda_\beta C(x_\alpha - x_\beta) = C(x_\alpha - x_0) \quad (2.41)$$

The left-hand side of the equations contains the covariances between data locations. The right-hand side contains the covariance between each data location and the location where a prediction is sought. When x_0 gets far from the data locations, the weights tend to be of small magnitude, so that the simple kriging predictor gets closer to the known mean value.

- *Ordinary kriging (OK)*, for which the random field is either second-order stationary with an unknown mean value or intrinsically stationary. Considering the mean value as unknown allows generalizing the predictor to situations where this mean is not constant in space: the mean can vary from one region to another, provided that it remains (approximately) constant at the scale of the kriging neighborhood (see next subsection). The predictor of the random field Z at location x_0 is:

$$Z^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha^{OK} (Z(x_\alpha)) \quad (2.42)$$

where the weights are obtained by solving the following system of linear equations (Wackernagel, 2003; Chilès and Delfiner, 2012):

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_\beta^{OK} \gamma(x_\alpha - x_\beta) + \mu_{OK} = \gamma(x_\alpha - x_0) \\ \sum_{\beta=1}^{n} \lambda_\beta^{OK} = 1 \end{cases} \quad (2.43)$$

where λ_β^{OK} is the weight assigned to the data located at x_β and μ_{OK} is a Lagrange multiplier.

- *Kriging with a trend model*, for which the mean value m(x) varies in space, which reflects the presence of a systematic trend or "drift" in the spatial distribution of the

random field. This variant is also called non-stationary kriging because of the location-dependent mean, often modeled as a polynomial function of the spatial coordinates (*universal kriging*) or scaled from a secondary exhaustively known variable (*kriging with an external drift*).

The choice of the kriging method depends on the geological setting, the amount of available information, and the characteristics of the random field model envisioned. The most commonly used method is ordinary kriging (Rossi and Deutsch, 2014), as it allows more versatility in the model: the mean value can be constant at the scale of the local neighborhood, but variable at the scale of the entire field, a situation known as "local stationarity".

In addition to a prediction of the random field at any target location, kriging provides a measure of precision, through the variance of the prediction error (known as the "kriging variance"). This variance depends on geometric information (data and target locations) and on variogram information (nugget effect, range, sill, anisotropy…) but not on the actual data values.

As an illustration, Figure 2.12 shows an example of simple and ordinary kriging predictions and error standard deviations (square root of kriging error variances) for cobalt concentration data distributed over a region of about 4×5 km^2 (Emery, 2017).

2.2.4.2. Kriging neighborhood

In practice, the data selected for kriging are quite often a subset of all the available data (those located in a "moving neighborhood" around the target location), as using all the available data ("unique neighborhood") may make the computation time prohibitive, or increases this computation time without significantly improving the precision of the predictor.

The loss of precision is small when the moving neighborhood discards the data that would receive small weights under a unique neighborhood (in general, these are the data far from x_0) and/or the data redundant with (in particular, clustered with) non-discarded data (Emery, 2009).

The shape of the neighborhood should account for the anisotropy in the spatial correlation of the data. In general, the neighborhood is chosen as an ellipse (2D) or an ellipsoid (3D), which theoretically corresponds to the case of a geometric anisotropy. To improve the distribution of the data around the target location, it is common to split the neighborhood into angular sectors and to look for data in each sector (Figure 2.13).

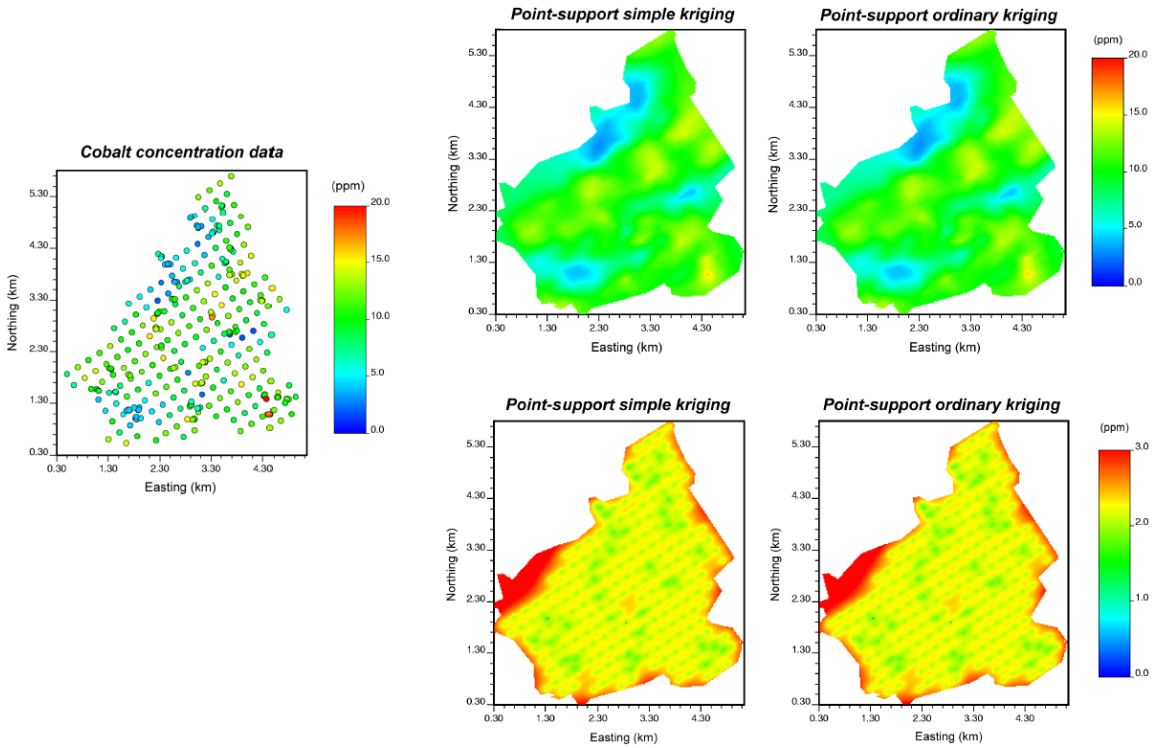


Figure 2.12: An example of simple and ordinary kriging predictions (right, upper part) and simple and ordinary kriging error standard deviations (right, down part) for cobalt concentration data (left) (Emery, 2017).

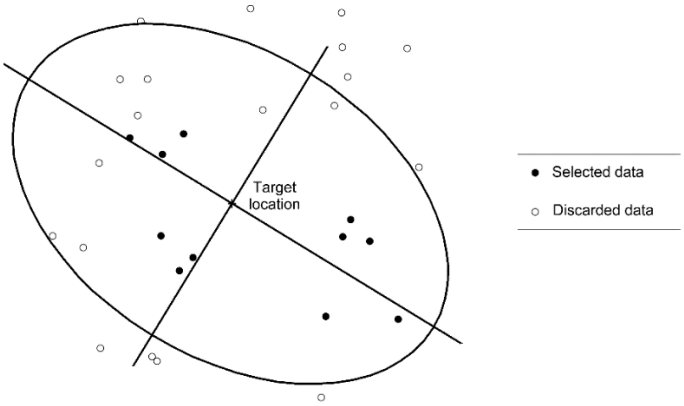


Figure 2.13: An example of moving neighborhood for kriging with up to 3 data per quadrant (Emery, 2017).

2.2.4.3. Cokriging

Cokriging is the multivariate version of kriging where one predicts a variable by accounting for data of this variable and of other correlated covariates. In other words, cokriging aims to perform the joint prediction of several coregionalized variables, taking into account their spatial dependences. Cokriging is advantageous when the target variable (primary variable) is under-sampled with respect to other cross-correlated variables (covariates or secondary variables). For example, the copper grade (primary variable) may be estimated from a combination of copper, gold and molybdenum (Au and Mo are secondary variables) samples values. For cokriging, there must be a spatial correlation between the primary and secondary variables that can be inferred from available information. As is the case when considering a single variable, there are several variants of cokriging, such as *simple cokriging* (SCK), *ordinary cokriging* (OCK), and *universal cokriging* (UCK). Conceptually these methods are the same as the ones explained with kriging; however, there is the additional complication of dealing with at least two variables. For example, cokriging requires a variogram (covariance) model for each variable, as well as the cross variograms (covariances) between the different pairs of variables, in order to measure the spatial cross-correlations between these variables (Rossi and Deutsch, 2014).

- Simple cokriging

This variant relies on the knowledge of the mean values $\{m_i: i = 1\dots N\}$ and direct and cross-covariances $\{C_{ij}: i = 1\dots N, j = 1\dots N\}$ of the N variables under study. In the isotopic case, the predictor of the entire set of variables at location x_0 can be written as (Chilès and Delfiner, 2012):

$$\boldsymbol{Z}^*(x_0) = \boldsymbol{a} + \sum_{\alpha=1}^{n} \boldsymbol{\Lambda}_\alpha^T \boldsymbol{Z}(x_\alpha) \tag{2.44}$$

where $\boldsymbol{a}$ is a $N \times 1$ vector and $\{\boldsymbol{\Lambda}_\alpha, \alpha = 1 \dots n\}$ are $N \times N$ weight matrices defined by:

$$\boldsymbol{a} = \left(\boldsymbol{I} - \sum_{\alpha=1}^{n} \boldsymbol{\Lambda}_\alpha^T\right)\boldsymbol{m}$$

$$\begin{pmatrix} \mathbf{C}(x_1 - x_1) & \cdots & \mathbf{C}(x_1 - x_n) \\ \vdots & \ddots & \vdots \\ \mathbf{C}(x_n - x_1) & \cdots & \mathbf{C}(x_n - x_n) \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \vdots \\ \boldsymbol{\Lambda}_n \end{pmatrix} = \begin{pmatrix} \mathbf{C}(x_1 - x_0) \\ \vdots \\ \mathbf{C}(x_n - x_0) \end{pmatrix} \tag{2.45}$$

$\boldsymbol{Z}$ the $N \times 1$ vector whose generic entry is z_i, $\boldsymbol{m}$ the $N \times 1$ vector whose generic entry is m_i, and $\mathbf{C}(x_\alpha - x_\beta)$ the $N \times N$ matrix whose generic entry is $\mathrm{C}_{ij}(x_\alpha - x_\beta)$.

In the heterotopic case, the rows and columns associated with missing data values in equation (2.45) should be removed to calculate the weight matrices.

- Ordinary cokriging

This variant relies only the knowledge of the direct and cross-covariances or direct and cross-variograms $\{\gamma_{ij}: i = 1\dots N, j = 1\dots N\}$ of the N variables under study, while the mean values are assumed unknown. In the isotopic case, the predictor of the entire set of variables at location x_0 can be written as (Chilès and Delfiner, 2012):

$$\boldsymbol{Z}^*(x_0) = \sum_{\alpha=1}^{n} \boldsymbol{\Lambda}_\alpha^T \boldsymbol{Z}(x_\alpha) \tag{2.46}$$

where $\{\boldsymbol{\Lambda}_\alpha, \alpha = 1 \dots n\}$ are $N \times N$ weight matrices defined by:

$$\begin{pmatrix} \boldsymbol{\Gamma}(x_1 - x_1) & \cdots & \boldsymbol{\Gamma}(x_1 - x_n) & \mathbf{I} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Gamma}(x_n - x_1) & \cdots & \boldsymbol{\Gamma}(x_n - x_n) & \mathbf{I} \\ \mathbf{I} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda_1} \\ \vdots \\ \boldsymbol{\Lambda_n} \\ -\mathbf{M} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Gamma}(x_1 - x_0) \\ \vdots \\ \boldsymbol{\Gamma}(x_n - x_0) \\ \mathbf{I} \end{pmatrix} \tag{2.47}$$

$\boldsymbol{\Gamma}(x_\alpha - x_\beta)$ is the $N \times N$ matrix whose generic entry is $\gamma_{ij}(x_\alpha - x_\beta)$,

M is a matrix of Lagrange multipliers, **I** the identity matrix, and **0** the zero matrix.

Again, in the heterotopic case, the rows and columns associated with missing data values in equation (2.47) should be removed to calculate the weight matrices.

Same as kriging, the variances of cokriging errors depend on geometric information (data and target locations) and on variogram information but they do not depend on the data values.

For a particular variable, cokriging improves the results of kriging, i.e., it yields a smaller error variance, when the variable is under-sampled with respect to other cross-correlated variables (heterotopic sampling). This case is advantageous when secondary variables are more accessible or less expensive to be sampled than the primary variable of interest (Wackernagel, 2003; Chilès and Delfiner, 2012).

2.2.4.4. Cokriging neighborhood

Cokriging with many variables and data generates a very large linear system to solve, which can be time consuming or prohibitive. This urges the choice of a small subset of data located around the target location, called a neighborhood or a plan, as a crucial step in cokriging.

Figure 2.14 shows three different neighborhoods for a given central target location (denoted by a star), primary data (denoted by full circles) as well as three alternate subsets of data from a secondary variable (denoted by squares). The neighborhood (A) which can be termed as the full neighborhood, uses all data available for the secondary variable. The neighborhood

(C), called a collocated neighborhood by Xu et al. (1992), uses the secondary data only at the location targeted for prediction, whereas the neighborhood (B) is called a multi-collocated neighborhood by Chilès and Delfiner (2012) and restricts the secondary information to the subset of locations where primary data is available as well as to the target location (Wackernagel, 2003). It has been observed that the latter option (multi-collocated, B) often yields predictions that are almost as precise as the ones obtained with the full neighborhood (A). In contrast, the collocated neighborhood (C) loses information and provides poorer predictions.

In a more general context where the secondary variable is not exhaustively known, it is advisable to separately search for data of each (primary and secondary) variable, in order to avoid omitting a relevant data for cokriging (Madani and Emery, 2018).

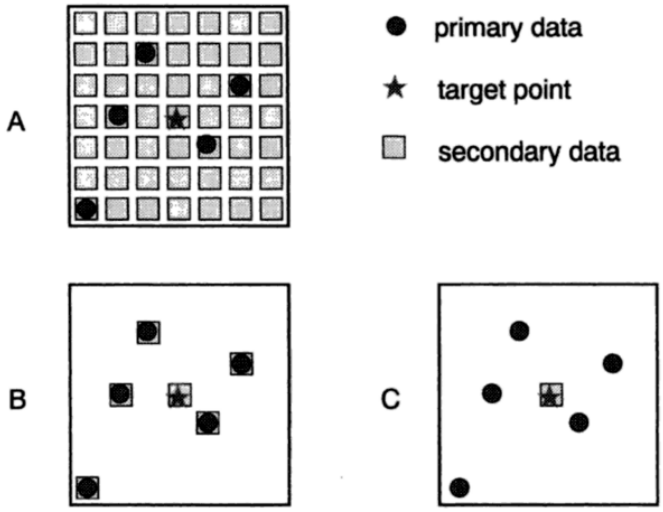


Figure 2.14: Three examples of neighborhoods (A) full neighborhood, (B) multi-collocated neighborhood, and (C) collocated neighborhood (Wackernagel, 2003).

2.2.4.5. Leave-one-out cross-validation and split-sample jack-knife

The many interdependent subjective decisions in a geostatistical study make somehow necessary to validate the entire geostatistical model and (co)kriging plan prior to any production run. The objective of leave-one-out cross-validation is to compare and validate two or more alternative theoretical variogram models, or alternative types of (co)kriging (ordinary kriging, universal kriging, etc.), or alternative (co)kriging neighborhoods (Deutsch and Journel, 1998). In this method, one data is removed at each time from the data set and is predicted by using the remaining data. To avoid inappropriate effects of the nearest datum in a drill hole, one or more of the closest data can be removed in the prediction exercise. Then one calculates the prediction error (true value minus predicted value) at each data location

and studies the quality of the prediction errors by means of statistical and graphical tools (Figure 2.15). As a complement, one can also examine the standardized errors (errors divided by their standard deviations).

In split-sample jack-knife techniques, a subset of the data (say, 40 or 50 % of the total) is removed completely from the data set, and the prediction is done using the remaining data. Jack-knife is a more interesting alternative as long as there is sufficient data to obtain a statistically meaningful set of errors (Rossi and Deutsch, 2014).

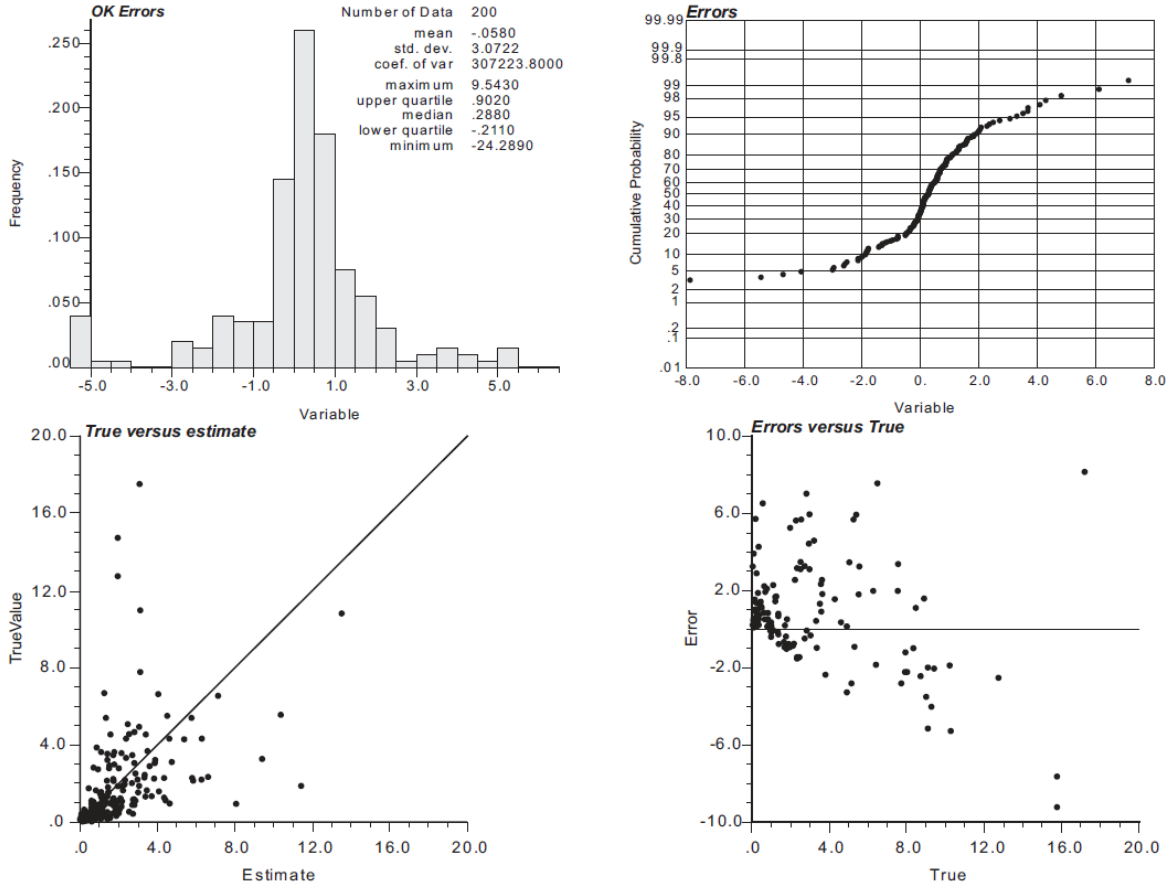


Figure 2.15: Error checks. Histogram of errors, cumulative distribution function of error, scatter diagram of true value vs. prediction, and scatter diagram of error vs. true values (Rossi and Deutsch, 2014).

2.2.5. Geostatistical simulation

2.2.5.1. Simulation versus prediction

Kriging provides a value that is, on average, as close as possible to the actual (unknown) value, according to some definition of goodness or quality (zero mean error and minimal error variance). The map of such best local predictions, however, may not be best as a whole. Kriging, like most other interpolation methods, has an unavoidable smoothing effect, which makes it useful for visualizing trends, but not local details. It does not reproduce the spatial variability of the variable, which can cause significant biases when nonlinear responses are of interest (Chilès and Delfiner, 2012; Rossi and Deutsch, 2014).

Another drawback of prediction is that the smoothing is not uniform; actually it depends on the number and configuration of the local data. Smoothing is minimal close to the data locations and increases as the target location gets farther away from data locations. A map of kriging predictions appears more variable in densely sampled areas than in sparsely sampled areas, so the kriged map may show artifact structures (Goovaerts, 1997).

This is where simulation comes into play. Unlike predicted models, simulated models (called "outcomes" or "realizations") aim to reproduce the true spatial variability and, also, at providing a model of uncertainty at every target location or jointly over several locations (Dimitrakopoulos, 1997). In particular, the extreme values of the original distribution are preserved and are not eliminated like in the case of prediction and its smoothing effect (Figure 2.16).

Simulated models can be used for several purposes, such as (Goovaerts, 1997; Rossi and Deutsch, 2014):

(i) risk analysis, by examining the most optimistic and the most pessimistic realizations;

(ii) prediction, by averaging the realizations;

(iii) estimation of the probability of an event, by calculating the frequency of occurrence of this event among the realizations;

(iv) assessment of uncertainty, by checking how different are the realizations.

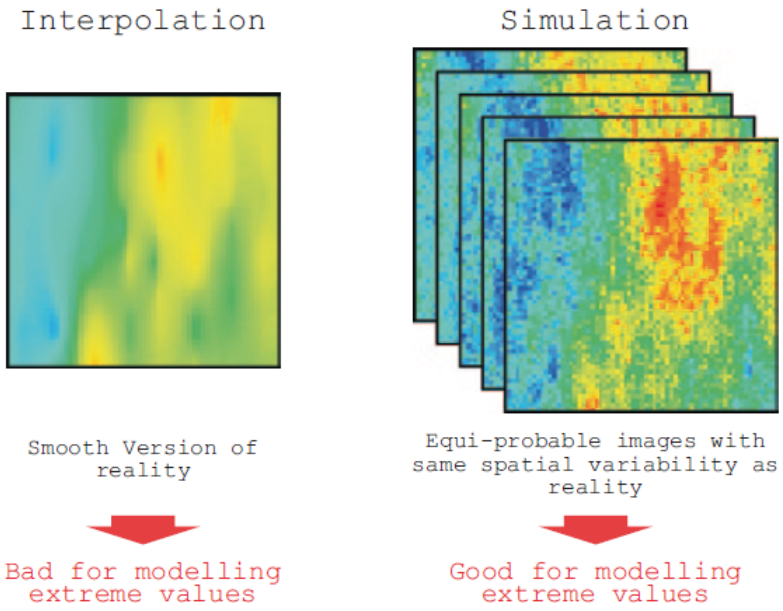


Figure 2.16: Comparison of predicted and simulated models (Rossi and Deutsch, 2014).

2.2.5.2. Principles of simulation

The formalism of random fields involves a family of alternative realizations similar in their spatial variability to the observed reality but different otherwise. Simulation aims to construct these realizations to visualize heterogeneity, to reproduce the true variability of the variable under study and to quantify spatial uncertainty, each realization representing a possible outcome or "scenario".

Simulation relies on the interpretation of the regionalized variable as one realization of a random field and consists in constructing other realizations of this random field. Then it can be computed a result for each realization and the statistical distribution of the results can be studied. There are three main challenges here: (1) defining the spatial distribution of the random field, insofar as knowing just the mean value and the variogram is not sufficient (Eq. 2.3), (2) drawing realizations of the random field (Leuangthong et al., 2008; Chilès and Delfiner, 2012), and (3) conditioning these realizations to the available information, i.e., forcing the realizations to reproduce the data values observed at the data locations (concept of *conditional* simulation).

In practice, high-order distributions cannot be defined from a limited set of sampling data, so a parametric model or some simplifying assumption is required. There are few distributions which can be parameterized simply, the multivariate Gaussian distribution being one of them, which is remarkable in its tractability and is the most common random field for representing regionalized variables measured on continuous quantitative scales. About random fields with

discrete or categorical distributions, truncated Gaussian and plurigaussian models are among the most commonly used, as they are based on the multivariate Gaussian model.

Preferably, a large number of realizations are needed to better describe the conditional cumulative distribution functions at unsampled locations, thus to measure the uncertainty at these locations. However, according to the practical limitations, a smaller number is generally used. Based on authors' experience, between 20 and 100 realizations are generally sufficient to characterize the range of possible values for the unsampled locations (Rossi and Deutsch, 2014).

Each realization constitutes a possible outcome, so that one can work on it as if it were the reality. Accordingly, each realization provides an unbiased response to the problem under consideration. And, with a set of realizations, one obtains a set of responses that reflect the uncertainty on the true unknown response. In particular, one can determine the most favorable response (best case), the least favorable response (worst case), and the average response (average case).

Geostatistical simulation has become popular especially in the mining industry as a tool that provides models of uncertainty at different spatial scales and different stages of a mining project. It is often built on fine grids, fine enough to provide a sufficient number of nodes within the block size of interest. The vertical resolution of the grid should be a function of the support data, for example in modeling a variable mined by open pit, the size of the mining bench is considered. Sometimes larger grid sizes should be used, because of the limitations in processing ability of the computers (Rossi and Deutsch, 2014). Simulation can be applied in related fields of ore body modeling with different aims and purposes, such as: grade control tools in daily operations (Rossi, 1999; Verly, 2005), to analyze risk related to resource classifications (Rossi and Camacho, 2001; Emery et al., 2006; Dimitrakopoulos et al., 2009), to assess the uncertainty of minable reserves at the project feasibility stage (Guardiano et al., 1995; Glacken, 1996; Journel and Kyriakidis, 2004; Leuangthong et al., 2006), mine planning (Jewbali and Dimitrakopoulos, 2009; Contreras et al., 2010), analysis of financial risk of a mining project (Dimitrakopoulos, 2009; Godoy, 2009), and to assess mineralization potential in certain settings. Other applications include assessment of recoverable reserves, drill hole spacing optimization studies and sampling design.

2.2.5.2.1. Reproducing model statistics

As explained previously, instead of a map of local best predictions, simulation generates maps of realizations, say, $\{z^{(l)}(x): x \in D\}$ with l denoting the l-th realization, which reproduce the statistics supposed most consequential for the problem in hand. Typical requisites for such simulated maps are as follows (Goovaerts, 1997):

1. The realization is conditional to the data values:

$$z^{(l)}(x) = z(x_\alpha) \qquad \forall\, x = x_\alpha\,,\ \alpha = 1, \ldots, n \tag{2.48}$$

2. The histogram of simulated values reproduces closely the sample histogram.

3. The covariance model C(h) or, better, the set of indicator covariance models $C_I(h; z_k)$ for various thresholds are also reproduced.

More complex features, such as the spatial correlation with a secondary attribute or high-order statistics, may also be reproduced. The set of alternative realizations $\{z^{(l)}(x): x \in D, l = 1, \ldots, L\}$, which meets the mentioned requisites, provides a visual and quantitative measure of spatial uncertainty.

2.2.5.2.2. Non-conditional vs. conditional simulation

Non-conditional simulation just aims to construct realizations with the same variability as the variable under study, but without reproducing the data values at the data locations. So different realizations have, up to statistical fluctuations, the same histogram, same variogram, etc., but not the same values (even at data locations). However, to be realistic, each realization should also reproduce the known values at the data locations. The simulation is said to be conditional when it restitutes the data values at the sampling locations $\{x_\alpha: \alpha = 1, \ldots, N\}$.

Geostatistical conditional simulation reproduces the posterior spatial distribution of the regionalized variable of interest, i.e., the distribution conditional to the available data (Rossi and Deutsch, 2014). In particular, far from the conditioning data, posterior and prior (non-conditional) distributions are the same, whereas, at a data location, there is no uncertainty and one obtains the data value in all the realizations.

2.2.5.3. Multi-Gaussian model

2.2.5.3.1. Definition and key properties

The probability density function of the univariate Gaussian or normal distribution is:

$$g(z;\, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \tag{2.49}$$

where μ and σ^2 are the mean or expectation and the variance of the distribution. The so-called "standard normal distribution" is given by taking $\mu = 0$ and $\sigma^2 = 1$ in the general normal distribution.

A random field is assumed to be Gaussian if it fulfills the following equivalent properties:

- every linear combination of the variables at given locations $\{x_1, \dots, x_n\}$ has a Gaussian distribution;
- the joint probability density function of the random vector $Z = (Z(x_1), \dots, Z(x_n))^t$ is:

$$g(x_1, \dots, x_n; z_1, \dots, z_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\sum)}} \exp\{-\frac{1}{2}(z-\mu)^t \sum^{-1} (z-\mu)\} \qquad (2.50)$$

with $z = (z_1, \dots, z_n)^t$, μ is the n-dimensional mean vector of Z, and $\sum$ is the n×n variance-covariance matrix of Z.

In this model, the prior distribution of Z(x) is standard Gaussian. Given the values at some locations, the posterior or conditional distribution of Z(x) is still Gaussian, with mean equal to the simple kriging prediction of Z(x) and with variance equal to the simple kriging variance (Chilès and Delfiner, 2012):

{Z(x) | data} = $Z^{SK}(x) + \sigma^{SK}(x)\, U(x)$ with U(x) ~ N(0,1) independent of data.

2.2.5.3.2. Gaussian transformation or anamorphosis

In practice, the regionalized variables under study do not have a Gaussian univariate distribution, so that the direct fitting of a multi-Gaussian model is inadequate. However, the univariate distribution can be transformed into a standard Gaussian in order to become consistent with the multi-Gaussian model, a process known as Gaussian transformation or anamorphosis (Figure 2.17) (Chilès and Delfiner, 2012).

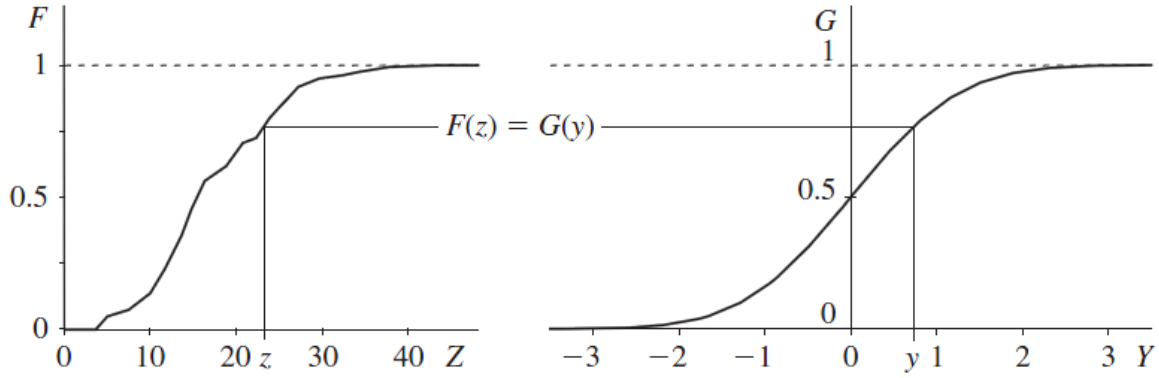


Figure 2.17: An example of graphical Gaussian transformation (Chilès and Delfiner, 2012).

The compatibility of higher-order distributions with the multi-Gaussian hypothesis should also be checked. In practice, to validate this assumption, one can look at the bivariate distributions, i.e., the distributions of pairs of data, via the examination of lagged scatter plots, indicator variograms, madograms and variograms of order less than 2 (Goovaerts, 1997; Emery, 2005).

2.2.5.3.3. Mahalanobis distance

In equation (2.50), $\sqrt{(z-\mu)^t \sum^{-1}(z-\mu)}$ is called the Mahalanobis distance, which gives a measure of the distance between a test point of z from the mean μ. It should be considered that, in the one-dimensional case, the distribution reduces to a univariate normal distribution and the Mahalanobis distance reduces to the standard score. Actually this distance is a multi-dimensional generalization of the idea of measuring how many standard deviations away z is from the mean μ, and it grows as z moves away from the mean. The Mahalanobis distance is dimensionless and scale-invariant, and takes into account the correlations of the data set.

If $\sum$= I (identity matrix), then the Mahalanobis distance changes to $\sqrt{(z-\mu)^t (z-\mu)}$ which is the Euclidean distance.

2.2.5.3.4. Geometry of the Gaussian

The geometry of the Gaussian can be understood through the eigen-decomposition of the variance-covariance matrix $\sum$. It should be considered a curve of constant probability, so the exponential part of equation (2.50) should be constant because the other parts are just for normalizing. Thus, the following should be constant:

$$\Delta^2 = (z-\mu)^t \sum^{-1}(z-\mu) \tag{2.51}$$

For knowing Δ^2, it should be considered the eigen-decomposition of the variance-covariance matrix $\sum$:

$$\sum = U \Lambda U^t \tag{2.52}$$

where U is a orthogonal matrix of eigenvectors and Λ is a diagonal matrix of eigenvalues. In the case of two components (bivariate Gaussian distribution), it can be written as:

$$\Sigma = \begin{bmatrix} u_1 & u_2 \\ \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} u_1 & \rightarrow \\ u_2 & \rightarrow \end{bmatrix} \tag{2.53}$$

Then one can define $k_i = u_i^T(z - \mu)$ which is the projection of vector $(z - \mu)$ onto the i-th eigenvector u_i (shown on Figure 2.18). So it can be seen that Δ^2 has a very simple form in terms of these z_i:

$$\Delta^2 = \frac{z_1^2}{\lambda_1} + \frac{z_2^2}{\lambda_2} \tag{2.54}$$

which is precisely the equation of an ellipse with one axis of size $\sqrt{\lambda_1}$ and the other axis of size $\sqrt{\lambda_2}$.

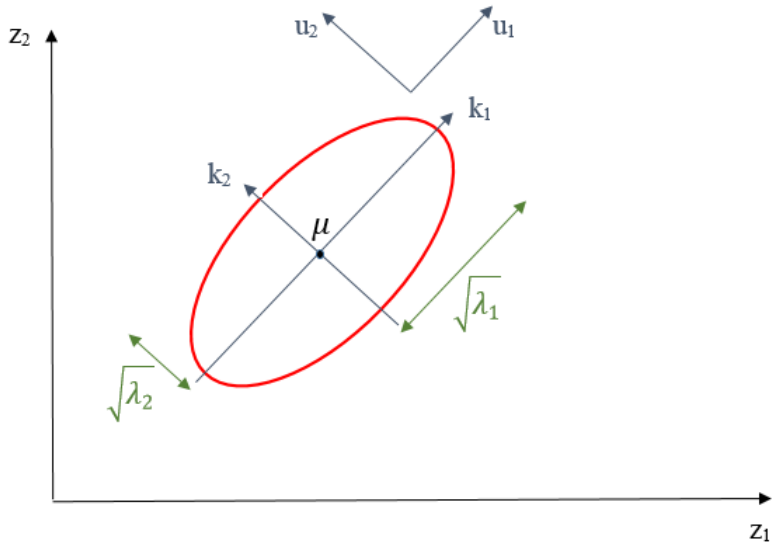


Figure 2.18: Geometry of the Gaussian

So the multi-variate Gaussian distribution can be generalized by:

- The vector mean

- The covariance matrix, which describes the shape (eigenvectors u_i) and the size (eigenvalues λ_i) of the bell-shaped model.

2.2.5.3.5. Multi-Gaussian non-conditional simulation algorithms

- **Sequential Gaussian simulation**

Sequential Gaussian simulation is one of the most widespread algorithms for simulating Gaussian random fields (Ripley, 1987; Isaaks, 1990; Gómez-Hernández and Journel, 1992;

Gómez-Hernández and Cassiraga, 1994; Goovaerts, 1997; Deutsch and Journel, 1998; Emery, 2004, 2010b).

At each step, one simulates the random field at one location and incorporates the simulated value to the set of conditioning data for simulating the random field at the next locations. This method allows directly constructing realizations conditional to existing data, is simple and easy-to-run, and also allows the refining of an existing simulation (i.e., increasing its resolution). But it has several drawbacks, such as being computationally slow and expensive (Goovaerts, 1997; Deutsch and Journel, 1998; Emery, 2004; Chilès and Delfiner, 2012).

- **Matrix decomposition method**

Simulation through the Cholesky decomposition of the covariance matrix provides the fastest solution when the total number of conditioning data plus the number of locations targeted for simulation is small (fewer than a few thousands) and a large number of realizations is requested (Luster, 1985; Alabert, 1987; Davis, 1987). This is an exact algorithm, i.e., it is theoretically correct and without any approximation. Once the Cholesky decomposition of the variance-covariance matrix is obtained, one can generate many realizations very quickly. However, the decomposition is difficult to obtain when the total number of conditioning data plus number of target locations is large (Chilès and Delfiner, 2012).

- **Continuous spectral method**

This method is based on Bochner's theorem, according to which every continuous covariance function in $\mathbb{R}^d$ is the Fourier transform of a positive and integrable measure. It consists in simulating the random field as a mixture of cosine waves with random frequencies (distributed according to the spectral distribution of the covariance) and phases (distributed uniformly). This algorithm is an extremely fast and computationally parallelizable method, and one can simulate as many locations as desired without much computational requirement. But the simulated random field is not exactly multi-Gaussian, although it possesses the correct covariance function. For obtaining a simulation that is approximately multi-Gaussian, one has to sum numerous independent cosine waves (central limit theorem) (Shinozuka, 1971; Mejia and Rodríguez-Iturbe, 1974; Lantuéjoul, 2002; Chilès and Delfiner, 2012).

- **Turning bands**

The turning bands method aims to simplify the simulation in multidimensional spaces, by generating a 3D simulation from several independent 1D simulations along lines that can be rotated in the 3D space (Matheron, 1973; Journel, 1974; David, 1977; Mantoglou and Wilson,

1982; Lantuéjoul, 2002; Emery and Lantuéjoul, 2006; Emery, 2008a). As the continuous spectral method, this algorithm is an extremely fast method, computationally parallelizable, and one can simulate as many locations as desired at a low cost. Also this algorithm exactly reproduces the desired covariance (or variogram). But the simulated random field is not exactly multi-Gaussian. To get close to the multi-Gaussian distribution, it is necessary to consider several hundreds or thousands of lines equi-distributed onto the unit sphere of the 3D space (Figure 2.19).

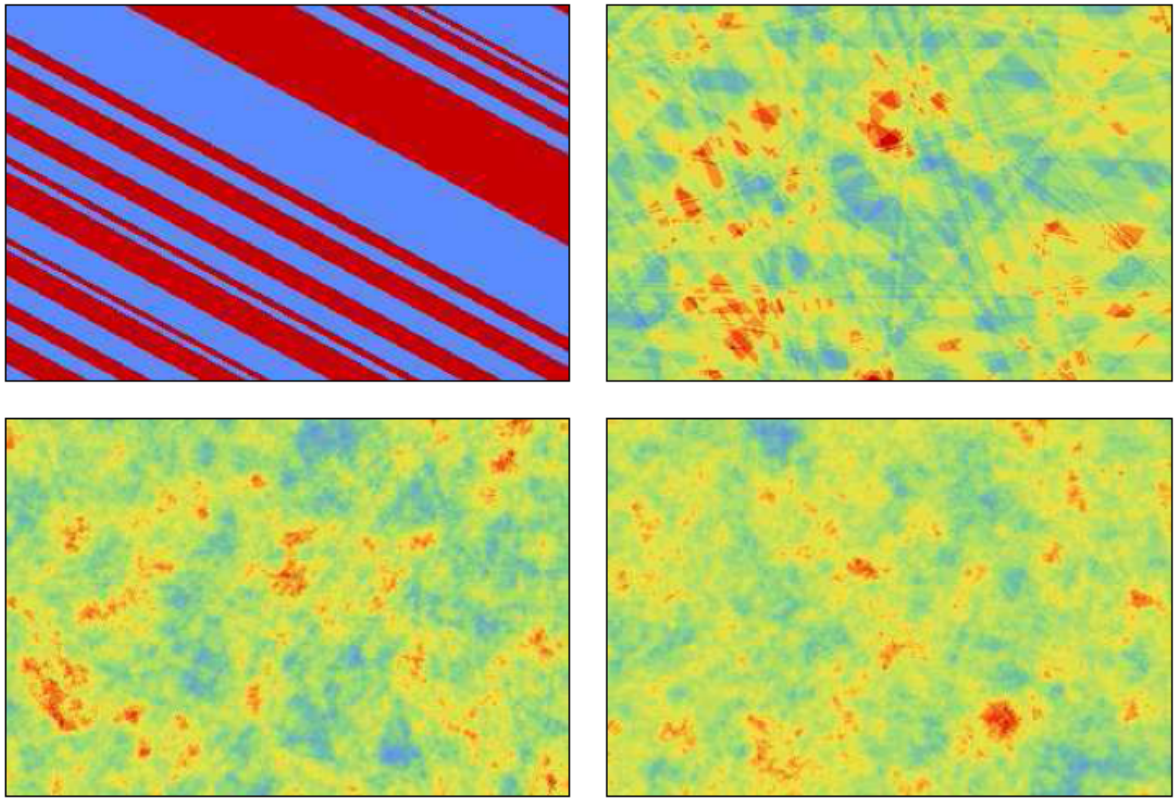


Figure 2.19: Simulations using 1, 10, 100 and 1000 turning bands (Wackernagel, 2013). Multivariate normality is almost reached in the last case

2.2.5.3.6. Conditioning to data

Two simulation algorithms allow directly obtaining conditional simulation: (i) sequential method and (ii) matrix decomposition method. The other algorithms only allow generating non-conditional realizations. However, it is possible to convert a non-conditional realization into a conditional one thanks to an additional step based on simple kriging. Indeed, consider a Gaussian random field Z known at n sample points $\{x_\alpha: \alpha = 1, 2, \ldots, n\}$, let $Z^*(x)$ denote the simple kriging predictor of Z at point x based on the conditioning data $(Z(x_\alpha))$ (Chilès and Delfiner, 2012). It can be written:

$$Z(x) = Z^*(x) + [Z(x) - Z^*(x)] \quad (2.55)$$

true value = kriging predictor + kriging error

Since Z(x) is not known, the kriging error is unknown too. Now it can be considered the same equality for a non-conditional simulation $Z_S(x)$, where $Z_S^*(x)$ is the kriging predictor obtained as if the simulation were known only at the same sampling points x_α:

$$Z_S(x) = Z_S^*(x) + [Z_S(x) - Z_S^*(x)] \quad (2.56)$$

This time, the value ($Z_S(x)$) is known and so the error ($Z_S(x) - Z_S^*(x)$) is known too. Hence the idea of substituting (in 2.47) the unknown error by the simulation of this error, as shown in Figure 2.20 gives us conditional simulation (Chilès and Delfiner, 2012):

$$Z_{CS}(x) = Z^*(x) + [Z_S(x) - Z_S^*(x)] \quad (2.57)$$

where:

$Z^*(x)$ is the simple kriging predictor of Z(x) at the point x based on the conditioning data.

$Z_S(x)$ is a non-conditional simulation at location x.

$Z_S^*(x)$ is the simple kriging predictor obtained as if the non-conditional simulation were known only at the sampling points x_α.

Since kriging is an exact interpolator, at the sampling points x_α one has $Z^*(x_\alpha) = Z(x_\alpha)$ and $Z_S^*(x_\alpha) = Z_S(x_\alpha)$, so $Z_{CS}(x_\alpha) = Z(x_\alpha)$. Conversely, the effect of conditioning decreases and finally vanishes when moving away from the data and the conditional simulation coincides with the non-conditional one.

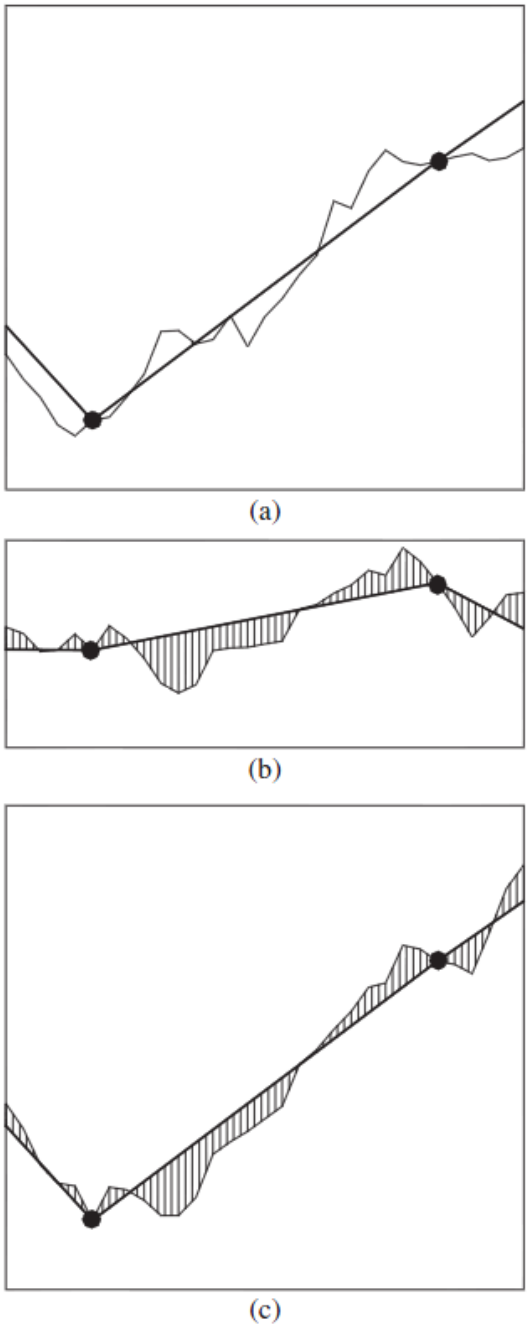


Figure 2.20: Conditioning a simulation: (a) real curve (unknown), sampling points and kriging; (b) non-conditional simulation (known), sampling points, and simulation of the kriging error; (c) kriging errors are picked from the simulation and added to the kriged curve (Chilès and Delfiner, 2012).

In practice, it suffices to perform a single kriging by writing:

$$Z_{CS}(x) = Z_S(x) + [Z^*(x) - Z_S^*(x)] = Z_S(x) + \Delta Z^*(x) \quad (2.58)$$

where $\Delta Z^*(x) = [Z(x) - Z_S(x)]^*$, which is the difference between the kriging prediction of the data and the kriging prediction obtained as if the non-conditional simulation were known only at the conditioning points. This re-expression shows that one can obtain a simulated value by (i) performing a non-conditional simulation to obtain $Z_S(x)$, and (ii) kriging the difference between the non-conditional simulated values at the data locations and the actual data values (Leuangthong et al., 2008).

Some comments about the process of the conditioning should be considered:

- For a given target location x, the kriging weights are the same for all the realizations. Therefore, solving a single kriging system is sufficient for conditioning several realizations.

- Kriging only uses the original conditioning data, not the previously simulated values (this is a parallel simulation paradigm, not a sequential one). So, the calculations can be parallelized.

- Usually, one uses simple kriging with zero mean. This can be replaced by ordinary kriging if one assumes the mean as unknown (Emery, 2008b). The resulting simulation still reproduces the prior variogram.

- In the multivariate case, one just has to replace kriging by cokriging.

2.2.5.4. Plurigaussian model

The plurigaussian model (Galli et al., 1994) has been designed to simulate categorical random fields, i.e., random fields that assume a finite set of values (categories) in space and represent geological domains associated with lithology, alteration or mineralogical assemblage.

This model has been initially used in the petroleum industry to simulate lithological facies in oil reservoirs, but over recent years the mining industry has also started using this approach for simulating rock types or mineralogical domains in ore bodies (Skvortsova et al., 2000, 2002; Deraisme and Field, 2006; Emery and González, 2007a, 2007b; Emery et al., 2008; Armstrong et al., 2011; Yunsel and Ersoy, 2013). Applications have been carried out in a variety of ore deposits such as:

- Granite-hosted uranium deposits (Skvortsova et al., 2000, 2002)
- Roll front uranium deposits (Fontaine and Beucher, 2006)
- Diamond pipes (Deraisme and Field, 2006)
- Porphyry copper deposits (Carrasco et al., 2007; Emery and González, 2007a, 2007b; Emery et al., 2008)

- Nickel laterite deposits (Rondon, 2009)
- Gold deposits (Yunsel and Ersoy, 2011)
- Epithermal lead-zinc deposits (Yunsel and Ersoy, 2013).

Application of the plurigaussian model in the geosciences is not limited to the petroleum and mining industries, as it has been applied in other fields like hydrogeology and environmental sciences (Cherubini et al., 2009; Mariethoz et al., 2009).

The basic idea of the plurigaussian model is to consider two or more Gaussian random fields (usually independent), which are simulated at every location in the study area. Then, using a truncation rule, the Gaussian values are converted into geological domains (Figure 2.21).

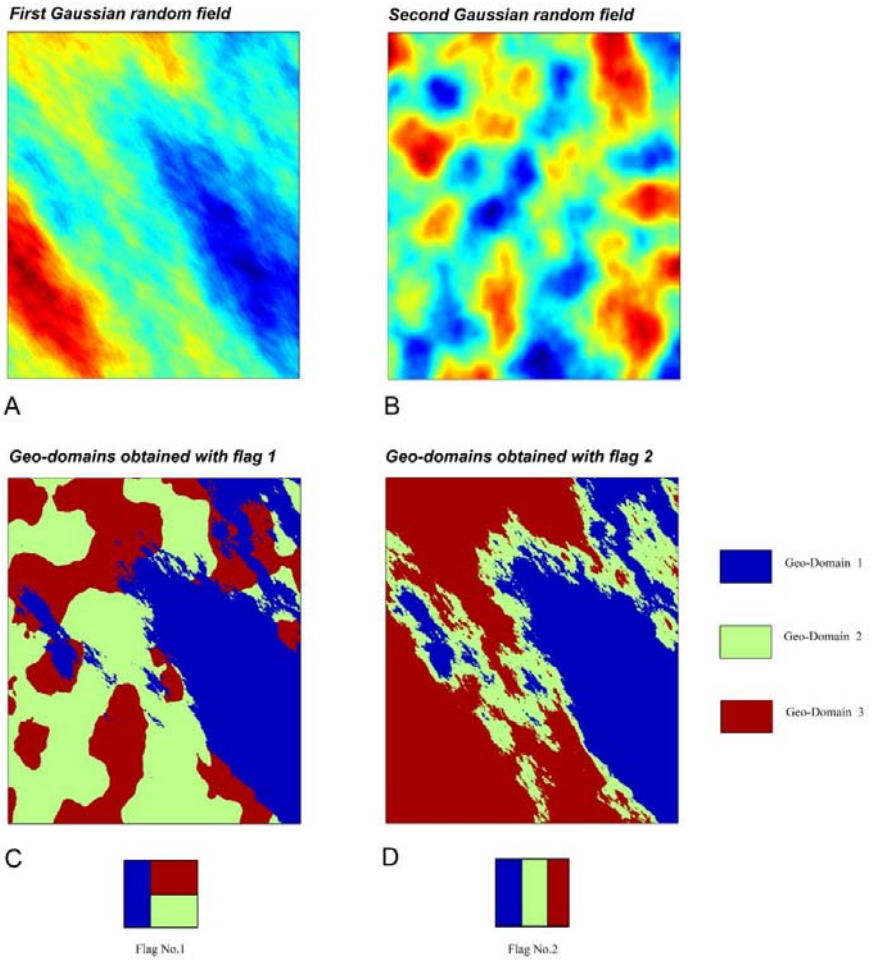


Figure 2.21: Examples of realizations of geological domains (bottom) obtained by truncating two independent Gaussian random fields (top). The truncation rule is represented by a flag below each realization, in which the abscissa axis represents the first Gaussian random field, the ordinate axis represents the second Gaussian random field, and the horizontal and vertical lines represent the truncation thresholds that define the partition of the bi-Gaussian space into geological domains.

2.2.5.4.1. Model parameters

The practical implementation of the model requires defining:

- a truncation rule: it has an impact on the contacts between domains
- truncation thresholds: they have an impact on the proportion of space covered by each domain
- the (direct and cross-)covariances or variograms of the underlying Gaussian random fields: they characterize the spatial continuity of these fields, hence, of the geological domains that will be obtained after applying the truncation rule.

- **Truncation rule**

The truncation rule consists in defining a partition of the Gaussian space, which associates the values of the Gaussian random fields with the geological domains (Le Loc'h et al., 1994; Armstrong et al., 2011; Lantuéjoul, 2002; Dowd et al., 2003; Xu et al., 2006).

Domains that are in contact in space should be touching each other in the truncation rule too (shown on Figure 2.21). And when there is no contact, the truncation rule should represent it, which can be shown by sequential contacts between domains, as illustrated in Figure 2.21D with domains 1 and 3. Also, chronological relationships can be shown in the truncation rule; for example, in Figure 2.21C, the first domain is seen to cross-cut the other two, therefore, from a geological point of view, it should represent a younger domain, while domains 2 and 3 are older. In conclusion, the layout of the truncation rule has implications on the spatial relationships between the geological domains, via the permissible and forbidden contacts between domains, and on their chronological ordering.

- **Truncation thresholds**

Given the geometry of the truncation rule, one has to specify the numerical values (thresholds) that delimit the different domains of the partition in the Gaussian space. For N underlying Gaussian random fields, grouped into a vector random field $Z = \{Z(x): x \in \mathbb{R}^N\}$ with N components, and M geological domains, one usually has with M – 1 thresholds that should be defined according to the proportion of each domain. Let us suppose that g(.) is the joint probability density function of the Gaussian random fields. For calculating the probability of occurrence of the i-th geological domain at a given location x, one needs to solve the following equation:

$$P_i(x) = \int_{D_i} g(z_1, \dots, z_N) dz_1 \dots dz_N \qquad (2.59)$$

where D_i is the region of the partition associated with the i-th geological domain (a subset of $\mathbb{R}^N$ delimited by the thresholds). Even with two Gaussian random fields, it is hard to solve this equation analytically, but most often, it can be solved numerically by trial and error. Trial and error methods and testing successively all the thresholds, lead it to be very slowly. But these calculations may be done faster by grouping the domains, as in Figure 2.22: the desired partition is shown on the top line, while the second line shows the order in which the thresholds are evaluated, starting from the top block.

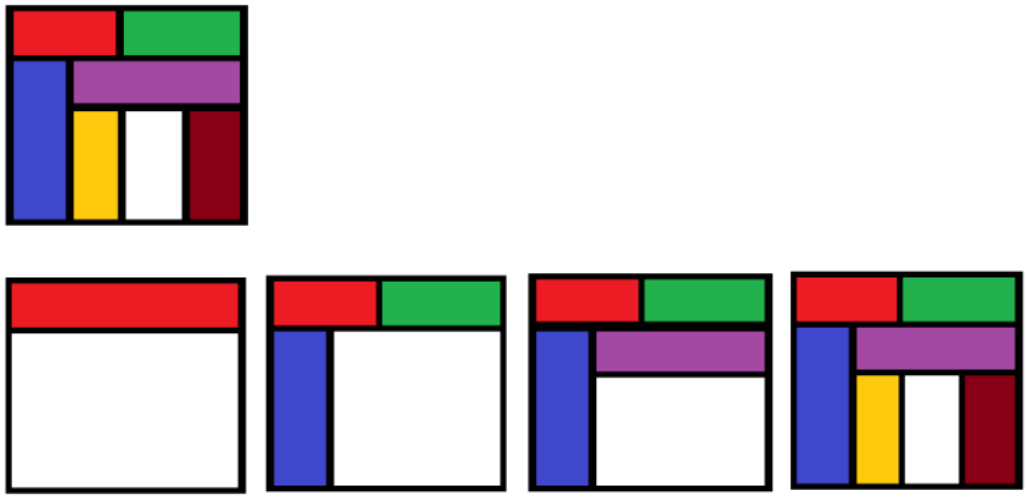


Figure 2.22: Successive groupings to obtain the truncation thresholds (modified from Armstrong et al., 2011)

- **Variograms of the underlying Gaussian random fields**

For any separation vector h, the indicator cross-variogram between two geological domains (with indices i and j) is derived from the corresponding non-centered covariance

$$\gamma_{ij}(h) = C_{ij}(0) - \frac{1}{2}\left(C_{ij}(-h) + C_{ij}(+h)\right) \tag{2.60}$$

with (Armstrong et al., 2011):

$$C_{ij}(h) = prob\left\{Z(x)\ \epsilon\ D_i, Z(x+h)\ \epsilon\ D_j\right\} \tag{2.61}$$

If D_i and D_j are rectangular parallelepipeds of $\mathbb{R}^N$ and the components of the vector random field Z are independent, the second member of equation (2.61) is a function of the direct covariances or variograms of the components of Z and can be calculated by numerical integration (Dowd et al., 2003) or by using expansions into Hermite polynomials (Emery, 2007). This establishes a link between the variograms of the underlying Gaussian random fields and the indicator variograms, which are accessible experimentally from the observed

geological domains at sampling locations. The former can therefore be determined according to the fitting of the latter, quite often, through a trial-and-error procedure (Emery, 2007; Armstrong et al., 2011).

2.2.5.4.2. Conditional simulation

Once the parameters of the plurigaussian model are specified as mentioned in the previous subsection, plurigaussian simulation can be performed in three main steps (Lantuéjoul, 2002; Dowd et al., 2003; Emery, 2007; Armstrong et al., 2011):

1. Simulate the underlying Gaussian random fields at the data locations, conditionally to the categorical data.

2. Simulate the Gaussian random fields at the target locations, conditionally to the values obtained in the previous step (this step can be performed by any multi-Gaussian simulation algorithm).

3. Apply the truncation rule to obtain the simulated domains.

The first step is the difficult one, which can be realized by an iterative algorithm known as the Gibbs sampler (Geman and Geman, 1984; Casella and George, 1992), which belongs to the family of Monte Carlo Markov Chain (MCMC) approaches (Ripley, 1987; Gentle, 2009; Chilès and Delfiner, 2012). The general procedure is the following:

- **Initialization**

For each data location x_α, generate a vector with N components z_α in $D_{i(\alpha)}$, where $i(\alpha)$ is the index of the geological domain observed at x_α.

- **Iteration**

1. Select a data location x_α, regularly or randomly.

2. Calculate the distribution of $Z(x_\alpha)$ conditional to the other data $\{Z(x_\beta): \beta \neq \alpha\}$. This is a Gaussian distribution, with mean equal to the simple kriging of $Z(x_\alpha)$ and covariance matrix equal to the covariance matrix of the simple kriging errors.

3. Simulate a vector z_α according to the previous conditional distribution.

4. If z_α is compatible with the domain prevailing at x_α (i.e., $z_\alpha \in D_{i(\alpha)}$), replace the current value of $Z(x_\alpha)$ by z_α.

5. Go back to step 1 and loop many times.

The Gibbs sampler so presented is an irreducible, aperiodic and reversible Markov chain, with the target Gaussian distribution as its ergodic limit. In other words, if the number of iterations increases infinitely, the distribution of the simulated vectors at the data locations converges to the conditional distribution of the desired Gaussian random fields.

There are some common strategies for selecting the data in Step 1 (Roberts and Sahu, 1997; Galli and Gao, 2001):

- Random sweep: the index of the data is chosen uniformly in {1, …, n}.
- Deterministic updating: the index of the data is increased by one unit at each iteration.
- Reverse updating: the index of the data is increased by one unit at the first iterations, then decreased by one unit at the next iterations, and so on.
- Random permutation: the index of the data follows a random permutation of {1, …, n} for each group of n successive iterations.

Arroyo et al. (2012) showed that the last strategy (random permutation) is getting much faster to convergence. A full neighborhood is required for kriging at Step 2; an alternative version of the sampler has been proposed recently to face the case when the number of data is too large to implement such a neighborhood (Lantuéjoul and Dessassis, 2012; Arroyo et al., 2012).

2.3. State-of-the-art on validation of geological loggings and interpretations

2.3.1. Validating and reclassifying geological logs

Quality control and validation of geological logging have been discussed in both the mining and petroleum industries using different methods and approaches based on geological or geophysical knowledge, data analysis or data mining, or geostatistics.

2.3.1.1. Geology and geophysics-based approaches

Approaches that do not use geostatistics are mainly focused on geological knowledge such as chronostratigraphy and sequencing methods for justifying how facies of different drill holes should be correlated (Figures 2.23 and 2.24). These approaches have been followed with computerizing previously manual methods and also providing better visual space for

decision making. Agterberg (1990) provides comprehensive studies according to these approaches.

Figure 2.23: Zonations based on average stratigraphic events (Agterberg, 1990).

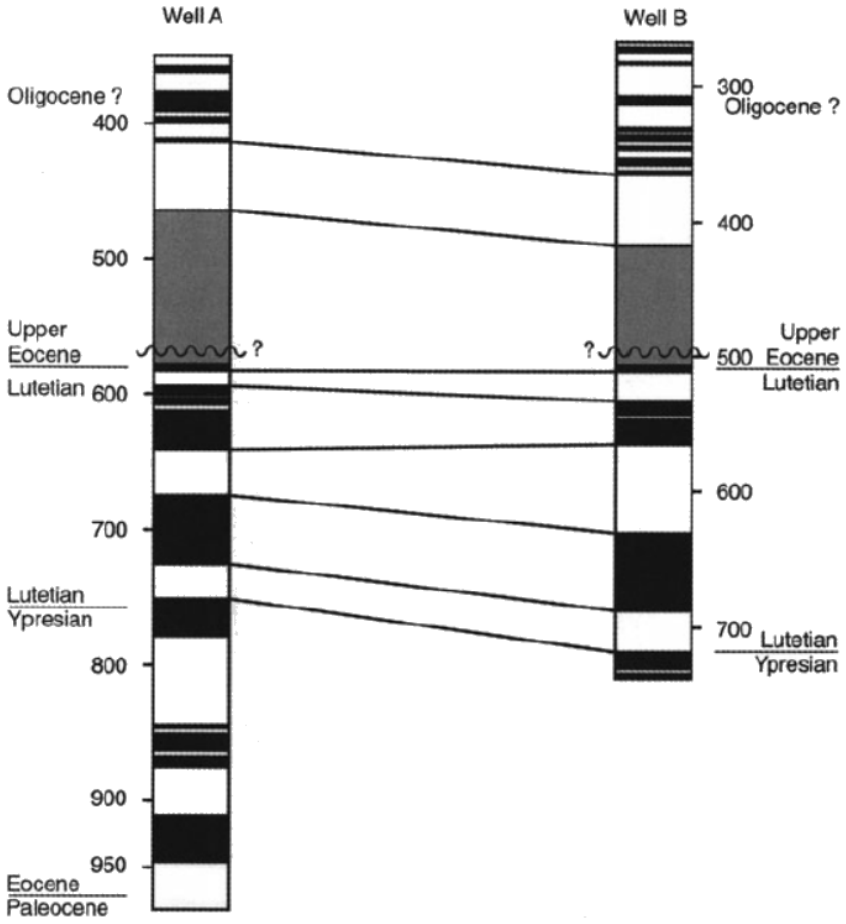


Figure 2.24: Chronostratigraphic correlation in two nearby wells (Luthi, 2001).

Furthermore, geophysical well logging allows mapping of petrophysical signature and can be used for cross-checking geological logs (Figure 2.25) (Hoyle, 1986; Spies, 1996; Hearst et al., 2000; Luthi, 2001; Dunn et al., 2002; Knödel et al., 2007; Soleimani et al., 2016).

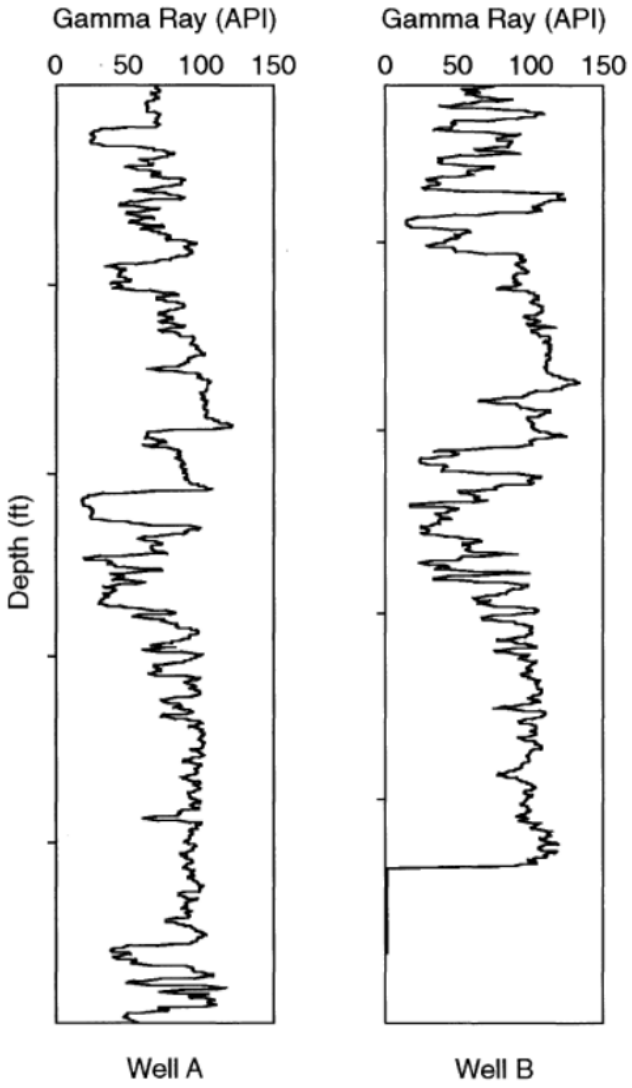


Figure 2.25: Gamma ray logs from two wells can easily be correlated except for some sandstone layers in the central part of the interval (Luthi, 2001).

2.3.1.2. Data mining-based approaches

Some studies propose the application of pattern recognition tools and neural networks to correlate geological logs and detect inconsistent logs (Figure 2.26) (Luthi and Bryant, 1997; Hassibi et al., 2003). Luthi (2001) covers both conventional and new approaches applied for logging correlation.

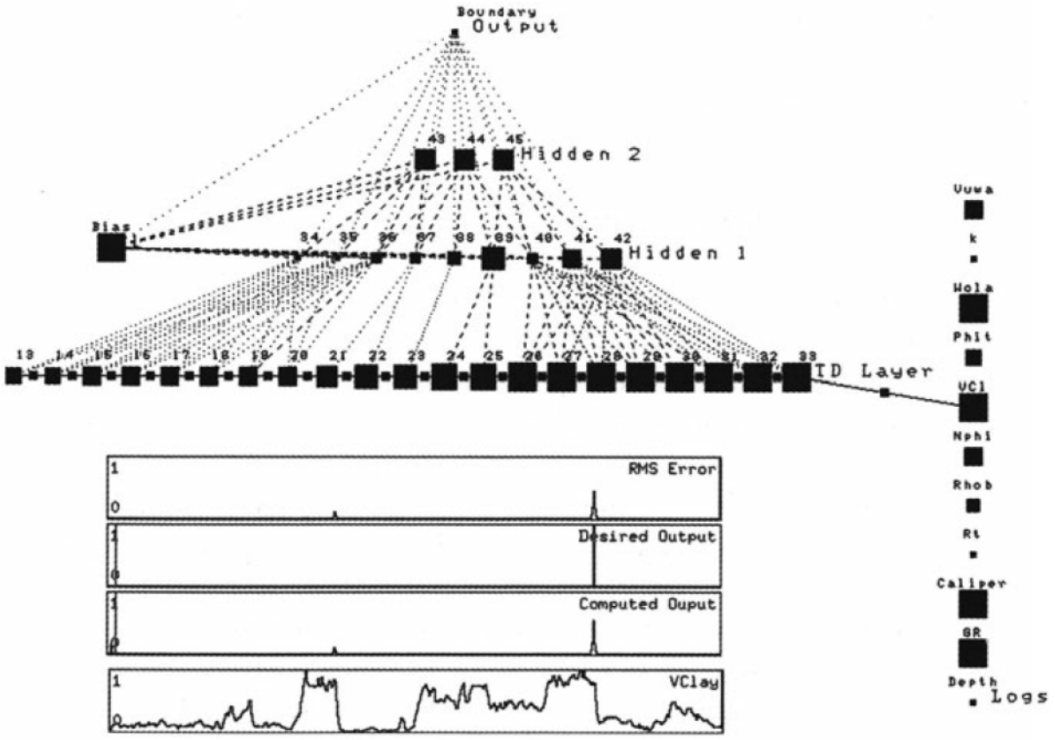


Figure 2.26: Neural network structure used for correlating seven wells in the Maracaibo area, Venezuela. This snapshot is from the training phase, whereby the clay volume (VCl, right) is presented in intervals of ten feet (21 samples) with the known output (1 for the marker to be present, 0 for it to be absent). The learning progress at this stage of the training is shown in the panels at the bottom (Luthi, 2001).

2.3.1.3. Spatial statistics and geostatistics-based approaches

In the field of geostatistics, there are actually few studies for facing this challenge, which are reviewed hereafter.

Bourgine et al. (2008, 2015) present a set of tools and methods designed and tested at BRGM (French Geological Survey) for (1) performing automatic consistency checks before and during modeling, and (2) facilitating the building of geological models (Figure 2.27). The approach is based on the following steps:

(a) Selection of a loose network of boreholes with logging or coring information, enabling a reliable interpretation. This first interpretation is based on the correlation of borehole log data and allows defining a 3D sequence stratigraphic framework with the acceptable difference between logs in terms of maximum variation in formation thickness, maximum variation in the top or bottom elevations, and minimum similarity of logs.

(b) Geostatistical analysis of characteristic geological interfaces, which is based on cross-validation tests. The goal is to identify quickly and semi-automatically potential errors among the data, so the geologist can check these potential errors and correct them.

(c) Consistency tests are also used to verify the appropriateness of the interpretations according to other constraints such as geological maps, maximal formation extension limits, or digital terrain models.

(d) Construction of a 3D geological model from the sum of reference boreholes and recently verified boreholes. Also standard-deviation maps are used for visualizing areas where data from available but not yet validated boreholes could be added to reduce uncertainty.

(e) At last, truncated plurigaussian facies simulations are realized using the validated data.

These approaches are really easier than the ones suggested in this thesis, because they work with only one variable and they should be improved by considering several variables. Another drawback is the risk of over-smoothing using "validated boreholes" and comparing new boreholes with them to be added or ignored. The authors have tested the presented procedure on two areas in France.

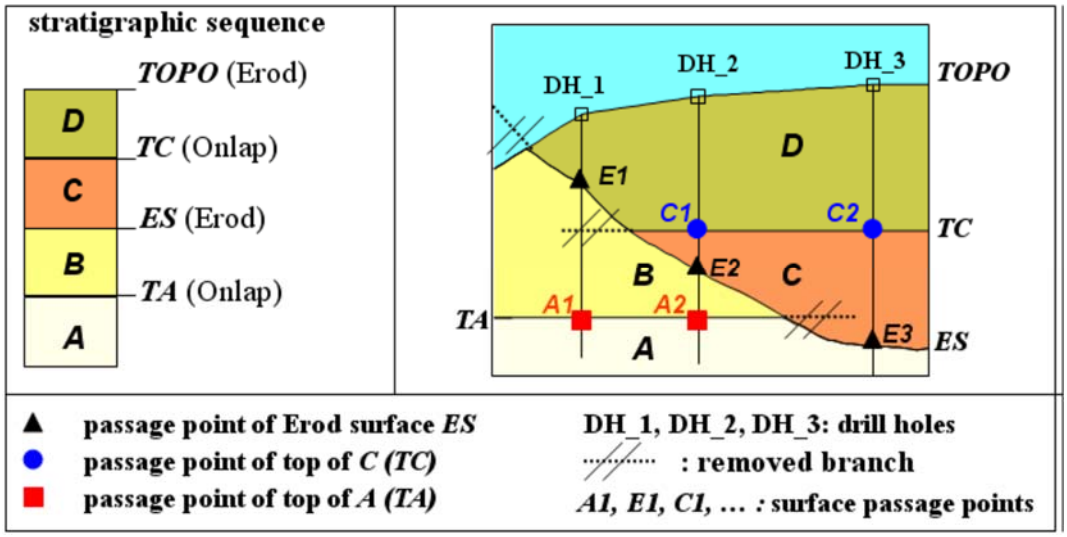


Figure 2.27: Modeling and intersecting stratigraphic surfaces (Bourgine et al., 2008).

More recently, Manchuk and Deutsch (2012) introduced a simple geostatistical measure of coherency based on covariance functions in facies between nearby boreholes. This measure of coherency is applied for quality control of borehole data and in geological zonations, specifically for categorical variables. Coherency measures the agreement between a borehole

and its immediate neighbors based on structural markers and facies interpretations. The authors applied the measure of coherency to a synthetic data set and identified potential problems with facies designations or borehole markers, maximized coherency and used the coherency measure for clustering data into groups having similar facies (Figure 2.28).

This method is proposed for checking coherency of facies between nearby boreholes and tested onto a synthetic data set. However, it does not look to be a practical method for a real case study, as interpreting results and making conclusions by checking relations of each borehole by all surrounding boreholes would be a very time-consuming and tedious work. Also, as mentioned previously, the method is proposed for checking coherency of facies with large-scale spatial variations; checking geological loggings with possibly short-scale variations is more complicated.

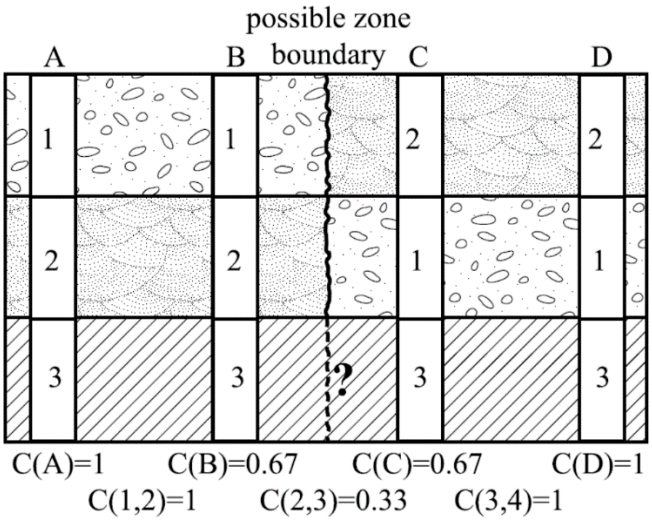


Figure 2.28: Basic concept of geological zonation using coherency. Wells are labeled A through D and facies are labeled 1 to 3. The difference between coherency of each well using the left and right neighbors and coherency between well pairs is shown. Determining if a zone boundary exists in facies 3 based on well pair coherency is problematic (Manchuk and Deutsch, 2012).

Cáceres and Emery (2013) proposed a methodology based on leave-one-out cross-validation, which consists in predicting the quantitative variables at each data location by using only the remaining data. All logged categories are successively assigned to each sample and the predictions are obtained by conditioning to the nearby samples with the same category. Finally, these predictions are compared with the actual values of the quantitative variables at the data location under consideration, determining whether or not the accuracy can be improved by assigning to the sample a category different from the one that has been logged.

Cáceres and Emery applied this methodology to a synthetic deposit by simulating a single quantitative variable (an assayed grade) and three logged classes (rock types), to which different errors are added. The results show that a small amount of logging errors can lead to

a severe distortion of the grade distribution per category, spatial structure and consequently mineral resources evaluation, which can be corrected to a great extent thanks to the proposed cross-validation procedure (Figure 2.29). The proposed approach is, however, applicable to mineral deposits with hard boundaries between logged classes, i.e., when there is no spatial correlation of the quantitative variables across the boundary between two classes, which is a restrictive assumption for disseminated deposits.

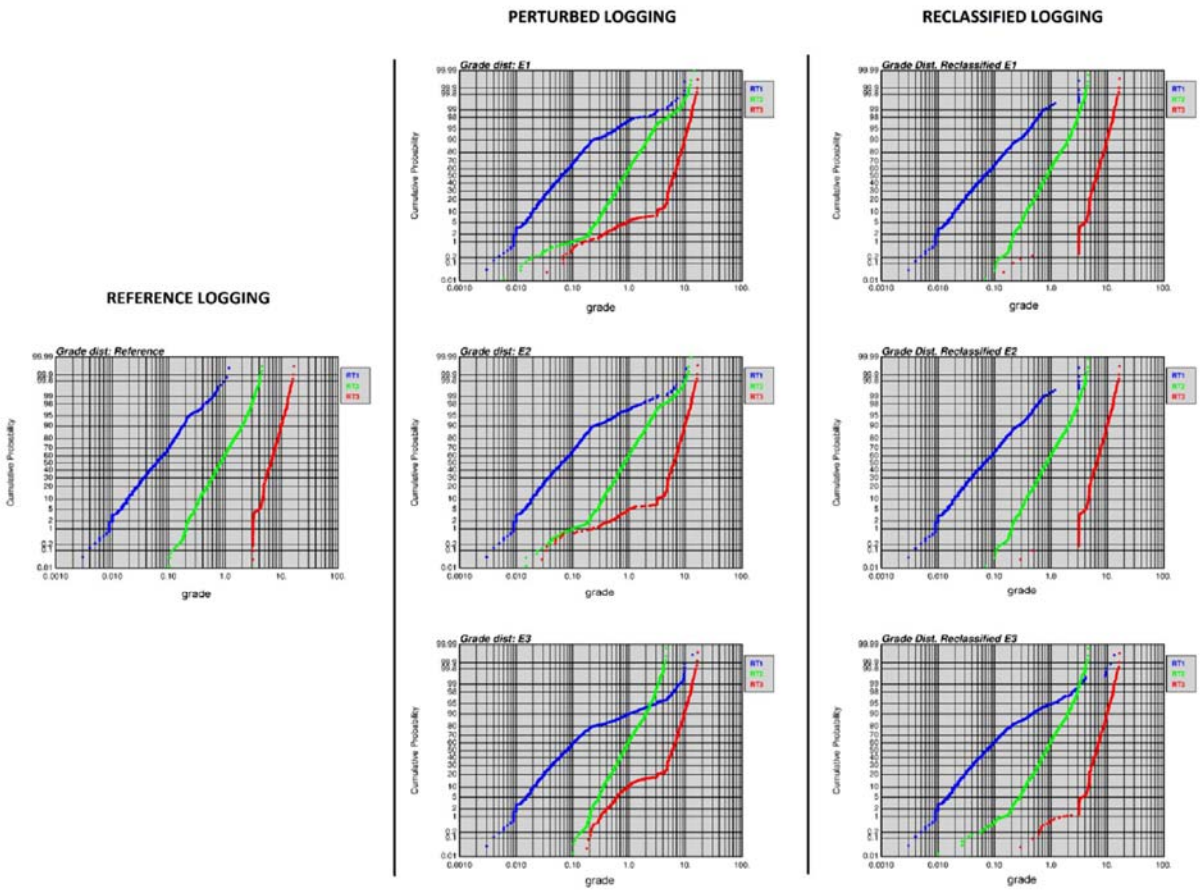


Figure 2.29: Grade distributions by rock type for reference, perturbed and reclassified loggings, for three kinds of logging errors: small random error (top), large random error (center) and systematic error (bottom) (Cáceres and Emery, 2013). The distributions observed with the perturbed loggings (middle column) are significantly different from the original distribution (left). The corrected distributions after cross-validation (right column) are much closer to the original distribution.

2.3.2. Validating and reclassifying geological interpretations

A geological interpretation consists of a three-dimensional representation of the extents of rock types or ore types constructed on the basis of the geological knowledge of the deposit, geological field observations, geophysical surveys and drill hole logs and assays. However,

a single representation of the ore deposit is often constructed, which lacks a quantification of the uncertainty in the actual rock types or ore type locations and extents.

Most published researches on geological interpretations focus on using all available data to generate a more accurate geological interpretation, e.g., by using structural and geophysical data together in addition to borehole data or by using data inversion methods to generate an interpreted model (Figure 2.30) (Guillen et al., 2008; Lelièvre, 2009). Studies on validating geological interpretations often concentrate on statistical and graphical analyses by comparing the models with the available data to detect inconsistencies (Rossi and Deutsch, 2014).

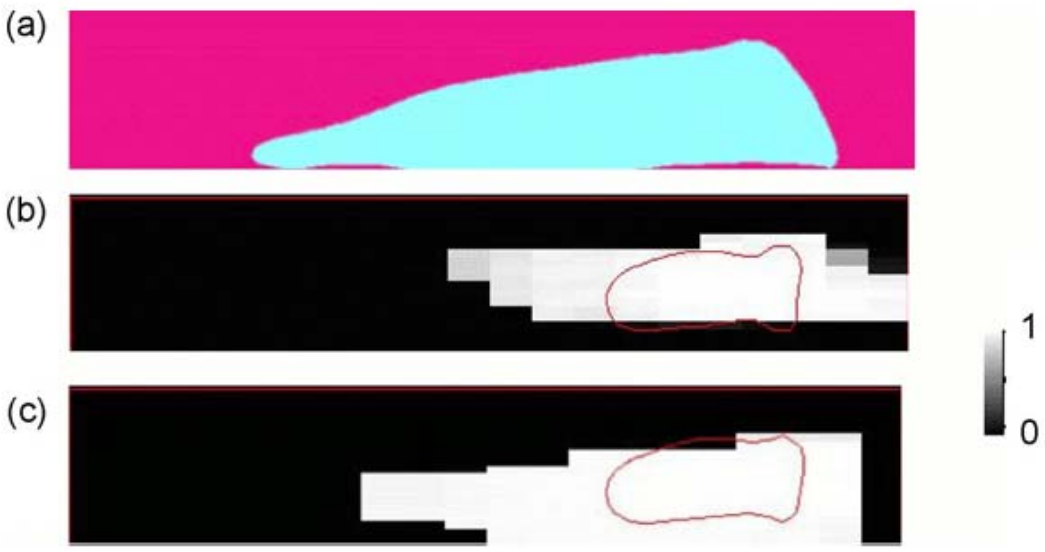


Figure 2.30: Comparison between a conceptual geological model and inversion results. (a) Conceptual model. (b) Probability to obtain the unit after inversion of the gravity field only. (c) Probability to obtain the unit after simultaneous inversion of the gravity field and the five tensors components (Guillen et al., 2008)

In the same line, Maleki et al. (2017) propose comparing the indicator direct and cross-variograms calculated on available borehole data with those calculated on the interpreted geological model. Such variograms convey information about the spatial and geometrical properties of geological domains and about their contact relationships, in particular in what refers to the smoothness of the domain boundaries, the contact area between two domains, and their propensity to be in contact with, or separated from, each other at short distances. The proposed tools, however, do not allow a local analysis to detect which part of a geological interpretation is likely to be mistaken. Rather, it provides a global overview of how much consistent is the interpretation with the logged data (Figures 2.31 and 2.32).

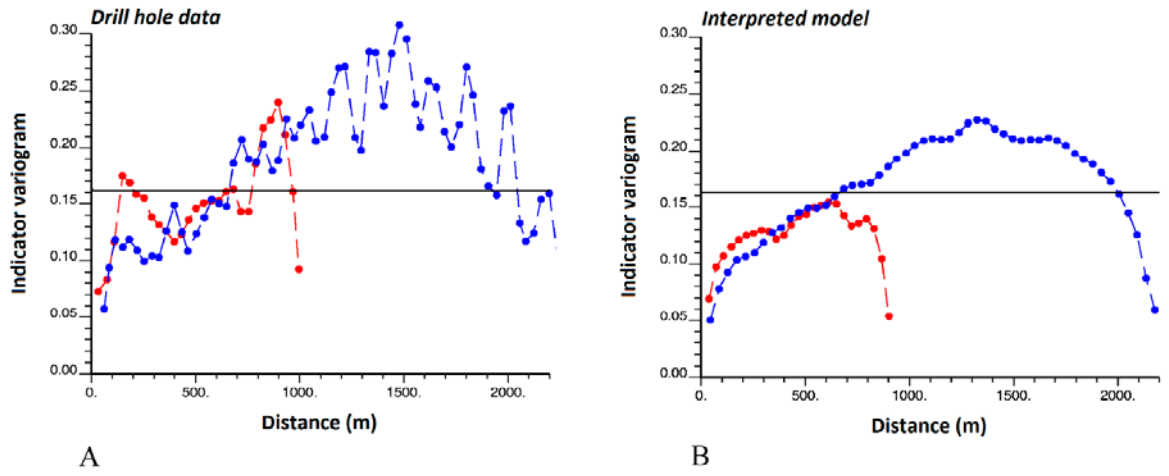


Figure 2.31: Indicator direct variogram for andesite rock type along two main anisotropy directions, calculated from (A) drill hole data, (B) interpreted model. Calculations consider the same angle tolerance of 20° around the target directions and same lag values. Variograms in (A) and (B) exhibit the same shape, slope near the origin and behavior at short, medium and large scales, indicating consistency between the interpreted geological model and the drill hole data for this particular rock type (Maleki et al., 2017).

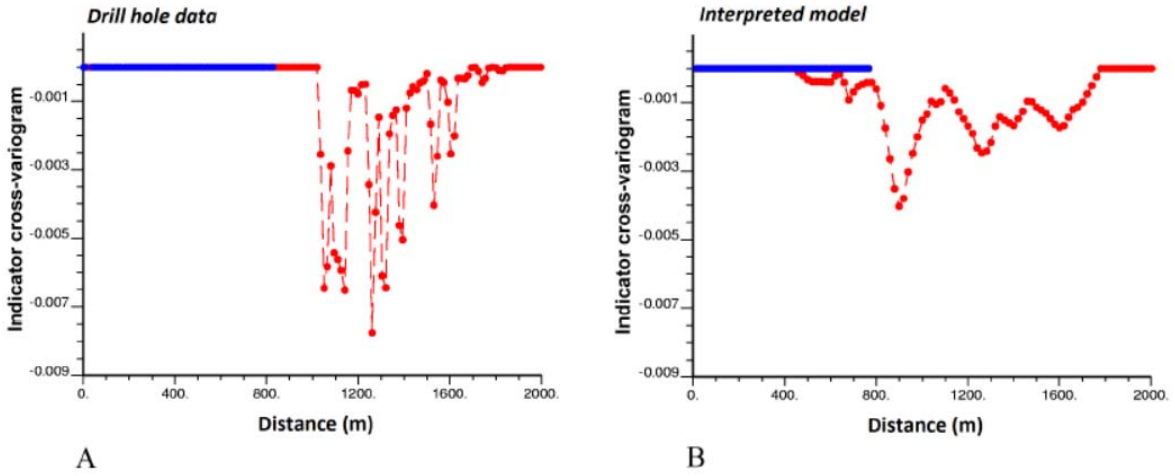


Figure 2.32: Indicator cross-variogram between granodioritic porphyry and hydrothermal breccia along two main anisotropy directions, calculated from (A) drill hole data, (B) interpreted model. Calculations consider the same angle tolerance of 20° around the target directions and same lag values. Indicator cross-variogram in (A) is zero until a distance of 1000 m, indicating the minimal separation distance between granodioritic porphyry and hydrothermal breccia. However, the indicator cross-variogram in (B) is zero until a distance of 500 m only, suggesting that the interpreted geological model underestimates the true separation distance between the two rock types (Maleki et al., 2017).

Chapter 3: Modeling the relationship between geological classes and quantitative variables

Suppose that one has a categorical variable, such as a rock type or a dominant alteration type, known (with some imprecision) from qualitative geological logs, and a set of quantitative continuous variables (metal grades, rock granulometry, rock density, metallurgical recoveries, etc.) known with precision from assays or metallurgical tests at borehole samples. Also suppose that one has a model of the categorical regionalized variable over space, based on an interpretation from resources geologists.

From this information, one is interested in solving two following problems:

(a) Validate the geological logs and identify the data for which the qualitative logged category is not in agreement with the quantitative continuous variables;

(b) Validate the geological interpretations and identify the data for which the interpreted category is not in agreement with the quantitative continuous variables.

Several models can be designed to address the aforementioned problems. The following three will be of interest in the thesis and will be explained in this chapter; models 2 (an extension of model 1) will be used in the next chapter for validating geological logs, while model 3 will be used in the subsequent chapter for validating geological interpretations.

3.1. Model 1: Hard boundary model

The model is hierarchical, with the quantitative continuous variables "subordinated" to the categorical variable. The space is partitioned into regions corresponding to the categories associated with the geological logs. The quantitative continuous covariates are defined in each category, with a joint correlation structure that can be represented by a linear model of coregionalization (Chilès and Delfiner, 2012). These continuous variables are independent from one category to another. To make the model operational, one can work with multi-Gaussian variables, i.e., in each category, the continuous variables can be transformed into Gaussian random fields, whose direct and cross-covariances are modeled through a linear coregionalization model (Figure 3.1).

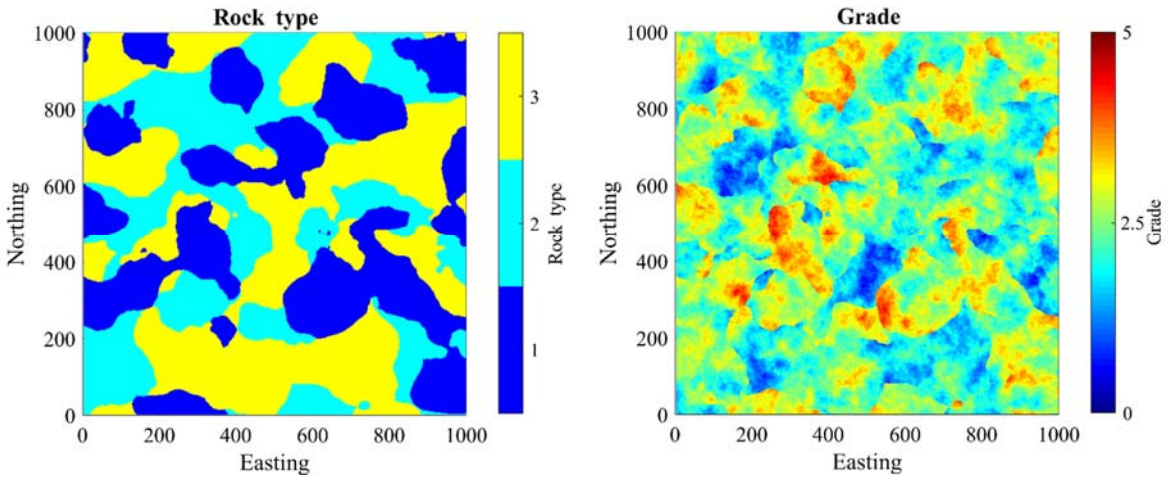


Figure 3.1: Left: rock type. Right: a quantitative continuous variable. The values of the quantitative variable are independent from one rock type to another, resulting in clear-cut discontinuities when crossing the rock type boundaries (hard boundary model).

In this model, the validation of geological logs (problem 1) can be tackled by leave-one-out cross-validation (Figure 3.2). Specifically, for each drill core sample, do the following:

a) Select a drill core sample and assume that it belongs to category k.

b) Transform the N quantitative continuous variables measured at this drill core sample into normal scores.

c) From the surrounding drill core samples of category k, perform cokriging of the continuous variables for the drill core sample under consideration. Obtain a vector of predictions and a variance-covariance matrix of cokriging errors.

d) Extract the vector of cokriging predictions and the variance-covariance matrix associated with category k (i.e., remove the prediction of the Gaussian random fields associated with geological domains that do not contain category k).

e) Based on the vector of cokriging predictions and variance-covariance matrix of cokriging errors, determine the Malahanobis distance corresponding to the vector of true values.

f) Repeat steps (a)-(e) for all the categories.

g) From the Malahanobis distances so obtained, construct a measure of consistency for each category for the drill core sample under consideration.

h) Determine whether or not the actual logged category is highly probable; identify other categories that may be more probable (possibly mislogged drill core sample).

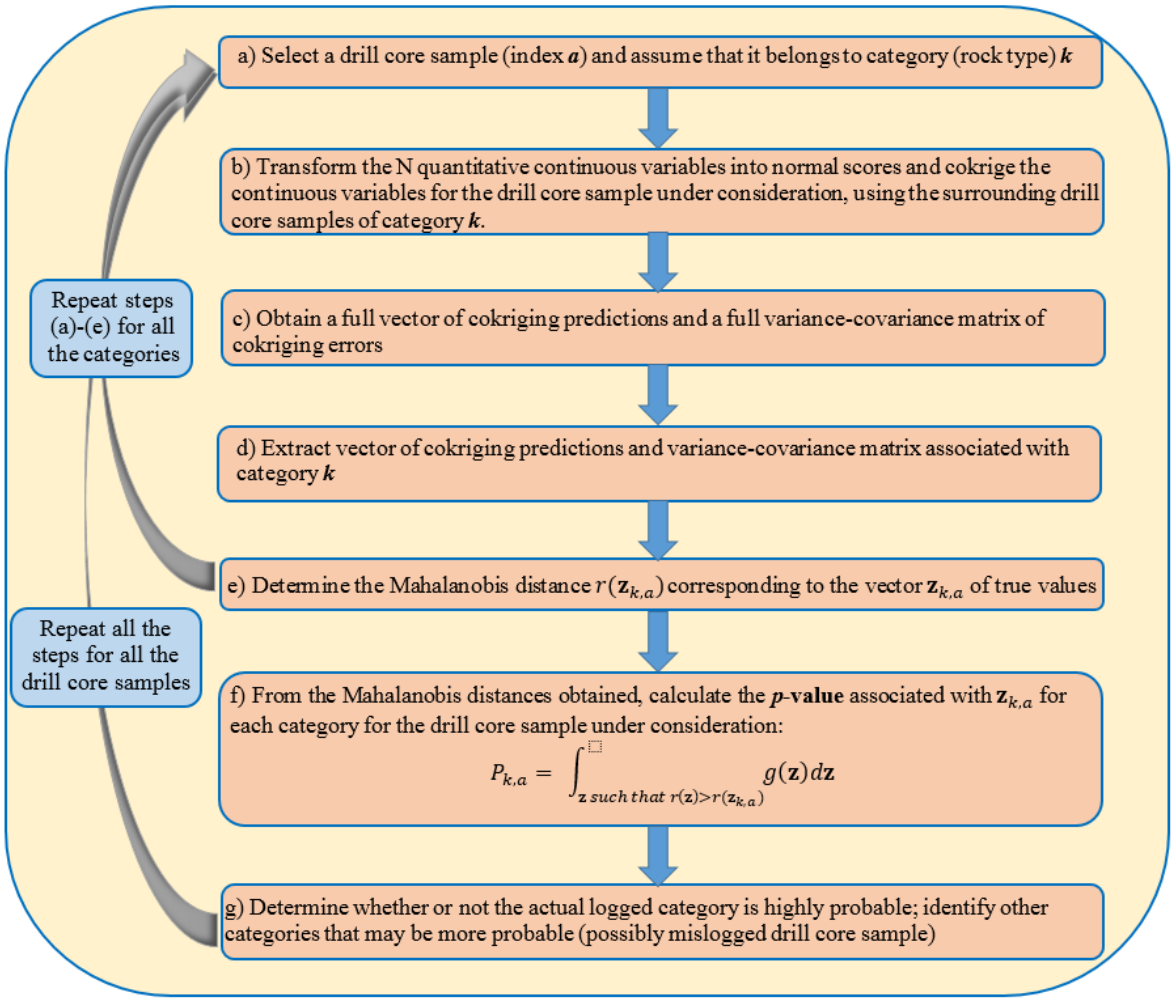


Figure 3.2: Workflow of the leave-one-out cross-validation steps based on model 1.

3.2. Model 2: Transitional (semi-hard) boundary model

Consider the previous hierarchical model, but now assume that the quantitative continuous variables are not independent from one category to another (Figure 3.3). This means that one has a complete coregionalization model that specifies the direct and cross-covariances (or variograms) of the continuous variables of the same categories, as well as of different categories.

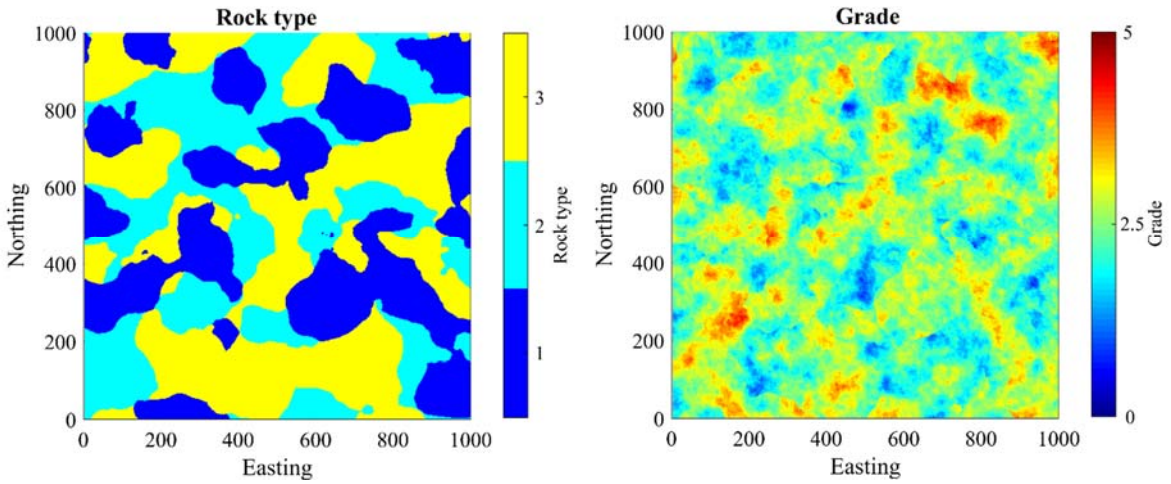


Figure 3.3: Left: rock type. Right: a quantitative continuous variable. The values of the quantitative variable are dependent between one rock type and another, resulting in cross-correlations across the rock type boundaries (transitional boundary model)

The solution of the previous problem is the same, except that cokriging at Step (c) should consider the surrounding data of the category under consideration as well as of other categories.

3.3. Model 3: Soft boundary model

Here, contrarily to the previous two models, the categorical variable is now subordinated to the quantitative continuous variables. The latter are represented by a global coregionalization model (without partitioning the space into regions). The category is then defined depending on the values of these continuous variables (Figure 3.4).

In such a model, the validation of geological interpretations (problem 2) can be done by jointly simulating the continuous variables over the entire deposit, then applying a classification technique from statistics or from data mining, such as discriminant analysis, neuronal network, support vector machine or classification trees, so as to predict the categories on the basis of the quantitative continuous variables for the different realizations. The determination of the prior (ignoring borehole data) and posterior (conditioned to borehole data) probabilities of categories provides a means of identifying the blocks or areas of the deposit for which the interpreted category disagrees with the simulated continuous variables, which leads to identifying potentially geological misinterpreted blocks (Figure 3.5).

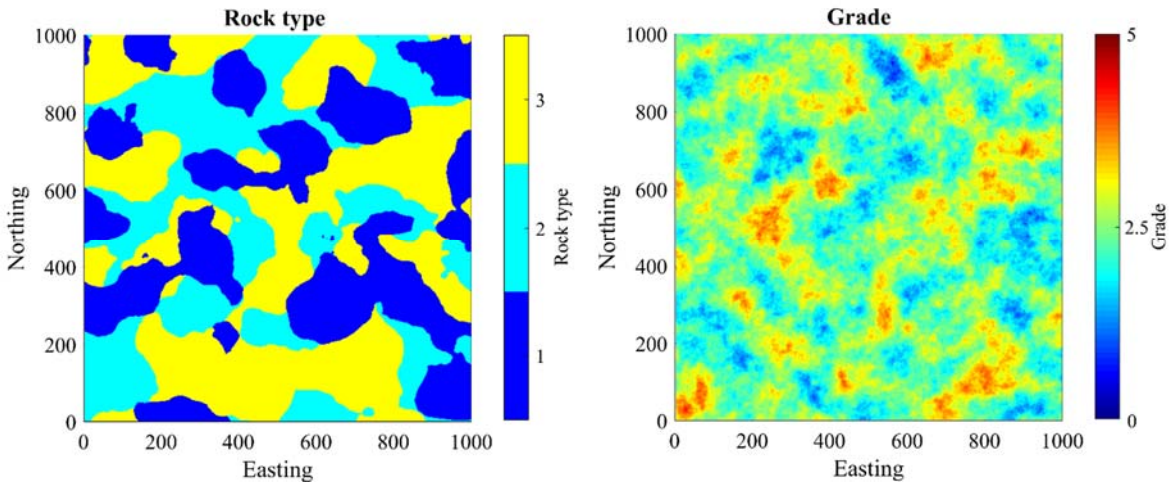


Figure 3.4: The quantitative variable (right) is defined over the entire region, prior to the partitioning into rock type domains (left), resulting in no discontinuity across the rock type boundaries (soft boundary model).

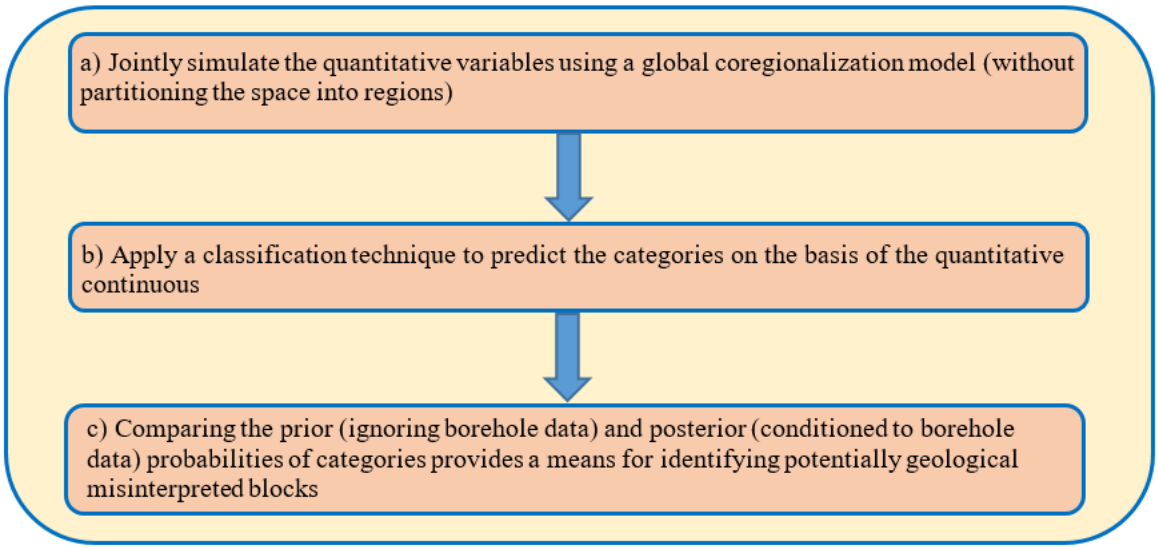


Figure 3.5: Workflow of the validation of geological interpretations based on model 3.

Chapter 4: Designing a geostatistical-based method for validating geological logs using quantitative covariates

Errors in geological core logs may lead to inconsistencies between the logged categories and other quantitative variables obtained from geochemical analyses or metallurgical tests. Based on the spatial dependence relationships between both types of variables (qualitative logs and quantitative covariates) and on the transitional boundary model (model 2) presented in the previous chapter, this chapter addresses the problem of developing an objective methodology for identifying potentially mislogged samples. The contents have been published in the journal Ore Geology Reviews:

Adeli, A., Emery, X., 2017. A geostatistical approach to measure the consistency between geological logs and quantitative covariates. Ore Geology Reviews 82, 160-169.

A geostatistical approach to measure the consistency between geological logs and quantitative covariates

Abstract

Core logging is the geological study, recording and classification of petrophysical attributes of drill hole samples, such as lithology, alteration or mineralogical assemblage. The geological logging is qualitative and subject to errors because of its visual nature and other factors inherent to logging, such as low drill hole recoveries, difficulties in estimating the volumetric contents of minerals, or different logging criteria among geologists. To date, different tools for quality control and validation of geological logging have been elaborated, based on geological knowledge, statistics, geostatistics, image analysis, neural network and data mining.

This paper presents an alternative approach based on geostatistical modeling for identifying and reclassifying potentially mislogged samples when quantitative covariates from geochemical analyses or metallurgical tests are available. The principle of this approach is to: (i) define geological domains for each quantitative variable by an adequate grouping of the log classes; (ii) transform the quantitative variables into normal scores, accounting for the previously defined domains, (iii) model the spatial correlation structure of the normal scores, (iv) perform leave-one-out cross validation and obtain predictions of the normal variables and the associated variance-covariance matrices of prediction errors; (v) calculate a measure of consistency for each sample and each possible logged class under a multivariate normal assumption; and (vi) compare these measures of consistency with the actual logged classes to detect suspicious logs. The methodology is demonstrated in a case study from an iron ore deposit, with data of rock type logged by geologists and seven quantitative variables (grades of elements of interest, loss on ignition and granulometry).

Key words: core logging; geological domaining; inconsistent data; misclassification.

4.1. Introduction

Drilling is the most expensive procedure in mineral exploration. Information from drill holes can be extracted by different methods, such as assaying, down-the-hole geophysical logging or geological core logging (Knödel et al., 2007; Marjoribanks, 2010). The latter is the geological study, visual recording and classification of petrophysical attributes of drill cores (essentially extracted by diamond drilling), such as lithology, alteration or mineralogical assemblage. Relevant information gathered from geological core logging is the basis for

constructing geological and geo-metallurgical models for mineral resources evaluation and classification, ore reserves definition and mine planning (e.g., Soltani and Hezarkhani, 2011). It is also often used for partitioning heterogeneous deposits into geological or geo-metallurgical domains in which the regionalized properties of interest can be interpreted as stationary fields (Sinclair and Blackwell, 2002; Moon et al., 2006; Yunsel and Ersoy, 2011; Haldar, 2013; Rossi and Deutsch, 2014).

However, due to the visual nature of logging, the classification of petrophysical attributes is qualitative and subject to errors, which may be explained by several factors (Manchuk and Deutsch, 2012; Cáceres and Emery, 2013): presence of complex rock textures caused by overprinting processes; inherent difficulty to estimate mineral percentages and thresholds; lack of geochemical analyses during logging; lack of experience of mining geologists; non-unique logging criteria among geologists; and low core recovery because mineralized and altered rock zones are frequently more fragile and are the first parts that are lost during coring. Inaccurate logs generate data that are inconsistent with geochemical analyses and metallurgical tests, such as low iron grades in supposedly supergene hematite zones or low acid consumption in supposedly calcareous rocks. Due to limited time and resources for relogging, inconsistent logs are generally seen as outliers or ignored in the geological or geo-metallurgical modeling stage (Theys, 1999, Cáceres and Emery, 2013).

Quality control and validation of geological logging have been discussed in the fields of geosciences and resources engineering, where manual to automated procedures have been elaborated, based on geological knowledge, statistics, geostatistics, image analysis, neural network and data mining tools (Hoyle, 1986; Agterberg, 1990; Luthi and Bryant, 1997; Taylor, 2000; Luthi, 2001; Hassibi et al., 2003; Bourgine et al., 2008, 2015; Ewusi and Kuma, 2011). Geological knowledge-based approaches are mainly focused on chronostratigraphy and sequencing methods for correlating the logs of different drill holes (Agterberg, 1990). Furthermore, geophysical well logging allows mapping of petrophysical signature and can be used for cross-checking geological logs (Spies, 1996; Hearst et al., 2000; Luthi, 2001; Dunn et al., 2002; Knödel et al., 2007; Soleimani et al., 2016). In recent years, Manchuk and Deutsch (2012) introduced a geostatistical measure of coherency for geological logs, based on a spatial correlation model, which can be applied for quality control of drill hole data. Another proposal by Cáceres and Emery (2013) relies on leave-one-out cross-validation in order to identify log data that disagree with quantitative measurements. This proposal was applied to a synthetic deposit characterized by a single quantitative variable (an assayed grade) and three logged classes (rock types) to which different logging errors were added. The results showed good ability to identify mislogged rock types based on grade information. However, the proposed approach is applicable to mineral deposits with hard boundaries between logged classes, i.e., when there is no spatial correlation among quantitative variables across the boundary between two classes, which is a too restrictive requirement for disseminated deposits.

This work aims to present an improvement of the approach by Cáceres and Emery (2013) for validating geological logs and identifying data for which the qualitative logged class is not in agreement with quantitative variables available from assays or metallurgical tests. The proposed methodology is also based on leave-one-out cross-validation, but will be set in a fully multivariate framework and will be applicable to mineral deposits where quantitative variables are correlated across boundaries between logged classes (soft boundaries). It relies on the following four working assumptions: (1) geochemical analyses, metallurgical tests and geological features such as lithology, alteration or mineralogical assemblage should be consistent in spatial and statistical terms, insofar as they are related responses to the same geological processes; (2) geochemical analyses and metallurgical tests are more accurate than geological logging; (3) a small proportion of the drill cores is likely to be mislogged; (4) the quantitative variables are regionalized and can be interpreted as realizations of spatial random fields that, within one or more domains that partition the deposit, are transforms of stationary Gaussian random fields, i.e., random fields whose finite-dimensional distributions are multivariate normal and are invariant under a translation in space. The latter assumption is the basis of most of the current geostatistical approaches used for simulating quantitative variables and for quantifying geological uncertainty in mineral deposits (Verly, 1983; Chilès and Delfiner, 2012; Rossi and Deutsch, 2014; Deutsch et al., 2016).

The paper is outlined as follows. Section 2 describes our proposal from a methodological point of view and leads to the definition of a statistical measure (*p*-value) for each drill hole sample that indicates how closely the classes of the logged petrophysical attribute agree with the quantitative variables resulting from geochemical analyses and metallurgical tests. Section 3 presents an application to drill hole samples of an iron ore deposit, for which information of the logged rock type and seven quantitative covariates is available. Conclusions follow in Section 4.

4.2. Methodology

4.2.1. Problem statement

Let us consider a set of drill hole samples with spatial coordinates $\{x_1,\ldots, x_A\}$, for which geological logs are available. These logs provide a class number (between 1 and K) for a petrophysical attribute (e.g., rock type). In addition, there are N quantitative variables (for example, grade assays) for the same samples as that of the geological logs. The information from the quantitative variables is supposed to be accurate, due to a proper quality assurance and quality control program, while that of the logs is subject to errors. The proposed methodology is explained in the following subsections and summarized in the schematic diagram presented in Figure 4.1.

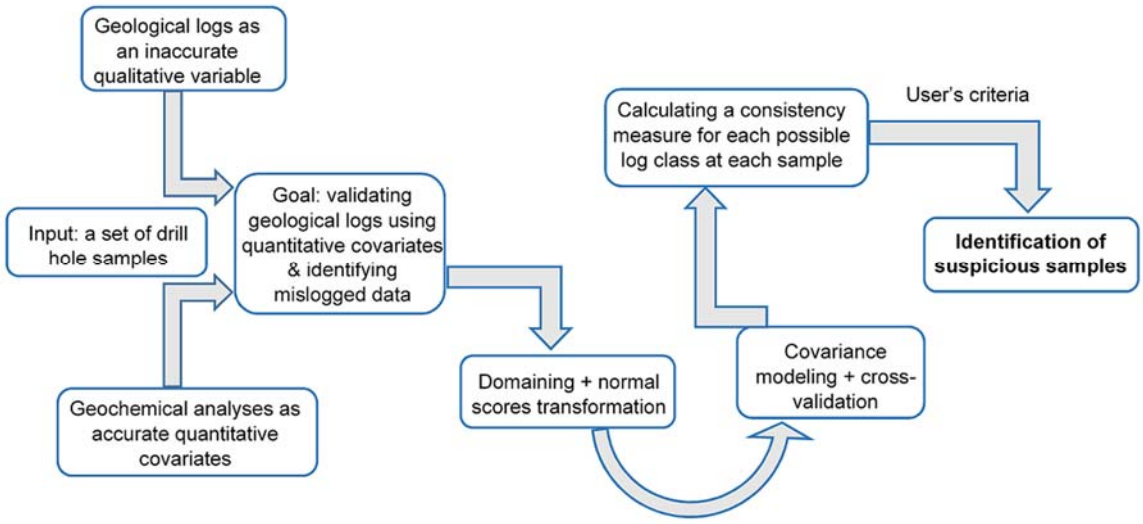


Figure 4.1: Schematic diagram of proposed methodology.

4.2.2. Geological domaining

For mineral resources evaluation in heterogeneous deposits, each quantitative variable is generally associated with geological domains that consist of a single log class or a group of log classes such that (i) the quantitative variable can be modeled as a stationary field (i.e., a field whose distribution is invariant under a spatial translation) within each domain and (ii) there is a change of distribution between one domain and another (e.g., one observes some discontinuity or some change in the spatial continuity when crossing the boundary between two domains). A contact analysis is helpful at this stage to determine the best grouping and therefore identify the geological domains (Glacken and Snowden, 2001; Rossi and Deutsch, 2014; Maleki and Emery, 2015). Also note that the domains can differ from one quantitative variable to another, depending on the geological characteristics that control the behavior of each variable; in other words, the geological controls may not be the same for all the quantitative variables.

4.2.3. Normal score transformation

The data of each quantitative variable within each of its geological domains are declustered in order to obtain a distribution corrected for possible biases caused by irregular sampling patterns, then normal score transformed, i.e., transformed into data that follow a standard Gaussian distribution (Deutsch and Journel, 1998; Chilès and Delfiner, 2012). The one-to-one relationships between the original variables and their normal scores can be stored in transformation tables.

As the geological domains associated with a given quantitative variable do not overlap, the normal score data associated with the different domains are totally heterotopic, i.e., they are not defined at the same locations. In contrast, the normal scores data associated with different quantitative variables may be isotopic, partially heterotopic or totally heterotopic, depending on whether the geological domains are the same, partially overlap or are disjoint and on the sampling design of the quantitative variables (isotopic or heterotopic).

4.2.4. Covariance analysis

Since sample cross-variograms cannot be calculated for heterotopic datasets (Wackernagel, 2003), the spatial correlation structure of the normal score data is inferred by calculating their sample direct and cross covariances. Direct covariances measure the spatial continuity of each quantitative variable within each of its geological domains, while cross-covariances measure the spatial cross-correlation that exists between two different quantitative variables or the spatial cross-correlation of a quantitative variable between two different geological domains. This way, our methodology is able to account for soft boundaries, i.e., when measurements of a quantitative variable exhibit some correlation across domain boundaries (a frequent situation for disseminated deposits).

On the basis of the calculated sample covariances, a linear coregionalization model can be fitted by using combinations of basic nested covariance models (Wackernagel, 2003; Chilès and Delfiner, 2012), so as to define a set of theoretical direct and cross covariances for the Gaussian random fields that represent the quantitative variables within their different geological domains.

4.2.5. Cross-validation

Leave-one-out cross-validation (Chilès and Delfiner, 2012) is then performed in order to get a consistency measure between the data associated with the quantitative variables and the logged classes. Specifically, one drill hole sample is removed at a time from the dataset and all the Gaussian random fields at this sample are predicted by cokriging, by using all the remaining data located in the same geological domain and in the other domains. The results of cokriging for the sample under consideration are a vector of the prediction of the Gaussian random fields and a variance-covariance matrix of the prediction errors. This process is performed for each sample successively.

4.2.6. Calculation of p-values

For each sample $a \in \{1 \dots A\}$ and each class index $k \in \{1 \dots K\}$:

(1) Assume that, if no logging error occurred, class k would prevail at the sample under consideration.

(2) Transform the N quantitative variables measured at this sample into normal scores, using the suitable normal score transformation tables obtained in subsection 4.2.3. Obtain a Gaussian vector $\mathbf{z}_{k,a}$ with N components.

(3) From the vector of cokriging prediction obtained in subsection 4.2.5, extract the sub-vector $\mathbf{z}_{k,a}^*$ corresponding to the Gaussian random fields that are defined for the class under consideration (i.e., remove the prediction of the Gaussian random fields associated with geological domains that do not contain class k).

(4) Similarly, from the variance-covariance matrix of cokriging errors, extract the sub-matrix $\mathbf{\Sigma}_{k,a}^*$ corresponding to the Gaussian random fields defined for the class under consideration.

(5) Based on the assumption made at stage (1), the conditional distribution of the Gaussian random fields at the sample location is multivariate normal with mean equal to the cokriging prediction and with variance-covariance matrix equal to that of the cokriging errors (Chilès and Delfiner, 2012). Accordingly, it has the following probability density function:

$$g(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{\Sigma}_{k,a}^*)}} \exp\{-\frac{1}{2}(\mathbf{z}-\mathbf{z}_{k,a}^*)^T (\mathbf{\Sigma}_{k,a}^*)^{-1}(\mathbf{z}-\mathbf{z}_{k,a}^*)\} \tag{4.1}$$

where N is the common size of vectors $\mathbf{z}_{k,a}$ and $\mathbf{z}_{k,a}^*$ (number of original quantitative variables), $\mathbf{z}$ is a generic point of R^N, while $\mathbf{z}_{k,a}^*$ and $\mathbf{\Sigma}_{k,a}^*$ have been obtained in the previous steps (3) and (4). Let us define the Mahalanobis distance as

$$r(\mathbf{z}) = \sqrt{(\mathbf{z}-\mathbf{z}_{k,a}^*)^T (\mathbf{\Sigma}_{k,a}^*)^{-1}(\mathbf{z}-\mathbf{z}_{k,a}^*)} \tag{4.2}$$

It provides a measure of the distance between the generic point $\mathbf{z}$ of R^N and the prediction $\mathbf{z}_{k,a}^*$ (expected value of the conditional distribution in equation (4.1)). It is dimensionless, scale-invariant and takes into account the correlations between the N Gaussian random fields defined for the class k under consideration. If $N = 1$, the distribution in equation (4.1) reduces to a univariate normal distribution and the Mahalanobis distance reduces to the standard score. For $N > 1$, the Mahalanobis distance is just a multidimensional generalization of the idea of measuring how many standard deviations away $\mathbf{z}$ is from $\mathbf{z}_{k,a}^*$.

The *p*-value $p_{k,a}$ associated with $\mathbf{z}_{k,a}$ (supposedly true Gaussian vector for sample *a*, which has been obtained at step (2)) is defined as the integral of the probability density in equation (4.1) over all the points of R^N with a Mahalanobis distance greater than $r(\mathbf{z}_{k,a})$. This *p*-value gives a measure of the remoteness between the expected value $\mathbf{z}_{k,a}^*$ and the supposedly true value $\mathbf{z}_{k,a}$, i.e.:

$$\begin{aligned} p_{k,a} &= \int_{\mathbf{z} \text{ such that } r(\mathbf{z})>r(\mathbf{z}_{k,a})} g(\mathbf{z})\,d\mathbf{z} \\ &= 1-\frac{1}{(2\pi)^{N/2}}\int_0^{r(\mathbf{z}_{k,a})} \exp\{-\frac{1}{2}t^2\}\frac{2\pi^{N/2}}{\Gamma(\frac{N}{2})}t^{N-1}\,dt \end{aligned} \tag{4.3}$$

where $2\pi^{N/2}/\Gamma(\frac{N}{2})$ is the surface area of the unit-radius sphere of R^N (Chilès and Delfiner, 2012). To calculate $p_{k,a}$, let us define

$$I_N(r) = \int_0^r \exp\{-\frac{1}{2}t^2\}t^{N-1}\,dt \tag{4.4}$$

This integral can be computed by using integration by parts. One finds:

$$\begin{cases} I_1(r) = \sqrt{2\pi}(G(r)-\frac{1}{2}) \\ I_2(r) = 1-\exp\{-\frac{r^2}{2}\} \\ I_N(r) = -r^{N-2}\exp\{-\frac{r^2}{2}\} + (N-2)I_{N-2}(r) \quad \text{for } N \geq 3 \end{cases} \tag{4.5}$$

where *G* stands for the standard Gaussian cumulative distribution function.

At the end of this process, one obtains, for each drill hole sample (*a*) and each class (*k*), a *p*-value $p_{k,a}$ that measures how "extreme" is the vector of *N* quantitative variables measured at this sample for the class under consideration. The analysis of the set of *p*-values will allow identification of the data for which the logged class is likely to be mistaken, as illustrated in the case study presented in the next section.

4.3. Case study

4.3.1. Presentation of the data

The proposed methodology is now applied to a data set from an iron ore deposit, the name and location of which are not disclosed for confidentiality reasons. The deposit is hosted by banded iron formations and is explored by diamond drill holes dipping from 60° to 90°. A total of 4,096 samples are taken from cores and sent for analysis, yielding data for seven quantitative variables: grades of iron (Fe), silica (SiO_2), phosphorus (P), alumina (Al_2O_3) and manganese (Mn), loss on ignition (LOI) representing the mass concentration of volatile materials, and granulometric fraction of fragments with size above 6.3 mm (G). The dominant rock type is also available for each drill hole sample; it is deduced by geological logging and is classified into 10 classes:

- **Code 1**: Friable hematite (FH)
- **Code 2**: Compact hematite (CH)
- **Code 3**: Alumina-rich hematite (ALH)
- **Code 4**: Alumina-rich itabirite (ALI)
- **Code 5**: Manganese-rich itabirite (MNI)
- **Code 6**: Compact itabirite (CI)
- **Code 7**: Friable iron-poor itabirite (FPI)
- **Code 8**: Friable iron-rich itabirite (FRI)
- **Code 9**: Amphibolitic itabirite (AI)
- **Code 10**: Canga (CG).

The deposit is divided into a supergene layer (surficial canga) and underlying ferruginous rocks. On the one hand, CG is a layer of well-consolidated rock composed mainly of goethite derived from weathering of the iron formation, with high iron grade, alumina grade and loss on ignition. On the other hand, the underlying rocks are subdivided on the basis of their granulometry and contents of iron, alumina, manganese and loss on ignition. Specifically, iron is mainly present in the form of hematite and itabirite. The former is an oxide-facies formation with high iron grade (most often above 62%), while the latter is a laminated, metamorphosed oxide-facies formation in which iron is present as thin layers of hematite,

magnetite or martite, with grade generally less than 62% (Dorr, 1964). The alumina and phosphorus grades are high in ALH, ALI, MNI and AI, with the last two rock types also having high manganese grade and high loss on ignition, respectively. Finally, the compact rock types CH and CI exhibit a coarse granulometry, with values of *G* most often above 50%, while the other rock types are associated with finer granulometry (*G* mostly below 50%) (Figure 4.2). These suggest that the quantitative variables are closely related to the rock type definition and that the proposed approach will provide an opportunity to detect inconsistencies between the former and the latter. In the following subsection, the methodology proposed in Section 4.2 is applied step-by-step. A general workflow is presented in Figure 4.3.

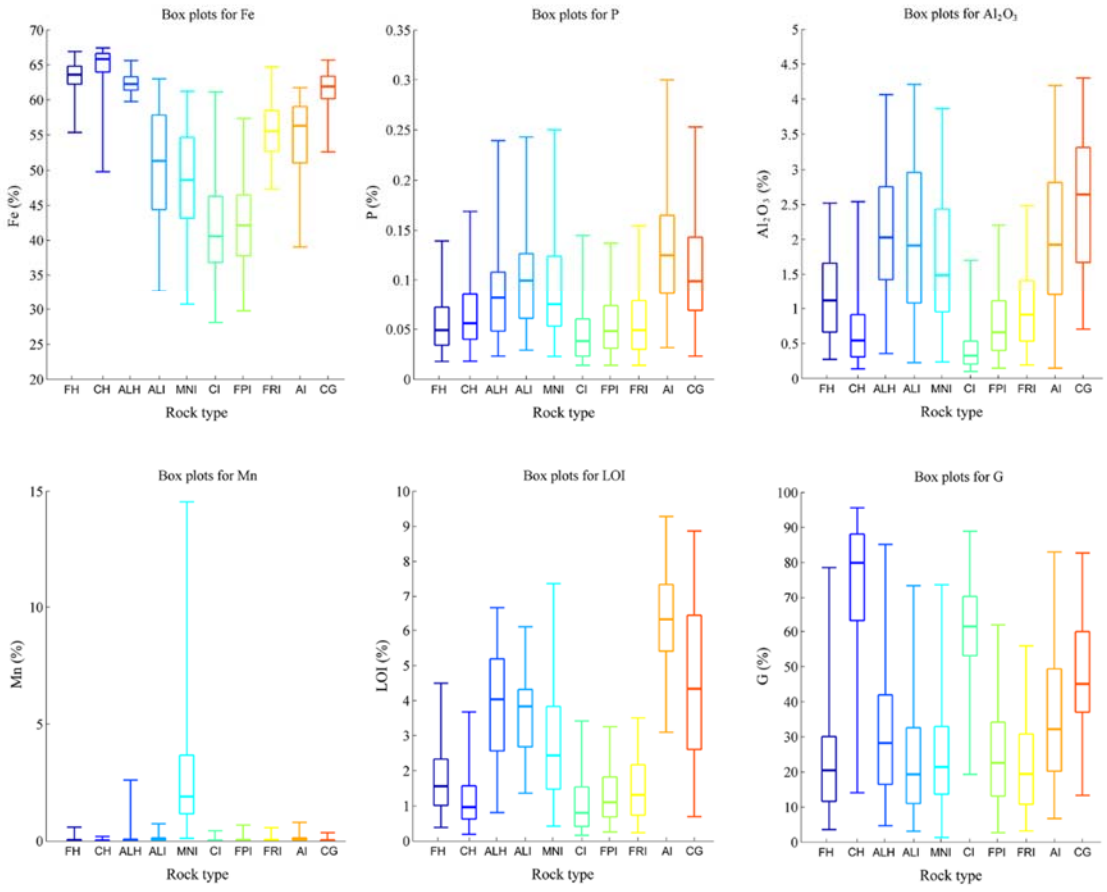


Figure 4.2: Box plots showing quantiles Q2.5, Q25, Q50, Q75 and Q97.5 of the distributions of iron grade, phosphorus grade, alumina grade, manganese grade, loss on ignition and granulometry, for each rock type.

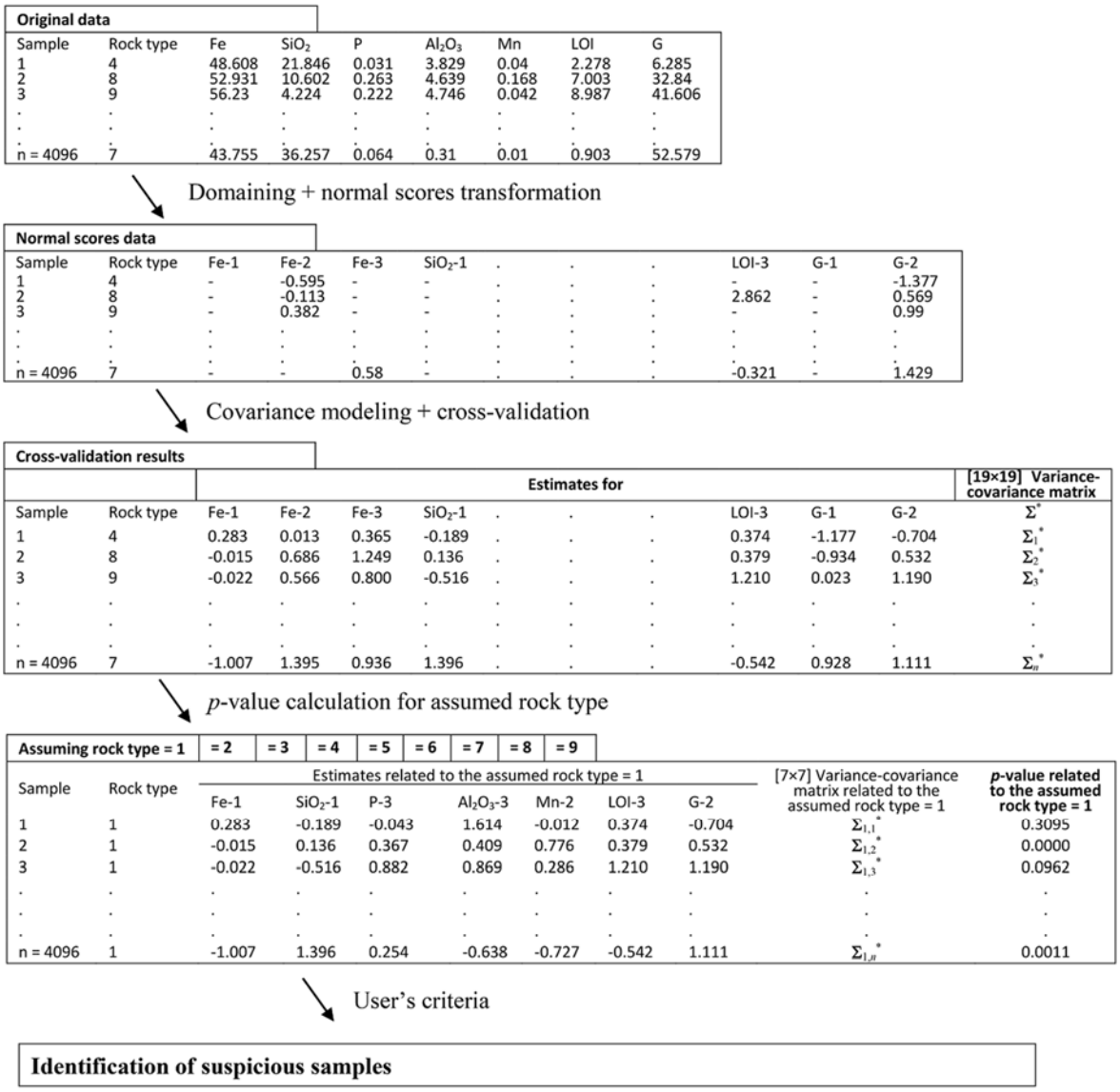


Figure 4.3: Workflow of proposed methodology for identifying mislogged rock types.

4.3.2. Geological domaining

The first modeling step involves partitioning the deposit into domains in which the quantitative variables are homogeneously distributed and can be represented by stationary random fields (Rossi and Deutsch, 2014). To this end, for each quantitative variable, the rock types are grouped into domains based on the following considerations:

- The results of contact analysis, which aims to determine the behavior of the mean value of a quantitative variable in the neighborhood of the boundary between two rock type domains. Some examples are provided in Figure 4.4, where the mean values

are seen to change abruptly when crossing a rock type boundary, indicating that the rock types on either side of the boundary belong to different domains.

- The data statistics (mainly the mean value and the standard deviation) within each rock type. The rock types belonging to the same domain are expected to have similar statistics. An example is given in Figure 4.5 for the iron grade and loss on ignition.

- The location and contact relationships of the rock types in the deposit. In particular, CG appears as a specific domain for all the quantitative variables because of its surficial position and because the data statistics in CG differ from those in any other rock type (Figure 4.2).

- The lithological controls on the variable under study. For example, hematites and itabirites are likely to form two different domains for the iron and silica grades, but not necessarily for the other quantitative variables. Likewise, the compact and friable rock types are likely to define two different domains for the granulometry, but not for the remaining quantitative variables.

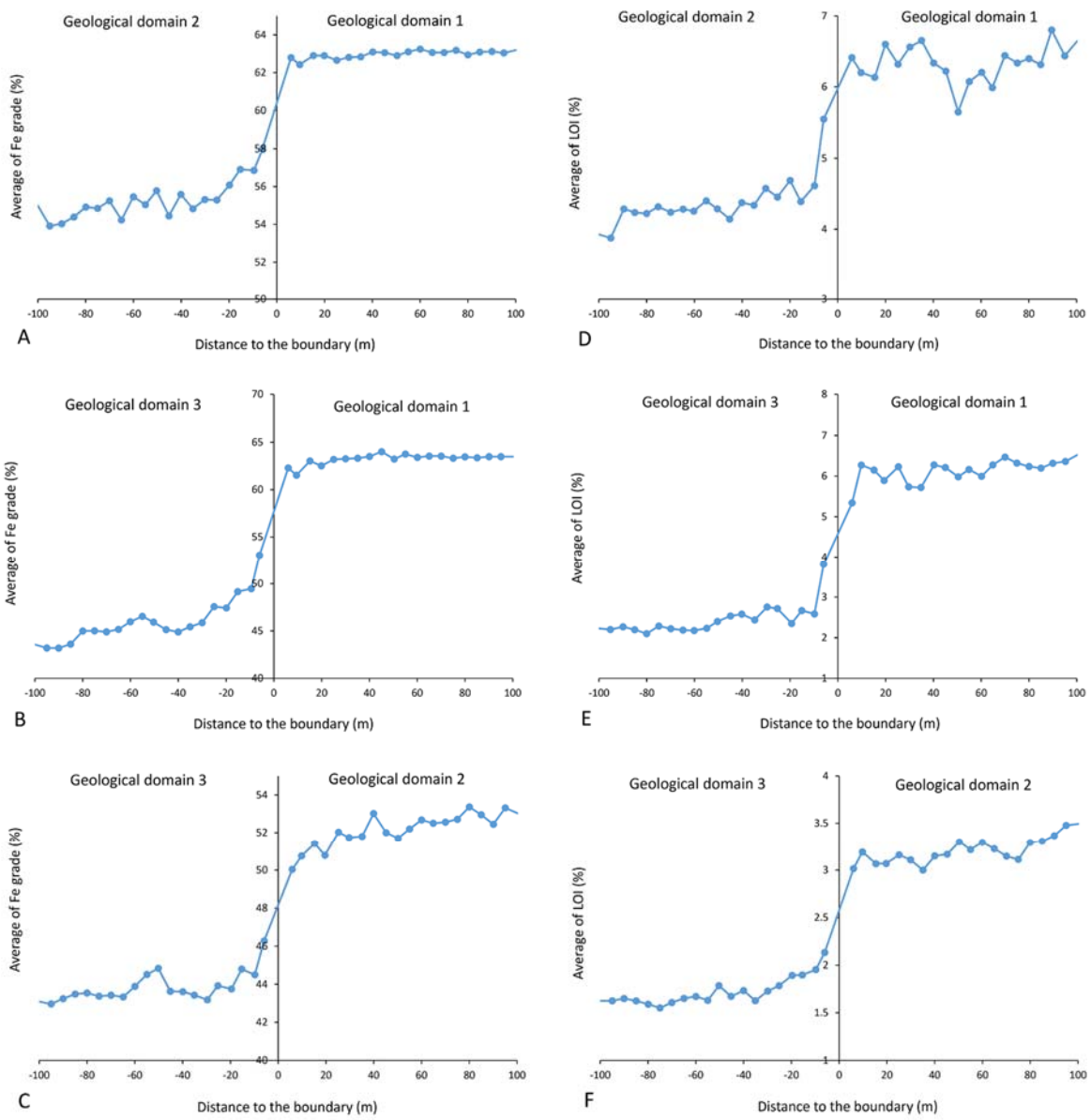


Figure 4.4: Contact analysis for Fe (A, B and C) near the boundaries of geological domains 1 (FH + CH + ALH), 2 (ALI + MNI + FRI + AI) and 3 (CI + FPI), and for LOI (D, E and F) near the boundaries of geological domains 1 (AI), 2 (ALH + ALI + MNI) and 3 (FH + CH+ CI + FPI + FRI).

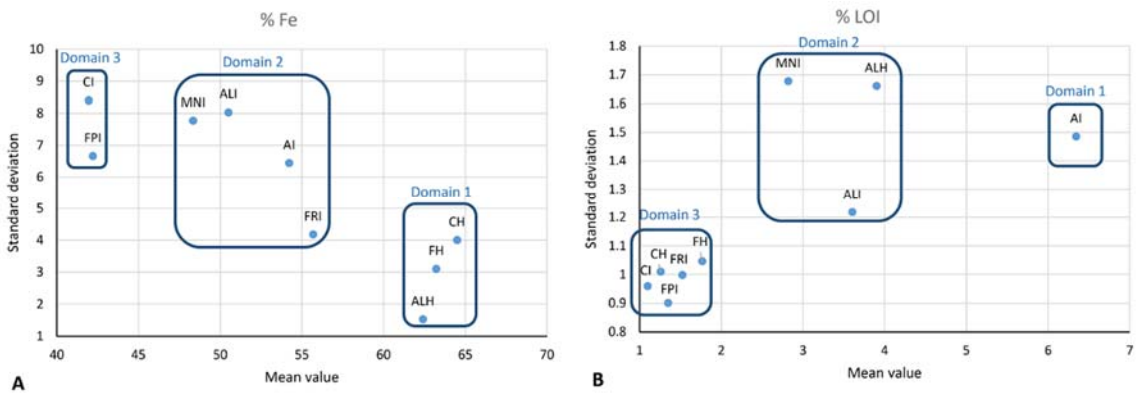


Figure 4.5: Mean values and standard deviations of Fe (A) and LOI (B) for the underlying rock types (excluding canga), and grouping of these rock types into geological domains.

In the following, the study will focus on the underlying ferruginous rocks, i.e., we will no longer consider surficial canga, as its characterization depends more on the geographical position than on the quantitative variables. Table 4.1 gives a summary of the geological domains defined in the underlying rocks, where each quantitative variable is associated with two or three domains.

Table 4.1: Geological domaining for each quantitative variable

	FH	**CH**	**ALH**	**ALI**	**MNI**	**CI**	**FPI**	**FRI**	**AI**
Fe	1	1	1	2	2	3	3	2	2
SiO_2	1	1	1	2	2	3	3	2	2
P	3	3	2	2	2	3	3	3	1
Al_2O_3	3	3	2	2	2	3	3	3	1
Mn	2	2	2	2	1	2	2	2	2
LOI	3	3	2	2	2	3	3	3	1
G	2	1	2	2	2	1	2	2	2

4.3.3. Normal scores transformation and coregionalization modeling

Overall, there are 19 combinations of quantitative variables and geological domains: Fe-1, Fe-2, Fe-3, SiO_2-1, SiO_2-2, SiO_2-3, P-1, P-2, P-3, Al_2O_3-1, Al_2O_3-2, Al_2O_3-3, Mn-1, Mn-2, LOI-1, LOI-2, LOI-3, G-1 and G-2. For each variable and each domain, the corresponding data are normal score transformed. The direct and cross covariances of the normal scores data are then calculated along the horizontal and vertical directions, identified as the main anisotropy directions, and fitted with a full linear coregionalization model thanks to a semi-automated fitting algorithm (Emery, 2010a). The model uses a nugget effect and a set of nested exponential covariances with a geometric anisotropy between the horizontal and vertical directions, as its basic structures (Figure 4.6).

Note that the normal scores associated with different quantitative variables and/or with different geological domains are assumed to be dependent, i.e., the linear coregionalization model assumes the existence of spatial cross correlations between all the 19 underlying Gaussian random fields. Accordingly, even if a quantitative variable is discontinuous across the boundary between two different domains, its values within these two domains are not necessarily independent (a situation sometimes referred to as a "transitional" boundary).

The above model parameters (normal score transformation tables and theoretical direct and cross covariances) will not change in all the subsequent stages of the study, insofar as we assumed that only a small proportion of the samples is mislogged, hence their effect on the model parameters should be marginal.

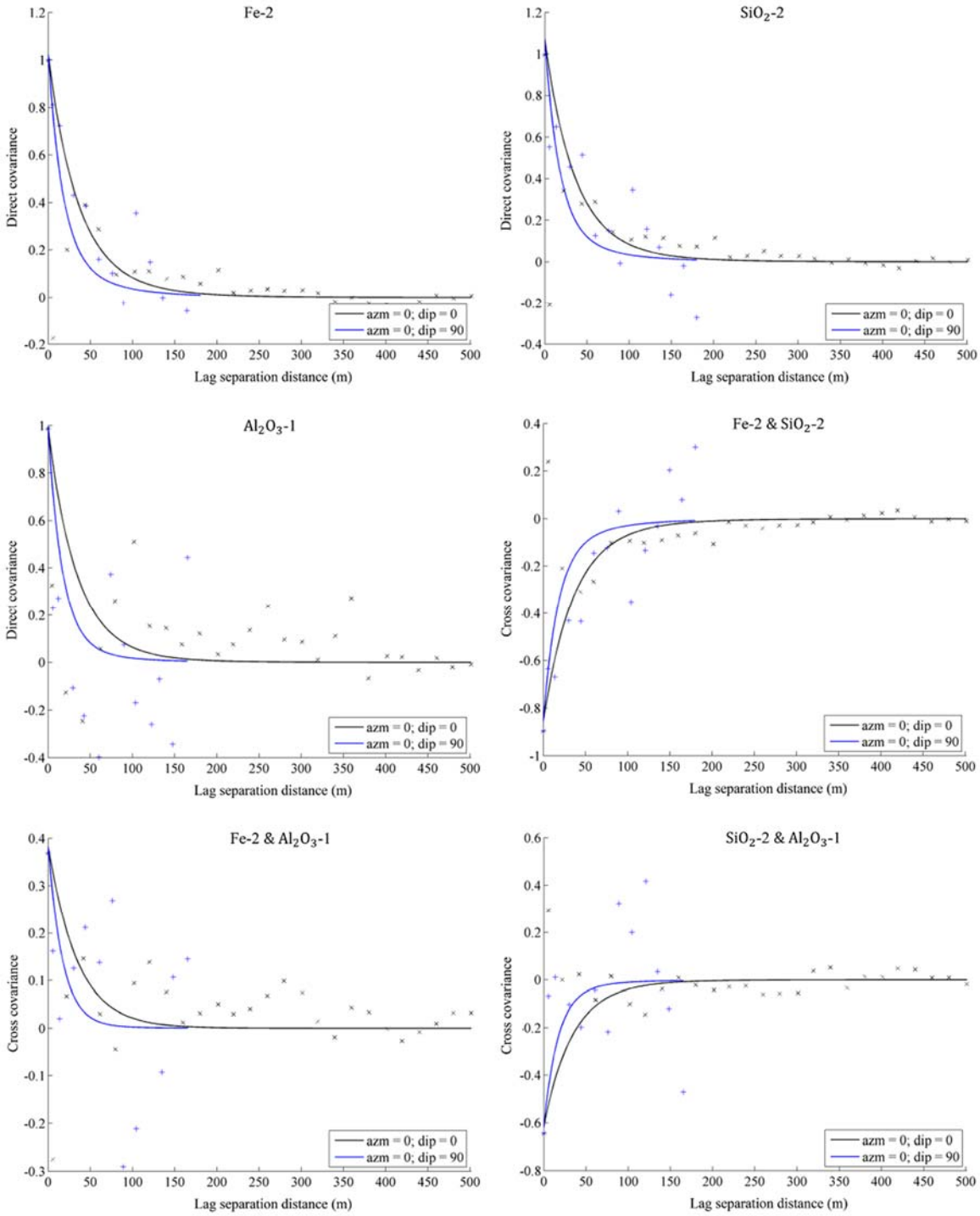


Figure 4.6: An example of fitted direct covariances and cross-covariances for the transformed iron grade in domain 2 (Fe-2), transformed silica grade in domain 2 (SiO_2-2) and transformed alumina grade in domain 1 (Al_2O_3-1).

4.3.4. Cross validation

For each drill hole sample, the 19 Gaussian random fields are predicted by simple cokriging, using a vector of zero means, the previously fitted linear coregionalization model and a moving neighborhood (in the present case, a ball with a radius of 300 meters, divided into octants, in each of which the 4 closest samples are searched) as input parameters, and the normal scores available at these closest samples as input data. Cokriging yields a 19×1 vector of predictions and a 19×19 variance-covariance matrix of prediction errors for each drill hole sample.

4.3.5. Definition of consistency measures (p-values)

For each drill hole sample (index $a = 1,\ldots, 4096$) and each rock type (index $k = 1,\ldots, 9$):

- Assume that rock type k prevails at the sample under consideration.
- In view of the geological domains associated with the quantitative variables (Table 4.1), identify which of the 19 Gaussian random fields are defined for sample a.
- Transform the measured values of the original quantitative variables (grades, loss on ignition and granulometry) into normal scores, using the transformation tables associated with the adequate geological domains. Obtain the "true" Gaussian vector $\mathbf{z}_{k,a}$ of size 7×1. For example, for $k = 1$ (friable hematite FH), the Gaussian values associated with Fe-1, SiO_2-1, P-3, Al_2O_3-3, Mn-2, LOI-3 and G-2 are obtained, while the remaining ones are not.
- From the cross-validation results, extract the 7×1 vector of predictions ($\mathbf{z}_{k,a}^*$) and the 7×7 variance-covariance matrix of prediction errors ($\mathbf{\Sigma}_{k,a}^*$).
- Calculate the p-value $p_{k,a}$, as per equation (4.3).

4.3.6. Detection of suspicious data

To detect samples that may be mislogged, the following criteria based on the calculated p-values are considered:

- The p-value associated with the class that has been actually logged by the geologist should be less than 0.01, i.e., the logged class is very unlikely.
- The p-value associated with some other class is more than 0.3 (quite likely).

In addition, to get results that are consistent with the geological zonation of the deposit, a geographical criterion is also considered, namely, that the alternative (likely) class has been logged at some drill hole sample distant less than 50 meters from the suspicious sample.

Accordingly, each sample meeting all three mentioned conditions represents a suspicious data that should be examined on a case-by-case basis. Note that the geographical distance limit (50 m) and the *p*-value limits (0.01 and 0.3) can be freely modified by the practitioner, so that one could be more or less conservative in finding such suspicious data.

4.3.7. Results and discussion

The application of the previous criteria yields a total of 2.42% of the drill hole samples identified as suspicious. In the following, a few of these suspicious samples are discussed in more details, in the light of their logged rock types, values of the quantitative variables (grades, loss on ignition and granulometry) (Table 4.2), *p*-values (Table 4.3) and information of the 10 closest samples (Table 4.4).

Table 4.2: Geochemical assay results for specified suspicious samples

Sample number	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G
50	58.334	3.351	0.23	4.184	0.034	7.273	61.465
61	58.575	3.34	0.221	4.051	0.034	7.139	62.23
110	61.293	5.146	0.026	3.12	0.02	1.935	13.9
138	36.643	4.958	0.069	2.431	24.729	7.077	41.5
146	59.735	3.414	0.058	3.738	0.014	6.263	84.18
1910	62.103	2.395	0.151	1.011	2.878	2.453	13.691

Table 4.3: Calculated *p*-values for all possible rock types of specified suspicious samples

Sample number	FH (1)	CH (2)	ALH (3)	ALI (4)	MNI (5)	CI (6)	FPI (7)	FRI (8)	AI (9)
50	0.316	0.292	0.983	0.865	0.000	0.006	0.007	0.114	0.003
61	0.027	0.021	0.000	0.000	0.000	0.001	0.001	0.069	0.988
110	0.006	0.001	0.881	0.143	0.010	0.000	0.001	0.000	0.062
138	0.000	0.000	0.000	0.000	0.304	0.000	0.000	0.000	0.000

146	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.913
1910	0.599	0.000	0.327	0.142	0.175	0.001	0.218	0.302	0.003

Table 4.4: Distance, sample number and logged rock type of the 10 closest samples to the specified suspicious samples

Sample Number		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
50	Distance (m)	5.1	10.2	20.2	30.2	40.2	45.8	48.8	50.8	53.9	57.6
	Sample no.	61	936	1042	568	614	615	563	591	321	917
	Rock type	3	3	3	3	3	7	3	3	3	3
61	Distance (m)	5.1	5.1	15.1	25.1	35.1	40.7	44.8	49.9	52.3	53.5
	Sample no.	50	936	1042	568	614	615	563	591	321	917
	Rock type	9	3	3	3	3	7	3	3	3	3
110	Distance (m)	5.1	5.2	13.5	13.7	18.8	20.3	25.6	37.7	41.5	44.6
	Sample no.	2472	109	118	664	604	830	446	32	44	45
	Rock type	8	3	4	7	1	4	1	3	9	9
138	Distance (m)	6.1	6.1	14.3	16.1	25.9	43.0	50.6	58.8	191.3	191.4
	Sample no.	137	1287	721	661	1076	714	1026	1934	2169	1115
	Rock type	5	1	1	1	1	1	8	7	5	5
146	Distance (m)	5.3	7.2	11.8	17.2	18.2	27.2	34.0	40.9	49.2	50.1
	Sample no.	145	1373	516	2433	358	2370	3358	1328	13	979
	Rock type	9	1	3	1	3	2	2	3	3	3
1910	Distance (m)	6.9	9.8	19.7	21.8	29.0	31.8	35.7	39.6	43.9	46.9
	Sample no.	1942	2716	3126	2311	2043	3382	3257	3374	3182	1083
	Rock type	3	2	2	3	2	3	2	3	1	8

Sample n°61: This sample has been logged by the geologist as rock type 3 (ALH), but the *p*-value of this rock type is close to 0 and the *p*-value of rock type 9 (AI) is very high (about 0.99) (Table 4.3). Looking at the assayed grades in Table 4.2, the measured iron grade appears to be abnormally low for an ALH, while the high loss on ignition agrees well with an AI. Therefore, it seems that sample n°61 is mislogged and should be relogged as rock type 9.

Sample n°50: This sample is the opposite case of sample n°61, as it has been logged as rock type 9 (AI), but the *p*-value of this rock type is close to zero and that of ALH (rock type 3) is close to 0.98 (Table 4.3). However, the measured grades (Table 4.2) are very similar to that of sample n°61 and agree well with rock type 9. Actually, sample n°61 (mislogged as rock

type 3) turns out to be the closest one (Table 4.4), therefore the most influential in the cross-validation procedure, which explains why rock type 9 appears so likely. One concludes that sample n°50 is logged correctly and that the calculated consistency measures (*p*-values) are distorted due to the mislogged neighboring sample.

Sample n°110: This sample has been logged as rock type 8 (FRI), but our criteria suggest changing it to rock type 3 (ALH). Based on the assays (Table 4.2) and the existence of samples with rock type 3 in the neighborhood (Table 4.4), it seems that sample n°110 is effectively mislogged and should be changed to rock type 3.

Sample n°138: This sample has been logged as rock type 1 (FH), but may be relogged as 5 (MNI). This relogging would definitely agree with the measured manganese grade (more than 24%) (Table 4.2) and with the fact that the closest sample (sample n°137, less than 10 meters away) is also logged as rock type 5 (Table 4.4). Note that another very close sample (n°1287) is logged as rock type 1 (Table 4.4), but the manganese grade measured for this sample is substantially lower (0.26%) and the iron grade substantially higher (60.95%), consistent with FH (code 1). Therefore, sample n°138 should be relogged as MNI (code 5), the same as its neighbor sample n°137, while sample n°1287 is correctly logged as rock type 1.

Sample n°146: This sample has been logged as rock type 1 (FH), but could be relogged as rock type 9 (AI). From Table 4.2, it seems that rock type 9 is more likely (owing to the high value of LOI) but not decisively (owing to the high value of Fe and low value of SiO_2, more compatible with a hematite). From Table 4.4, one observes that the closest sample (sample n°145) whose assays are very close to that of sample n°146 is coded as rock type 9, which may explain the *p*-value obtained for rock type 9. But, as for sample n°146, it cannot be said decisively whether rock type 9 or 5 is correct for sample n°145, so it is advised to physically check and relog these two samples.

Sample n°1910: This sample has been originally logged as rock type 2 (CH) but, according to the calculated *p*-values, it could be rock type 1 (FH), 3 (ALH) or 8 (FRI) (Table 4.3). From Table 4.2, it can be understood that codes 2 and 8 are not adequate for this sample (because of the too low granulometry and the too high iron grade, respectively). Finally, it is suggested to discard codes 2 and 8 and to physically check this sample to choose between codes 1 and 3.

4.4. Conclusions

There is an increasing need in the mining industry for high performance of mineral resource models. A portion of model deviations are caused by mislogged samples, for which the logged value of a petrophysical attribute such as the lithology, alteration or mineralogical assemblage is erroneous.

Considering the regionalized nature of petrophysical attributes and their dependence relationships with quantitative variables from geochemical analyses or metallurgical tests, a geostatistical approach based on leave-one-out cross-validation has been proposed for identifying possible mislogged samples. The proposal is aimed at calculating, for each sample, a measure of consistency between the logged classes and the quantitative covariates. It is worthwhile to mention that the application is not black-boxed and allows the practitioner to include additional criteria (e.g., *p*-value limits and geographical criteria) to detect suspicious samples. Also, because some samples may have all their measures of consistency smaller than the chosen *p*-value limit and may therefore not be classified into any of the logged classes, the proposed methodology cannot be used blindly as a substitute for the original logs.

To illustrate the applicability of the proposal, a case study from an iron ore deposit has been presented, where the logged rock types are closely related with seven quantitative variables (grades of iron, silica, phosphorus, alumina, manganese, loss on ignition and granulometry) measured on the same set of exploration drill holes. The samples detected as suspicious have been carefully checked on the basis of their logged classes and quantitative covariates, as well as on the basis of the information of the neighboring samples, in order to confirm or reject the correctness of the original rock type logs.

The proposed approach can be applied in several geometallurgical contexts, for example by using geochemical data as covariates for defining lithologies, ore mineralogical data from QEMSCAN (Quantitative Evaluation of Minerals by SCANning electron microscopy) analyses for finding mineral zones, or gangue mineralogical data derived from spectroscopy for recognizing alterations, or all the previous types of data as well as metallurgical tests for identifying geometallurgical domains. In all these contexts, the identification of misslogged samples can be beneficial for the overall performance of the value chain of the mining business and provide criteria to define samples that should be part of a relogging campaign.

Acknowledgments

The authors are grateful to Alejandro Cáceres from University of Chile, as well to the reviewers and editors of the journal, for their constructive comments on a previous version of this work. The funding from the Chilean Commission for Scientific and Technological Research, through Projects CONICYT / FONDECYT / REGULAR / N°1130085 and CONICYT PIA Anillo ACT1407, is also acknowledged.

4.5. Bibliography

Agterberg, F.P., 1990. Automated stratigraphic correlation. Elsevier, Amsterdam, 423 pp.

Bourgine, B., Prunier-Leparmentier, A.M., Lembezat, C., Thierry, P., Luquet, C., Robelin, C., 2008. Tools and methods for constructing 3D geological models in the urban environment. The Paris case. In: Ortiz, J.M., Emery, X. (eds.) Proceedings of the Eighth International Geostatistics Congress. Gecamin Ltda, Santiago, pp. 951-960

Bourgine, B., Lasseur, E., Leynet, A., Badinier, G., Ortega, C., Issautier, B., Bouchet, V., 2015. Building a geological reference platform using sequence stratigraphy combined with geostatistical tools. Geophysical Research Abstracts 17 EGU2015-8292

Cáceres, A., Emery, X., 2013. Geostatistical validation of geological logging. In: Ambrus, J., Beniscelli, J., Brunner, F., Cabello, J., Ibarra, F. (eds.) Proceedings of the Third International Seminar on Geology for the Mining Industry. Gecamin Ltda, Santiago, pp. 73-80

Chilès, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 699 pp.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, 369 pp.

Deutsch, J.L., Palmer, K., Deutsch, C.V., Szymanski, J., Etsell, T.H., 2016. Spatial modeling of geometallurgical properties: techniques and a case study. Natural Resources Research 25: 161-181.

Dorr, J.V.N., 1964. Supergene iron ores of Minas Gerais, Brazil. Economic Geology 59(7): 1203-1240.

Dunn, K.J., Bergman, D.J., Latorraca, G.A., 2002. Nuclear Magnetic Resonance – Petrophysical and Logging Applications. Pergamon, Amsterdam.

Emery, X., 2010a. Iterative algorithms for fitting a linear model of coregionalization. Computers & Geosciences 36(9): 1150-1160.

Ewusi, A., Kuma, J.S., 2011. Calibration of shallow borehole drilling sites using the electrical resistivity imaging technique in the granitoids of Central Region, Ghana. Natural Resources Research 20: 67-63.

Glacken, I.M., Snowden, D.V., 2001. Mineral resource estimation. In: Edwards, A.C. (ed.) Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice. The Australasian Institute of Mining and Metallurgy, Melbourne, pp. 189-198.

Haldar, S.K., 2013. Mineral Exploration: Principles and Applications. Elsevier, Oxford, 334 pp.

Hassibi, M., Ershaghi, I., Aminzadeh, F., 2003. High resolution reservoir heterogeneity characterization using recognition technology. Developments in Petroleum Science 51: 289-307.

Hearst, J.R., Nelson, P.H., Paillet, F.L., 2000. Well Logging for Physical Properties. Wiley, Chichester.

Hoyle, I.B., 1986. Computer techniques for the zoning and correlation of well-logs. Geophysical prospecting 34(5): 648-664.

Knödel, K., Lange, G., Voigt, H.J., 2007. Environmental Geology: Handbook of Field Methods and Case Studies. Springer, Berlin, 1357 pp.

Luthi, S.M., 2001. Geological Well Logs: Their Use in Reservoir Modeling. Springer, Berlin, 373 pp.

Luthi, S.M., Bryant, I.D., 1997. Well-log correlation using a back-propagation neural network. Mathematical Geology 29(3): 413-425.

Maleki, M., Emery, X., 2015. Joint simulation of grade and rock type in a stratabound copper deposit. Mathematical Geosciences 47: 471-495.

Manchuk, J.G., Deutsch, C.V., 2012. Applications of data coherency for data analysis and geological zonation. In: Abrahamsen, P., Hauge, R., Kolbjørnsen, O. (eds.) Geostatistics Oslo 2012. Springer, pp. 173-184.

Marjoribanks, R., 2010. Geological Methods in Mineral Exploration and Mining. Springer, Berlin, 238 pp.

Moon, C.J., Whateley, M.K.G., Evans, A.M., 2006. Introduction to Mineral Exploration. Blackwell Scientific Publications, Oxford, 481 pp.

Rossi, M.E., Deutsch, C.V., 2014. Mineral Resource Estimation. Springer, Heidelberg, 332 pp.

Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation. Cambridge University Press, Cambridge, 400 pp.

Soleimani, M., Shokri, B.H., Rafiei, M., 2016. Integrated petrophysical modeling for a strongly heterogeneous and fractured reservoir, Sarvak Formation, SW Iran. Natural Resources Research, doi:10.1007/s11053-016-9300-9.

Soltani, S., Hezarkhani, A., 2011. Determination of realistic and statistical value of the information gathered from exploratory drilling. Natural Resources Research 20: 207-216.

Spies, B.R., 1996. Electrical and Electromagnetic Borehole Measurements: A Review. Surveys in Geophysics 17: 517-556.

Taylor, G.R., 2000. Mineral and lithology mapping of drill core pulps using visible and infrared spectrometry. Natural Resources Research 9: 257-268.

Theys, P., 1999. Log Data Acquisition and Quality Control. Editions Technip, Paris, 480 pp.

Verly, G., 1983. The multigaussian approach and its application to the estimation of local reserves. Mathematical Geology 15(2): 259-286.

Wackernagel, H., 2003. Multivariate Geostatistics: an Introduction with Applications. Springer, Berlin, 387 pp.

Yunsel, T.Y., Ersoy, A., 2011. Geological modeling of gold deposit based on grade domaining using plurigaussian simulation technique. Natural Resources Research 20: 231-249.

Chapter 5: Designing a geostatistical-based method for validating geological interpretations using quantitative covariates

The layout of geological domains in an ore deposit is the result of an interpretation from the resources geologists and, as for the geological logs, can be inconsistent with the information carried by quantitative covariates measured at surrounding boreholes. Based on the spatial dependence relationships between the geological domains and the quantitative covariates and on the soft boundary model (model 3) presented in Chapter 3, we here propose a geostatistical approach to construct simulated geological scenarios and to validate an interpreted geological model by identifying the areas of a deposit where the interpreted category is not in agreement with the quantitative variables. The contents of this chapter have been published in the journal *Minerals*:

Adeli, A., Emery, X., Dowd, P., 2018. Geological modelling and validation of geological interpretations via simulation and classification of quantitative covariates. Minerals 8(1), 7.

Geological modelling and validation of geological interpretations via simulation and classification of quantitative covariates

Abstract

This paper proposes a geostatistical approach for geological modelling and for validating an interpreted geological model, by identifying the areas of an ore deposit with a high probability of being misinterpreted, based on quantitative coregionalised covariates correlated with the geological categories. This proposal is presented through a case study of an iron ore deposit at a stage where the only available data are from exploration drill holes. This study consists of jointly simulating the quantitative covariates with no previous geological domaining. A change of variables is used to account for stoichiometric closure, followed by projection pursuit multivariate transformation, multivariate Gaussian simulation, and conditioning to the drill hole data. Subsequently, a decision tree classification algorithm is used to convert the simulated values into a geological category for each target block and realisation. The determination of the prior (ignoring drill hole data) and posterior (conditioned to drill hole data) probabilities of categories provides a means of identifying the blocks for which the interpreted category disagrees with the simulated quantitative covariates.

Key words: geological uncertainty; geological modelling; geological misinterpretation; geostatistical simulation; classification.

5.1. Introduction

A geological model consists of a three-dimensional representation of an ore deposit constructed by resource geologists on the basis of their knowledge of the deposit, geological field observations, geophysical surveys, and drill hole logs and assays. A geological model that represents the spatial locations and extents of rock types or ore types is an essential input for mineral resources evaluation and mine planning and, as such, affects all subsequent stages of the mining process (Duke and Hanna, 2001; Sinclair and Blackwell, 2002; Knödel et al., 2007; Marjoribanks, 2010; Rossi and Deutsch, 2014). The typical workflow for assessing mineral resources consists of grouping the rock types or ore types into geological domains in which the quantitative variables of interest (geochemical, geometallurgical, and/or geomechanical variables) are assumed to be homogeneously distributed and then interpolating these variables within each domain using geostatistical techniques. This hierarchical workflow accounts for geological controls on the distributions of the quantitative

variables but produces clear-cut discontinuities in the values of the quantitative variables when crossing the domain boundaries (Ortiz and Emery, 2006; Rossi and Deutsch, 2014). Several alternatives have been proposed to mitigate these discontinuities and to account for spatial correlation across the domain boundaries (Larrondo et al., 2004; Ortiz and Emery, 2006; Vargas-Guzmán, 2008; Séguret, 2013; Mery et al., 2017). Another approach to produce gradual transitions near the domain boundaries is to model the quantitative variables of interest with no previous geological domaining by considering the controlling rock types or ore types as cross-correlated covariates (Dowd, 1994; Dowd, 1997; Emery and Robles, 2009; Emery and Silva, 2009; Maleki and Emery, 2015; Maleki and Emery, 2017).

Geostatistical simulation approaches have been designed to construct several geological scenarios in order to quantify uncertainty in the actual locations and extents of rock types or ore types, accounting for their spatial continuity and proportions (which may vary in space), and contact relationships, including chronological associations, allowable and non-allowable contacts, edge effects (preferential contacts), and directional effects (asymmetrical spatial relationships) between rock types or ore types ((Xu et al., 2006; Armstrong et al., 2011; Mariethoz and Caers, 2014; Beucher and Renard, 2016) and references therein). However, these approaches are still in their infancy in practical orebody modelling where the geological model often corresponds to a single interpretation of the deposit, rather than multiple scenarios, which does not allow geological uncertainty to be measured. This motivates the need for quantitative methods to validate the model and to identify the areas of the deposit that have higher probabilities of being misinterpreted.

Most published research on geological modelling focuses on using all available data to generate a more accurate geological model, e.g., by using structural and geophysical data together in addition to drill hole data or by using data inversion methods to generate an interpreted model (Guillen et al., 2008; Lelièvre, 2009). Studies of validating geological models often concentrate on statistical and graphical analyses by comparing the models with the available data to detect inconsistencies (Rossi and Deutsch, 2014).

In this work we propose a geostatistical approach to construct simulated geological scenarios and to validate an interpreted geological model by identifying the areas of a deposit that are likely to be misinterpreted. The approach relies on the analysis of quantitative variables that are measured at sampling locations and are cross-correlated with the geological categories obtained from geological core logging. It includes the geostatistical modelling and simulation of the quantitative variables, followed by their classification into geological categories. Comparing the prior (without sampling information) and posterior (accounting for sampling information) probabilities of categories for each target location provides a means of identifying the locations that are most likely to be incorrectly interpreted.

The paper is outlined as follows: Section 5.2 comprises the case study and the methodology used to model, simulate, and classify the quantitative variables of interest. Section 5.3

presents the results of comparing the prior and posterior probabilities of geological categories to identify potentially misinterpreted blocks. Conclusions and perspectives follow in Section 5.4.

5.2. Materials and methods

5.2.1. Case study presentation

The case study is a banded-iron formation (BIF)-hosted iron ore deposit for which there are 4177 diamond drill core samples. For reasons of confidentiality, the name and location of the deposit are not disclosed and a local coordinate system is used. Seven quantitative variables have been analysed for each sample: the grades of iron (Fe), silica (SiO_2), phosphorus (P), alumina (Al_2O_3), manganese (Mn), loss on ignition (LOI), and the granulometric fraction of fragments with size above 6.3 mm (G). In addition, for each sample, the dominant rock type is available from geological logging, which is coded into ten categories: friable hematite (code 1), compact hematite (code 2), alumina-rich hematite (code 3), alumina-rich itabirite (code 4), manganese-rich itabirite (code 5), compact itabirite (code 6), friable iron-poor itabirite (code 7), friable iron-rich itabirite (code 8), amphibolitic itabirite (code 9), and canga (code 10). There are dependent relationships between the quantitative variables and the rock codes (Maleki et al., 2016), as summarised in Table 5.1. This implies that information about the former may help to detect inconsistencies in the interpretation of the latter, which is the basis of the proposed geostatistical methodology.

Based on the drill hole information and their knowledge of the deposit, the resource geologists constructed two-dimensional representations of the rock type distribution in specific plan views and cross-sections, then interpolated these representations with indicator kriging to construct a rock type model (the most probable rock type, obtained by post-processing the indicator kriging results) on a 3D grid with a regular spacing of 10 m × 10 m × 10 m (Figure 5.1). This 3D model is the basis for mineral resource evaluation and for mine planning, but does not provide any quantification of the uncertainty in the actual rock type assigned to each block. It is therefore of interest to design a method for validating the interpreted rock type assigned to each grid block and for finding the blocks for which the interpretation is likely to be mistaken.

Table 5.1. Associations between rock types and quantitative variables (for each variable, "poor" and "fine" refer to the rock types with the lowest values, and "rich" and "coarse" to the rock types with the highest values) (Maleki et al., 2016).

G	Fe	SiO_2	Al_2O_3	Mn	LOI	P	Rock Code
Coarse	Rich	Poor					2
Coarse	Poor	Rich					6
Fine	Rich	Poor	Rich				3
Fine	Rich	Poor	Poor				1
Fine	Intermediate	Intermediate	Rich	Rich			5
Fine	Intermediate	Intermediate	Rich	Poor	Rich	Rich	9
Fine	Intermediate	Intermediate	Rich	Poor	Poor	Poor	4
Fine	Intermediate	Intermediate	Poor				8
Fine	Poor	Rich					7
Intermediate	Rich	Poor	Rich	Poor	Rich		10

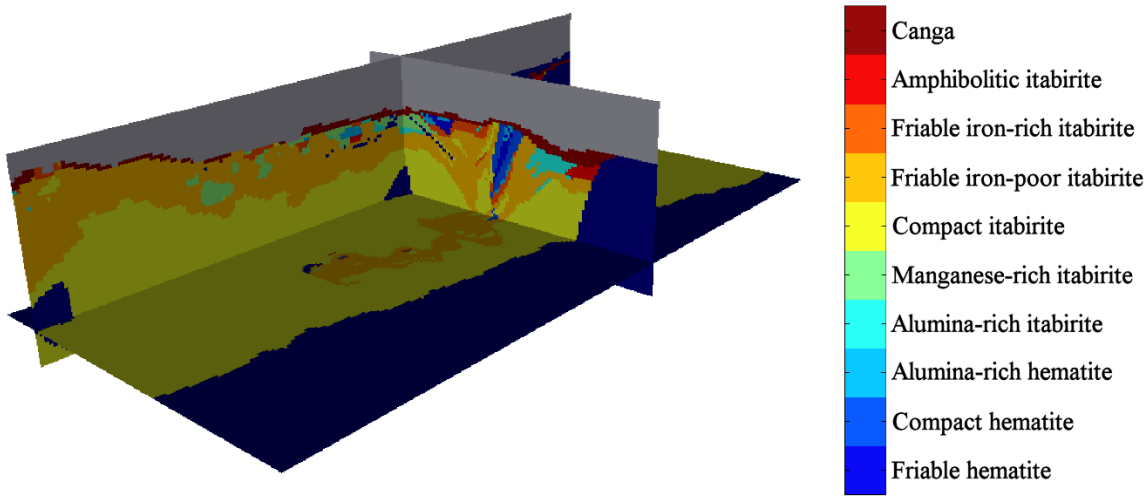


Figure 5.1: Isometric view of the interpreted rock type model, showing the plan view and vertical cross-sections passing through the origin (local coordinate system). Waste and air are shown in dark blue and grey, respectively.

In the following, we will exclude the waste rocks in the outer parts of the deposit, as well as the canga in the superficial part, as their locations and extents depend more on geographical position than on the quantitative variables. Accordingly, the following stages of the study will be restricted to the underlying ferruginous rock (rock types 1–9).

5.2.2. Modelling and simulation of quantitative variables

Adeli and Emery (2017) presented a hierarchical model for this deposit, in which the rock type controls the distribution of the quantitative variables (Fe, SiO_2, P, Al_2O_3, Mn, LOI, G) and the spatial correlation structure of these variables depends on the prevailing rock type domain. In the following, we will reverse this point of view and assume that the rock type is subordinate to the quantitative variables. In other words, the quantitative variables will be modelled and simulated throughout the deposit without any previous geological domaining. Because of the relationship between the grades, granulometry, and rock types (Table 5.1), the rock type will then be allocated on the basis of the simulated values of these quantitative variables by means of a classification algorithm. Unlike the aforementioned hierarchical model, this approach does not produce discontinuities in the values of the quantitative variables near the rock type boundaries, which conforms with the concept of a disseminated ore deposit (Emery and Silva, 2009; Maleki and Emery, 2015). In this deposit, the quantitative variables are spatially correlated across the rock type boundaries, as shown in (Adeli and Emery, 2017; Mery et al., 2017).

5.2.2.1. Change of variables based on stoichiometric closure

A joint simulation approach is required to reproduce the dependence relationships among the quantitative variables. In particular, the grade variables are linked through the following stoichiometric closure formula:

$$1.4297\ Fe + SiO_2 + 2.2913\ P + Al_2O_3 + 1.2912\ Mn + LOI = 100 \qquad (5.1)$$

in which the coefficients 1.4297, 2.2913, and 1.2912 are used to rescale the masses of iron (Fe), phosphorus (P), and manganese (Mn) to the masses of hematite (Fe_2O_3), phosphorus pentoxide (P_2O_5), and manganese monoxide (MnO), respectively. A convenient way of reproducing the stoichiometric closure in the simulated grade values is to conduct a change of variables. Some alternatives for such a change of variables are the additive logratio (alr), centred logratio (clr), or isometric logratio (ilr) transformations that are often used in compositional data analysis (Pawlowsky-Glahn and Buccianti, 2011), but these transformations are not suitable for variables that can take zero values, as is the case in the present case study. We therefore opt for a ratio transformation that does not use logarithms, as proposed in (Mery et al., 2017), where the quantitative variables are successively normalised by the residual of the closure:

$$\begin{cases} Z_1 = \dfrac{\mathrm{P}}{100} \\ Z_2 = \dfrac{1.2912\,\mathrm{Mn}}{100 - 2.2913\,\mathrm{P}} \\ Z_3 = \dfrac{\mathrm{Al_2O_3}}{100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn}} \\ Z_4 = \dfrac{\mathrm{LOI}}{100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn} - \mathrm{Al_2O_3}} \\ Z_5 = \dfrac{\mathrm{SiO_2}}{100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn} - \mathrm{Al_2O_3} - \mathrm{LOI}} \end{cases} \tag{5.1}$$

In the above equation, the original variables have been ordered from the variable with the lowest mean value (P) to the variable with the highest mean value (SiO_2) in order to minimise the distortion induced by the ratio transformation (the correlation coefficients between Z_1 and P, Z_2 and Mn, Z_3 and Al_2O_3, Z_4 and LOI, and Z_5 and SiO_2 are all greater than 0.995 (Mery et al., 2017)). The transformed variables have no stoichiometric constraint and take their values in the interval [0–1). Note that there are only five unconstrained transformed variables (Z_1–Z_5) instead of six constrained grade variables (Fe, SiO_2, P, Al_2O_3, Mn, LOI). The back-transformation is obtained from equations (5.1) and (5.2):

$$\begin{cases} \mathrm{P} = 100\,Z_1 \\ \mathrm{Mn} = \dfrac{Z_2\,(100 - 2.2913\,\mathrm{P})}{1.2912} \\ \mathrm{Al_2O_3} = Z_3\,(100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn}) \\ \mathrm{LOI} = Z_4\,(100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn} - \mathrm{Al_2O_3}) \\ \mathrm{SiO_2} = Z_5\,(100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn} - \mathrm{Al_2O_3} - \mathrm{LOI}) \\ \mathrm{Fe} = \dfrac{100 - 2.2913\,\mathrm{P} - 1.2912\,\mathrm{Mn} - \mathrm{Al_2O_3} - \mathrm{LOI} - \mathrm{SiO_2}}{1.4297} \end{cases} \tag{5.2}$$

5.2.2.2. Projection pursuit multivariate transformation

The data for the five unconstrained variables (Z_1–Z_5) and granulometry (G) are transformed into multivariate Gaussian data, hereafter called "normal scores". Because of the heteroscedastic dependence relationships between the variables prior to transformation (Figure 5.2), the normal scores transformation of each variable separately does not provide

truly multivariate Gaussian data (Chilès and Delfiner, 2012). For instance, the scatter diagram between any two transformed variables does not have an elliptical shape, which indicates that these transformed variables do not correspond to jointly Gaussian random fields. To avoid this inconvenience, a joint normal scores transformation can be used, such as stepwise conditional transformation (SCT) (Leuangthong and Deutsch, 2003), flow transformation (FT) (Van den Boogaart et al., 2016), or projection pursuit multivariate transformation (PPMT) (Friedman and Tukey, 1974; Friedman, 1987; Barnett et al., 2014). All these methods require all the variables to be known at all the data locations (isotopic sampling), which is the case in the present case study; otherwise, the data set should be completed by multivariate imputation techniques before joint normal scores transformation (Silva and Deutsch, 2016). In practice, the first two approaches are still limited to few variables (SCT) or to small data sets (FT), and for this reason we chose the third approach (PPMT) here. The PPMT transformation is based on an iterative algorithm and allows the complex dependence relationships (such as nonlinearities and heteroscedasticities) between cross-correlated variables to be removed, providing a set of new variables that are normally distributed and uncorrelated at collocated locations (Friedman and Tukey, 1974; Friedman, 1987; Barnett et al., 2014). The transformation uses declustering weights to account for the uneven positions of the drill hole data in space. For each rock type, the weights are obtained by considering the ratio of the rock type proportion in the interpreted geological model and the rock type proportion in the drill hole data. It is assumed here that the interpreted model, which is constructed from the drill hole information and geological knowledge of the deposit, is globally accurate, i.e., it provides a reliable estimate of the true rock type proportions, although it may be locally inaccurate as some blocks may be misinterpreted.

Figure 5.2 shows how PPMT transforms the joint distribution of the quantitative variables (Z_1, Z_2, Z_3, Z_4, Z_5, G) into a multi-Gaussian one. The marginal distributions (histograms) are bell-shaped, while the bivariate distributions (scatter plots) exhibit the typical circular shape of uncorrelated Gaussian variables.

5.2.2.3. Spatial continuity modelling

The PPMT transformed variables are represented by jointly stationary Gaussian random fields within the studied area. By construction, these random fields have a mean of zero, so that their finite-dimensional distributions are fully characterised by their direct and cross-covariance functions. Under an additional assumption that the cross-covariances are even functions, one can use the direct and cross-variograms as an alternative to the covariances; this additional assumption implies the absence of asymmetries, such as spatial shifts or delay effects, in the spatial cross-correlation between variables (Wackernagel, 2003).

In the first step, the spatial correlation structure of the normal scores data is inferred by calculating their experimental direct and cross-variograms (six direct variograms and fifteen

cross-variograms) along the horizontal and vertical directions, which were identified as the main anisotropy directions. The cross-variograms indicate a low correlation (not necessarily zero) between two different random fields taken at different locations (separation distances greater than zero). The direct variograms tend to a sill value close to one, which corroborates the validity of the stationarity assumption, at least at a local scale (quasi-stationarity) (Chilès and Delfiner, 2012). Based on this observation, a linear model of coregionalisation consisting of nested exponential models is fitted to the direct and cross-variograms, by using a semi-automated algorithm to find the sill matrices associated with the nested structures that minimise the squared differences between experimental and theoretical variograms (Goulard and Voltz, 1992; Wackernagel, 2003; Emery, 2010a) (Figure 5.3). A simplified model, in which the cross-variograms are exactly zero and the PPMT-transformed variables are spatially independent, could also be considered, which amounts to neglecting the cross-correlation between these variables. The full model (with non-zero cross-variograms) is used in the following, as it is not significantly more complex.

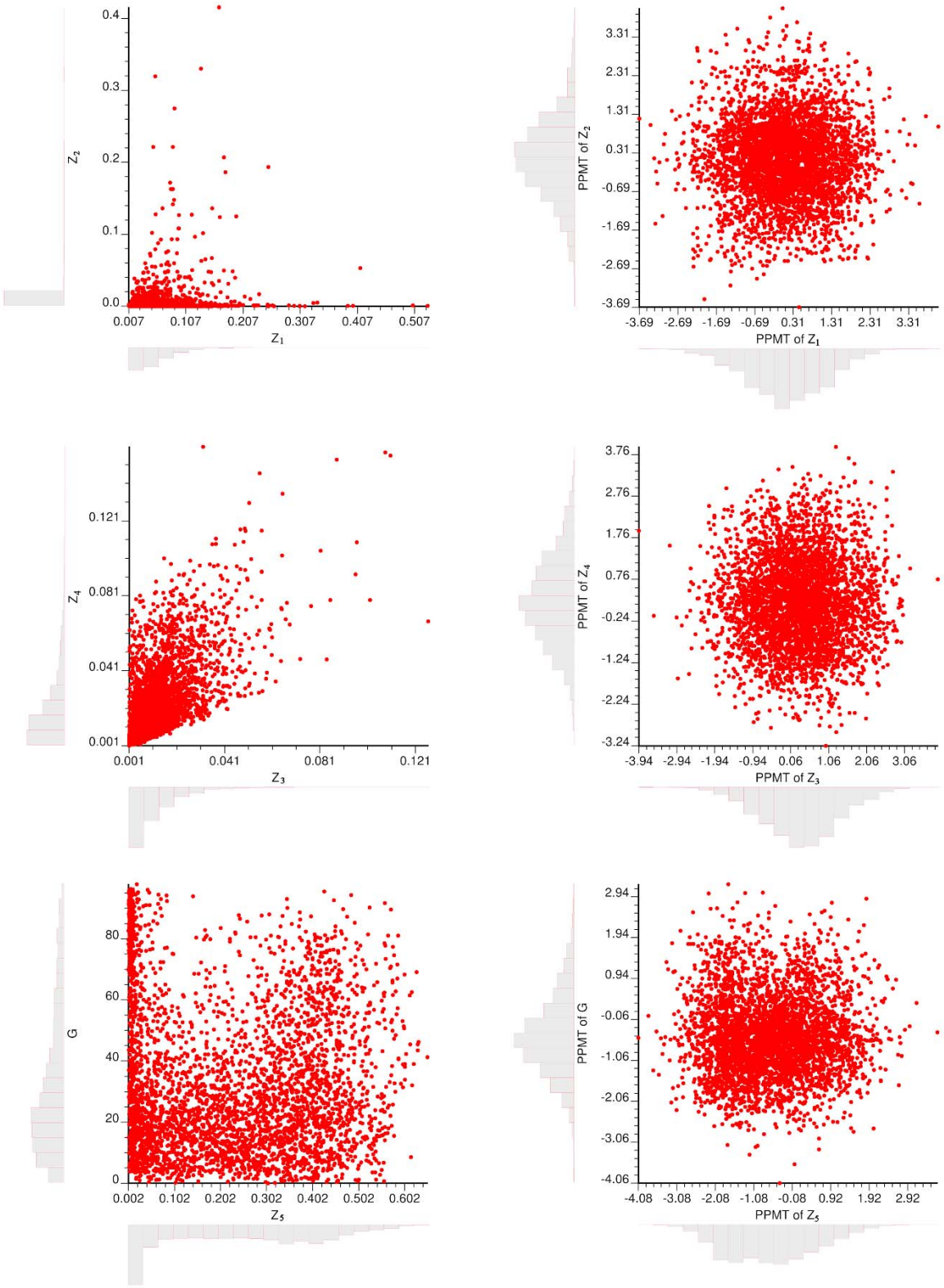


Figure 5.2: Histograms and scatterplots of Z1 vs. Z2, Z3 vs. Z4, and Z5 vs. G before (left) and after (right) PPMT transformation.

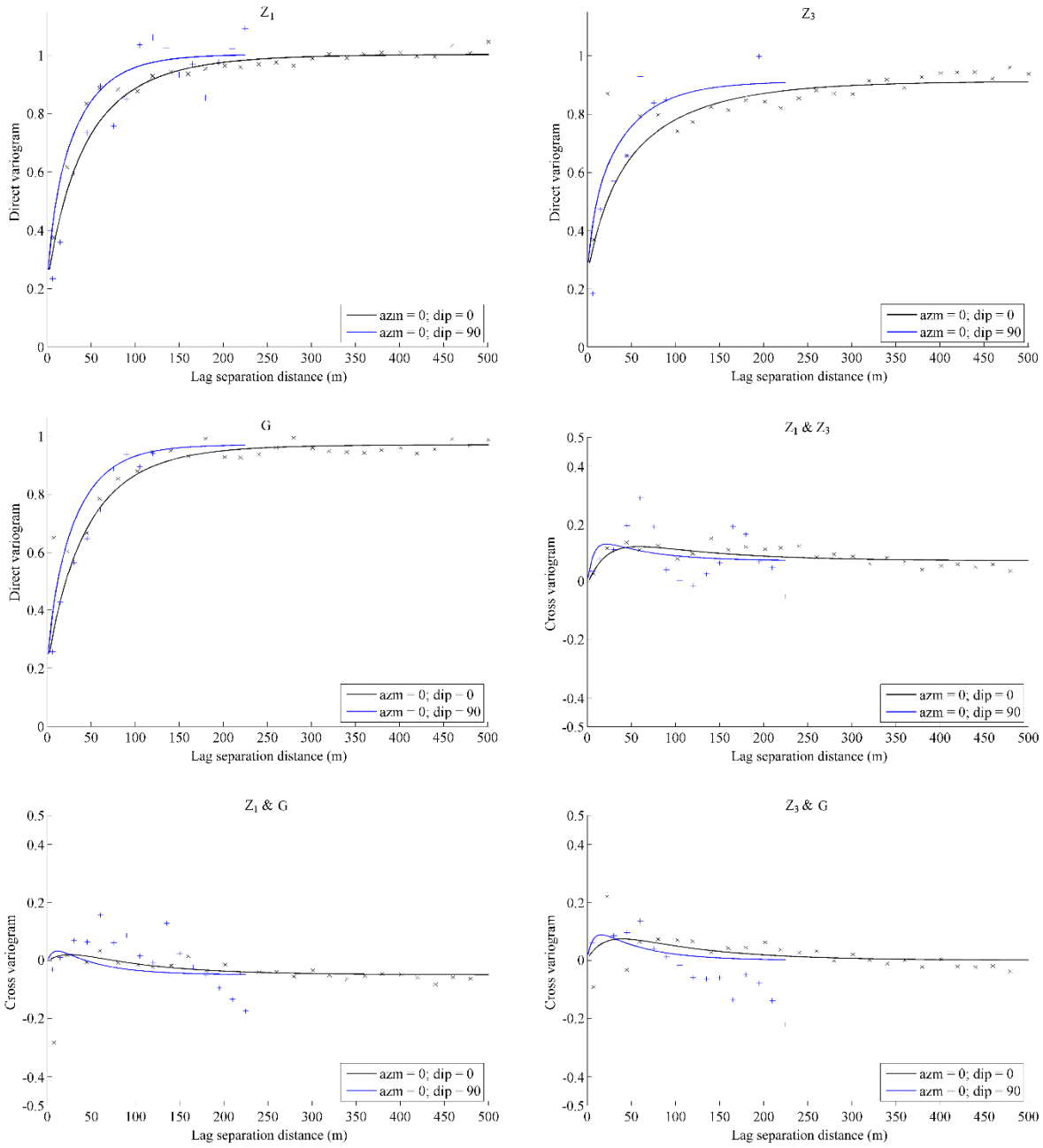


Figure 5.3: An example of experimental (crosses) and fitted (solid lines) direct and cross-variograms for the PPMT-transformed variables of Z1, Z3, and G, along the horizontal (black) and vertical (blue) directions.

5.2.2.4. Conditional simulation

The Gaussian random fields are jointly simulated using a spectral turning-bands algorithm (Emery et al., 2016). This algorithm is preferred to other alternatives, such as sequential, covariance matrix decomposition or circulant-embedding algorithms ((Chilès and Delfiner, 2012) and references therein) because of its accuracy, versatility, unequalled computational speeds, and low memory storage requirements, being able to simulate highly-multivariate

random fields and to reproduce exactly the desired spatial correlation structure (Emery et al., 2016).

Twenty realisations are constructed on the same grid as the interpreted rock type model. These realisations are then conditioned to the PPMT-transformed data known at the drill hole locations by using post-conditioning cokriging and finally back-transformed into grades and granulometry (Chilès and Delfiner, 2012). To account for possible deviations from strict stationarity, the conditioning to data is performed by ordinary cokriging, which allows the mean values of the Gaussian random fields to vary locally (i.e., at the scale of the cokriging neighbourhood) and to reproduce the spatial trends exhibited by the conditioning data (an exception would be for extrapolation situations, but grid nodes located far away from the data are not the target of the proposed methodology) (Journel and Rossi, 1989; Emery and Robles, 2009; Emery, 2010b). As an illustration, two realisations are shown in Figure 5.4. More than 20 realisations could have been constructed, but this would increase not only the computational time to run the simulation (a few hours on a common desktop) and the memory requirements to store the simulated values, but also the combination of the results to be treated (support of the multinomial distribution used to model the combinations of rock type occurrences among the realisations, see equation (5.4) in Section 5.2.5).

5.2.2.5. Checking the realisations

The correlation coefficients of drill hole data and simulated grades and granulometry are shown in Table 5.2, while Figure 5.5 displays the scatter diagrams of three pairs of variables. Both tools show that the simulated values accurately reproduce the bivariate distributions of the data values (Leuangthong et al., 2004). Furthermore, by construction, the realisations also reproduce the stoichiometric closure (imposed by the change of variables), the spatial continuity (imposed by the direct and cross-variograms), and the conditioning data (imposed by post-conditioning cokriging).

Table 5.2: Correlation coefficients of drill hole data and simulated outcomes of grades and granulometry (correlation observed on drill hole data: bold entries above main diagonal; average correlation over 20 outcomes: regular entries above main diagonal; minimum correlation over 20 outcomes: bold entries under main diagonal; maximum correlation over 20 outcomes: regular entries under main diagonal).

Variable	Fe	Si	P	Al	Mn	LOI	G
Fe	1	**−0.98**/−0.99	**0.21**/0.18	**0.27**/0.30	**−0.04**/−0.01	**0.23**/0.23	**−0.11**/−0.07
Si	**−0.99**/−0.98	1	**−0.30**/−0.29	**−0.39**/−0.42	**−0.08**/−0.08	**−0.36**/−0.37	**0.14**/0.11
P	**0.14**/0.22	**−0.32**/−0.25	1	**0.33**/0.43	**0.11**/0.16	**0.72**/0.73	**−0.07**/−0.10
Al	**0.27**/0.32	**−0.44**/−0.39	**0.41**/0.45	1	**0.19**/0.17	**0.59**/0.62	**−0.38**/−0.32
Mn	**−0.03**/0.02	**−0.10**/−0.06	**0.13**/0.20	**0.14**/0.21	1	**0.19**/0.15	**−0.06**/−0.08
LOI	**0.20**/0.27	**−0.40**/−0.35	**0.71**/0.75	**0.60**/0.64	**0.12**/0.19	1	**−0.18**/−0.13
G	**−0.12**/−0.04	**0.07**/0.16	**−0.14**/−0.08	**−0.34**/−0.30	**−0.10**/−0.05	**−0.17**/−0.10	1

5.2.3. Construction of simulated geological scenarios by classification

A classification algorithm is now used to assign a rock type from 1 to 9 for each realisation and each target grid block depending on the values of the simulated quantitative variables. To choose the algorithm, several classifiers are trained on the drill hole data set and compared through a stratified 10-fold cross-validation. Specifically, the drill hole data set (containing information on both the rock type and the quantitative variables) is divided randomly into ten subsets; each subset is held out, the classifier is trained on the remaining nine subsets and tested on the holdout subset, and its error rate is calculated. This procedure is executed in its entirety 10 times on different training subsets; ultimately, the ten error rates are averaged to get an overall error estimate (Witten et al., 2016). A geographical selection could be used instead of a random selection to define the ten data subsets, so as to reduce the redundancies between the training and testing subsets. However, the direct variograms of the normal scores data (Figure 5.3) exhibit a significant nugget effect (more than 25% of the total sill), so that the redundancies are low, even when using a random selection. Note that this training stage is the only instance in our proposed approach that uses the rock type data logged on the drill hole samples.

The classifier that obtained the best results was the Simple Cart algorithm, with an error rate less than 18% (Table 5.3). This classifier is a decision tree algorithm; it is a logical choice given the associations between rock types and quantitative variables indicated in Table 5.1. Note that an error rate of 0% is not desirable since the logged rock type is a qualitative property obtained from geological core logging and is not error-free (unlike the quantitative measurements, which are assumed to be accurate) (Adeli and Emery, 2017). In addition, the incorrectly classified rock type data can be identified as the most likely to be mis-logged and be the priority candidate for checking (relogged) to ensure data consistency.

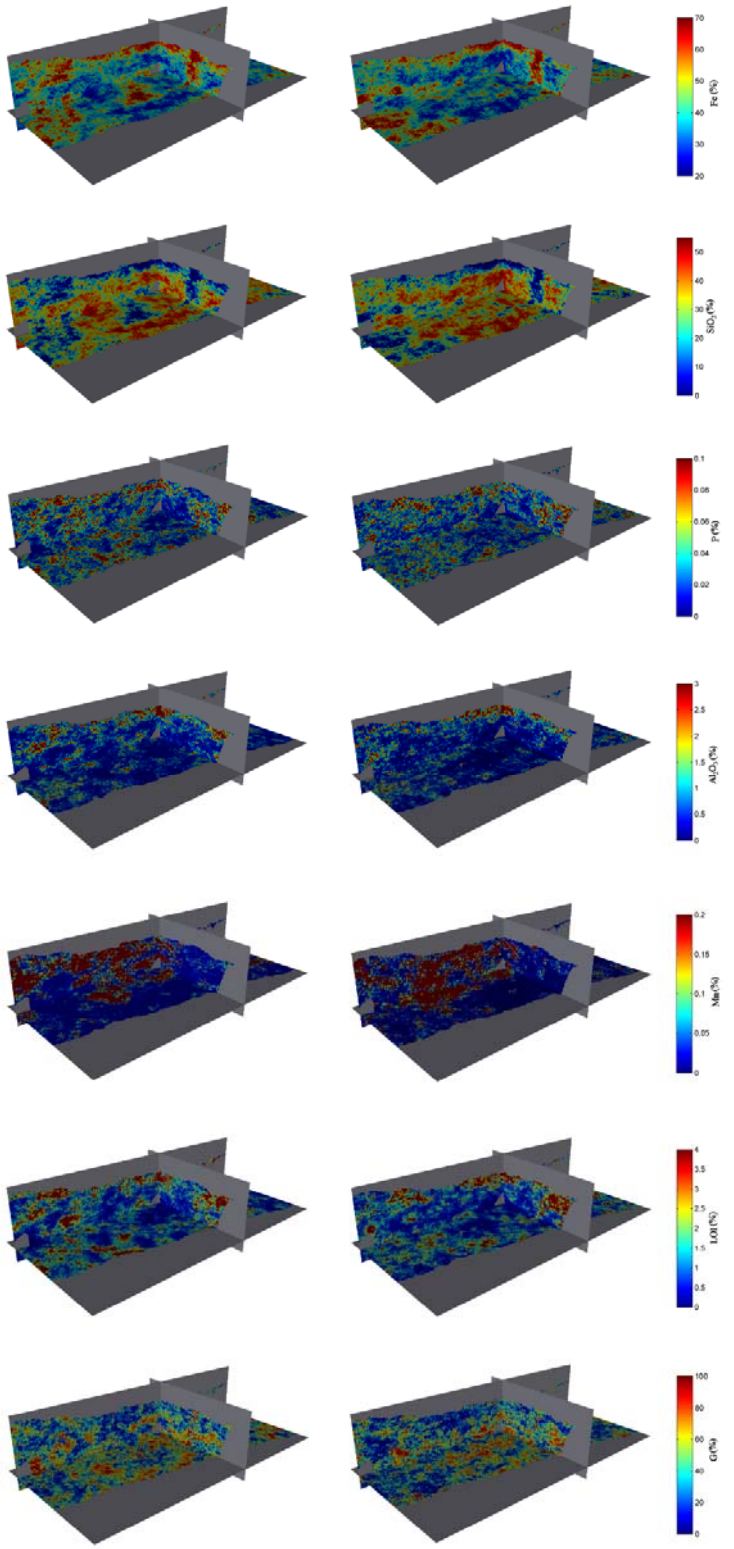


Figure 5.4: Isometric view of two realisations of the grades and granulometry (left: realisation 1 and right: realisation 2), showing the plan view and vertical cross-sections passing through the origin (local coordinate system). Waste and air are shown in grey. From top to bottom: iron grade, silica grade, phosphorus grade, alumina grade, manganese grade, loss on ignition, granulometry.

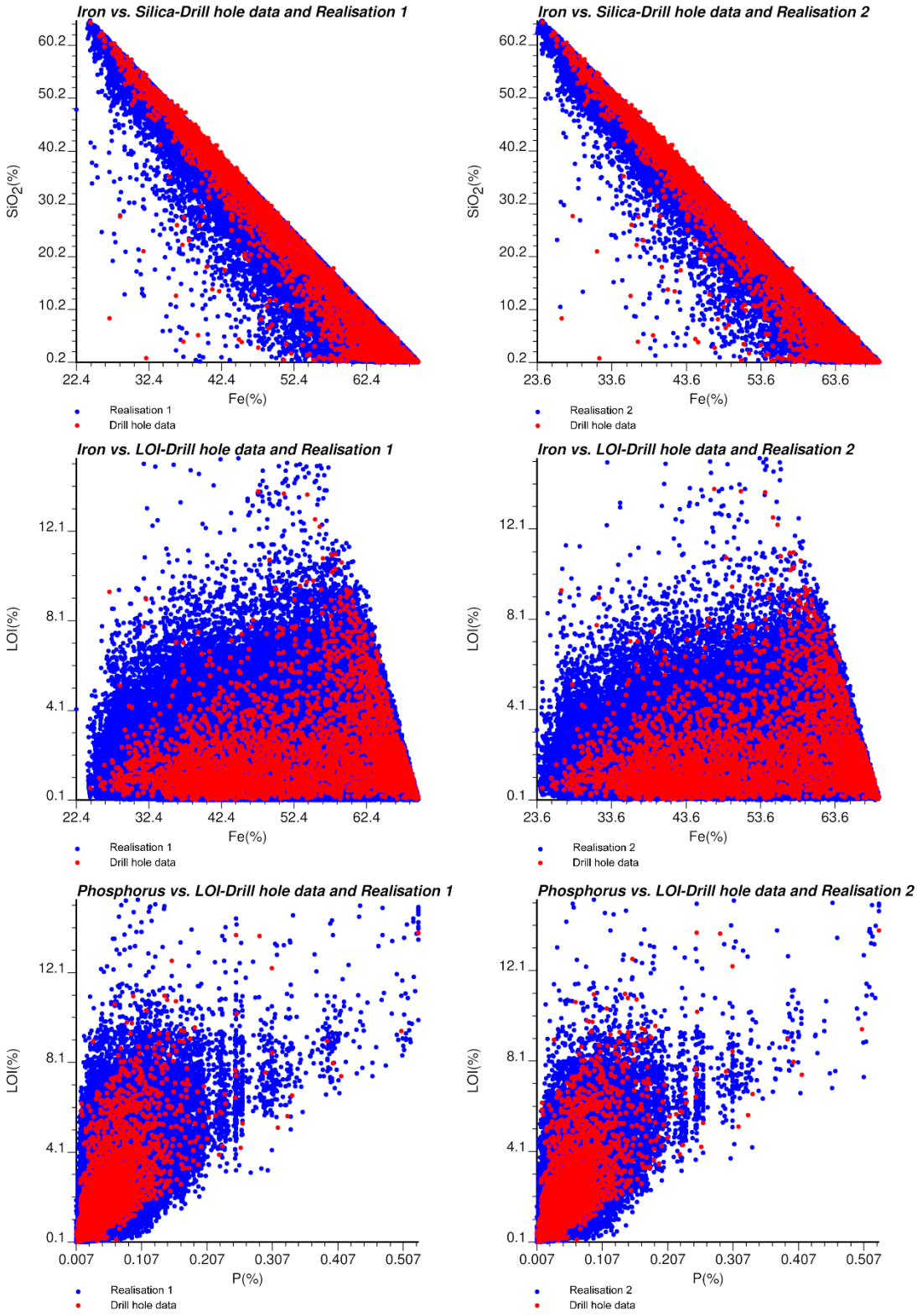


Figure 5.5: Scatter diagrams of Fe vs. SiO2, Fe vs. LOI, and LOI vs. P for drill hole data (red) and simulated values (blue). Left: realisation 1, right: realisation 2.

Table 5.3: Classifiers tested for the case study with their rates of correct classification on drill hole data.

Classification Algorithm	Algorithm Type	Correct Classification Rate (Cross-Validation)
Simple Cart	Decision tree	82.6
BF Tree	Decision tree	81.7
Classification via Regression	Meta-learning algorithm	81.7
REP Tree	Decision tree	81.6
Random Forest	Decision tree	81.1
Multilayer Perceptron (Neural Network)	Function	80.6
Bayes Network	Bayesian	77.5
RBF Network	Function	75.1
Naive Bayes	Bayesian	73.3
Random Tree	Decision tree	70.6

The classification applied to the 20 conditional realisations resulted in a number of occurrences of each rock type (from 0 to 20) for each target grid block. Figure 5.6 shows the classified rock type for two realisations.

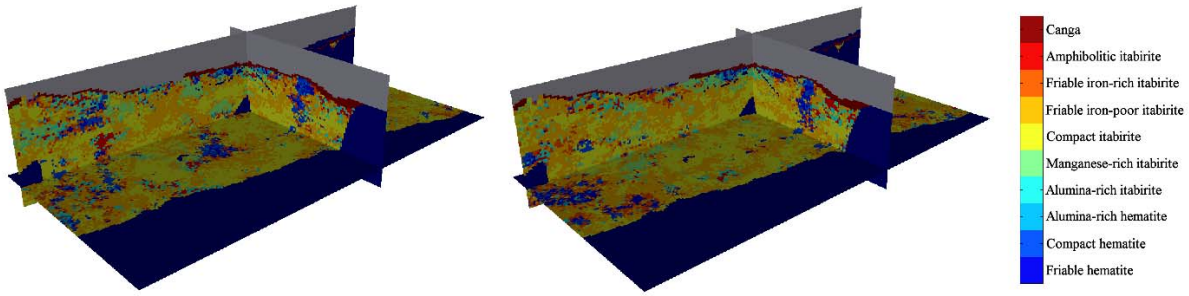


Figure 5.6: Isometric view of two realisations of the classified rock type (left: realisation 1 and right: realisation 2), showing the plan view and vertical cross-sections passing through the origin (local coordinate system). Waste and air are shown in dark blue and grey, respectively.

5.2.4. Determining the prior probability of occurrence of each rock type

It is of interest to determine whether the conditioning drill hole information influences the classification results for each target block. To do so requires the prior probabilities of all rock types to be determined by classifying the grades and the granulometry simulated in the absence of the conditioning drill hole data. In detail, the simulation process is then repeated,

this time without any conditioning data, to produce 1000 realisations of the grades and granulometry. These non-conditional realisations provide the prior probabilities (p_1, …, p_9) of the different rock types by counting the numbers of occurrences of each rock type across the 1,000 realisations (Table 5.4). This only requires one block to be simulated because the prior probabilities are the same for all the blocks in the deposit.

Table 5.4: Prior probability of each rock type.

Symbol	Rock Type Code	Prior Probability
p_1	1	0.023
p_2	2	0.051
p_3	3	0.003
p_4	4	0.048
p_5	5	0.025
p_6	6	0.345
p_7	7	0.384
p_8	8	0.090
p_9	9	0.031

5.2.5. Comparing the prior and posterior probabilities of rock type occurrences to identify potentially misinterpreted blocks

Knowing the prior probability of occurrence of each rock type, the prior distribution of the rock types that should be observed on a limited set of independent realisations can be modelled as a multinomial distribution (Papoulis, 1984):

$$\forall (n_1,...,n_9) \in \mathrm{N}^9 : n_1 + ... + n_9 = n,\ \ f(n_1,...,n_9) = \frac{n!}{n_1!... \ n_9!} p_1^{n_1} \times ... \times p_9^{n_9} \qquad (5.3)$$

In particular, this distribution gives the probability of any particular combination of the numbers of occurrences for the nine rock types among twenty (n = 20) realisations in the absence of effects of conditioning drill hole data.

If, for a given block, the numbers of rock type occurrences that are observed among the twenty conditional realisations constitute an unlikely combination of the prior multinomial distribution, then it can be concluded that the drill hole data has a significant effect on that block, i.e., there is statistical evidence that the drill hole data convey information about the rock type for this particular block. If, in addition, the rock type interpreted by the resource

geologists has a low frequency of occurrence (posterior probability) among the twenty conditional realisations, then the block can be identified as potentially misinterpreted.

The statements in the previous paragraph require a quantitative definition of "unlikely" or "low". To this end, the combinations of the prior multinomial distribution are ranked, from the least probable to the most probable, and the less probable combinations (up to a cumulative probability of 0.1) are classified as "unlikely" or "improbable". One then looks for the blocks in which such improbable combinations arose in the twenty conditional realisations, i.e., the combination is very unlikely in the non-conditional case (absence of drill hole data) but has occurred in the conditional case, showing that there is a significant effect of the conditioning drill hole data on the blocks. Note that there is no particular reason to find 10% of the blocks in the geological model with the above-specified cumulative probability (0.1): more blocks may exhibit an "unlikely" combination if the conditioning data have a strong effect (long-range correlation structure), while fewer blocks (possibly none) may be identified if the conditioning data have low spatial correlation.

Finally, the following three criteria are used to identify potentially misinterpreted blocks among the blocks that are significantly affected by the drill hole data: (i) the prior probability of the rock type interpreted by the geologists is greater than its posterior probability; (ii) the posterior probability of the rock type interpreted by the geologists is less than 0.15 (unlikely); and (iii) another rock type has a posterior probability higher than the posterior probability of the rock type interpreted by the geologists and has been logged at a drill hole sample less than 60 m away from the block. This last criterion is adopted to avoid extrapolating the drill hole information too much, bearing in mind that the geostatistical model is likely to be valid only at a local scale (quasi-stationarity assumption) and considering a distance lower than the spatial correlation range, for which the direct variograms reach about 70–80% of their sills (Figure 5.3). The particular values (0.15 probability and 60 m distance) chosen in criteria (ii) and (iii) can nevertheless be tuned by the user depending on his/her preferences, intuition, and expertise (or be modified in other case studies, depending on the observed correlation range of the quantitative data), which reflects more or less conservative detections of the misinterpreted rock types in the geological model.

5.3. Results and discussion

The application of the criteria in the previous section identifies 3.39% of all the blocks (20,999 blocks out of 619,919 blocks flagged with rock types 1–9 in the original geological model) to be potentially misinterpreted (condition +5), as shown in Figure 5.7. Conditions +1 to +5 are described in Table 5.5.

Except for a few blocks scattered across the deposit, the identified misinterpreted blocks (condition +5) are concentrated in the upper part of the ore deposit, in the margins of the

manganese-rich itabirite, alumina-rich itabirite, and hematite bodies, showing the necessity to check these blocks in order more accurately to separate these bodies located in the transitional parts of the deposit from high to low manganese, alumina, and iron grades.

Table 5.5: Defined conditions.

Symbol	Condition
+1	Block under consideration is not affected significantly by the drill hole data
+2	Prior probability of the rock type interpreted by the geologists is lower than its posterior probability (the interpreted rock type "agrees" with the posterior distribution)
+3	Block under consideration does not meet condition +2, but there is not any evidence for a misinterpretation (neither +4 nor +5)
+4	Prior probability of the rock type interpreted by the geologists is greater than its posterior probability, posterior probability of the interpreted rock type is less than 0.15 (unlikely), and another rock type has a higher posterior probability
+5	In addition to the criteria of condition +4, a rock type with higher posterior probability has been logged at some drill hole sample less than 60 m from the block

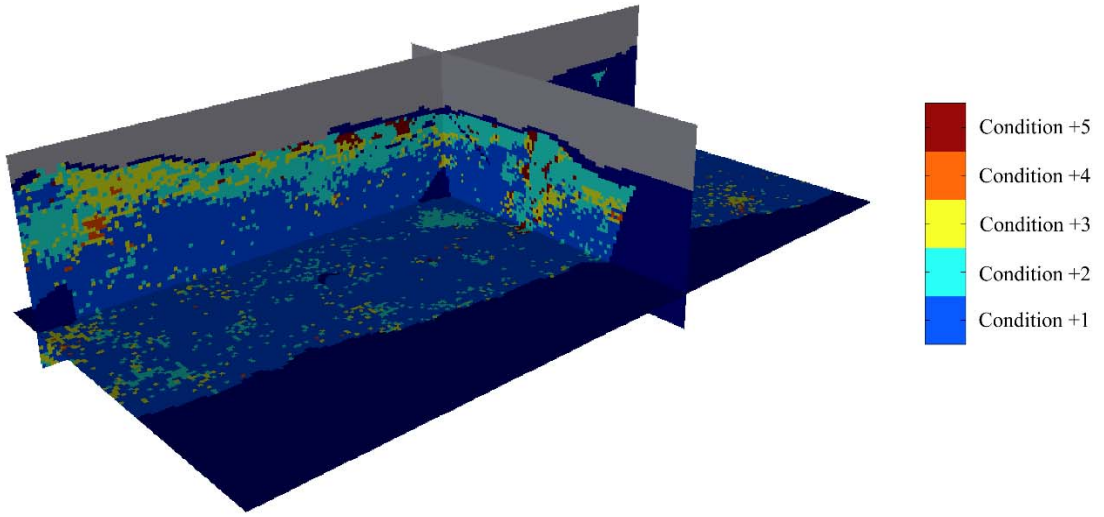


Figure 5.7: Isometric view of block classification according to criteria in Table 5.5, showing the plan view and vertical cross-sections passing through the origin (local coordinate system). Waste and air are shown in dark blue and grey, respectively.

The numbers of correctly interpreted and potentially misinterpreted blocks for each interpreted and suggested rock type are shown in Table 5.6. For example, 35 blocks are interpreted as rock type 1 (friable hematite) with the suggestion they be changed to rock type 2 (compact hematite).

Table 5.6: Numbers of blocks with no evidence of misinterpretation (conditions +1 to +4) (diagonal line) and numbers of potentially misinterpreted blocks (condition +5) (off-diagonal) for each interpreted rock type (row) and each suggested rock type based on classification of 20 realisations (column).

Rock Type	1	2	3	4	5	6	7	8	9
1	12,058	35	51	22	35	21	76	61	29
2	198	7712	45	14	8	41	239	223	27
3	70	61	5289	26	22	5	41	55	14
4	93	39	99	15,917	22	6	99	78	70
5	96	15	38	55	5788	3	85	89	4
6	535	216	73	410	165	349,723	2607	415	619
7	2292	990	1169	1458	776	1161	172,299	1623	2072
8	369	172	336	149	110	72	403	18,259	233
9	50	48	32	36	15	8	46	29	11,875

In the following, three of the potentially misinterpreted blocks are selected and discussed in light of the simulated values of grades and granulometry. For the first selected block (referred to as "block n°1"), with easting coordinate 340 m, northing coordinate 380 m, and elevation −40 m in the local coordinate system, the mining geologist interpretation corresponds to rock type 6 (compact itabirite). However, the simulated values of granulometry (mostly less than 50%) suggest that this block should actually be interpreted as rock type 7 (friable iron-poor itabirite) in 17 out of the 20 realisations (Table 5.7). This rock type code (7) is also the one logged in the three closest drill hole samples, all of which are less than 20 m from block 1. A similar situation occurs for block n°2 with easting coordinate −580 m, northing coordinate 330 m, and elevation 280 m, interpreted as rock type 7 (friable iron-poor itabirite), which is re-interpreted as rock type 1 (friable hematite) in 17 out of the 20 realisations, based on the high simulated iron grades (Table 5.8); the four nearest drill hole samples, less than 35 m from the block, have been logged as rock type 1, which corroborates the suggestion for a re-interpretation of the block. For block n°3 with easting coordinate 1390 m, northing coordinate 410 m, and elevation 390 m, 17 of the 20 realisations suggest classification as rock type 3 (alumina-rich hematite), based on the high simulated iron and alumina grades (Table 5.9), whereas the original geological interpretation corresponds to rock type 8 (friable iron-rich itabirite). This interpretation (rock type 8) coincides with the log of the closest sample, less than 30 m from the block, but examination of the grades of this sample suggests that it may actually be mis-logged: following the methodology proposed in (Adeli and Emery, 2017), the *p*-values of rock types 3 and 8 are 0.94 and 0.51, respectively, indicating that the former code is more plausible than the latter.

Table 5.7: Simulated grades and granulometry, and associated rock type, for 20 realisations of block n°1, interpreted as rock type 6 (compact itabirite) by mining geologists.

Realisation	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G	Classified Rock Type
1	47.22	30.21	0.019	1.411	0.026	0.798	55.92	6
2	40.56	40.97	0.017	0.418	0.011	0.566	20.90	7
3	35.79	47.56	0.017	0.678	0.010	0.533	16.69	7
4	39.74	41.97	0.022	0.540	0.010	0.613	40.73	7
5	36.71	46.70	0.010	0.416	0.010	0.368	23.99	7
6	37.02	46.01	0.016	0.451	0.010	0.573	11.24	7
7	41.50	38.96	0.017	0.536	0.020	1.113	32.39	7
8	59.11	14.72	0.010	0.384	0.010	0.343	10.60	8
9	52.29	20.23	0.072	1.703	0.029	3.113	31.50	8
10	38.68	44.03	0.012	0.330	0.010	0.301	49.71	7
11	37.93	43.16	0.036	0.570	0.010	1.950	43.22	7
12	37.42	45.59	0.010	0.311	0.010	0.570	37.23	7
13	30.67	55.29	0.014	0.320	0.010	0.502	14.52	7
14	32.97	52.01	0.011	0.254	0.010	0.563	42.12	7
15	30.54	54.77	0.012	0.702	0.010	0.814	34.56	7
16	44.76	33.78	0.017	1.146	0.027	1.013	16.65	7
17	45.96	32.21	0.017	0.988	0.014	1.032	33.37	7
18	40.43	41.54	0.013	0.303	0.010	0.314	45.39	7
19	47.19	30.01	0.032	1.164	0.018	1.263	41.08	7
20	42.17	37.34	0.048	0.901	0.015	1.343	44.33	7

Table 5.8: Simulated grades and granulometry, and associated rock type, for 20 realisations of block n°2, interpreted as rock type 7 (friable iron-poor itabirite) by mining geologists.

Realisation	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G	Classified Rock Type
1	67.59	1.07	0.013	1.193	0.010	1.054	26.67	1
2	66.88	2.50	0.016	1.240	0.010	0.584	46.37	1
3	66.17	3.46	0.015	1.275	0.010	0.613	48.08	1
4	64.99	5.14	0.014	1.317	0.010	0.576	3.89	1
5	67.46	1.68	0.018	1.020	0.010	0.805	19.45	1
6	68.66	1.24	0.010	0.346	0.010	0.216	57.84	2
7	62.30	7.40	0.015	2.319	0.010	1.161	10.35	1
8	66.55	2.07	0.020	1.663	0.010	1.061	25.53	1
9	67.89	2.27	0.010	0.361	0.010	0.272	2.92	1
10	68.74	0.83	0.015	0.388	0.010	0.448	31.09	1
11	62.57	9.15	0.012	0.851	0.010	0.503	10.12	1
12	61.69	9.59	0.024	0.970	0.010	1.168	18.37	8
13	65.41	3.59	0.016	1.816	0.010	1.021	35.01	1
14	67.70	2.10	0.012	0.679	0.010	0.389	10.40	1
15	67.63	0.65	0.028	1.240	0.011	1.339	37.19	1
16	67.66	1.84	0.021	0.704	0.010	0.659	14.63	1
17	59.19	14.67	0.010	0.397	0.010	0.272	1.93	8
18	62.59	8.48	0.012	1.255	0.010	0.743	22.11	1
19	66.79	1.55	0.032	1.533	0.016	1.338	31.18	1
20	68.48	0.71	0.016	0.581	0.010	0.748	32.40	1

Table 5.9: Simulated grades and granulometry, and associated rock type, for 20 realisations of block n°3, interpreted as rock type 8 (friable iron-rich itabirite) by mining geologists.

Realisation	Fe	SiO_2	P	Al_2O_3	Mn	LOI	G	Classified Rock Type
1	63.608	0.80	0.052	0.873	0.410	6.736	9.90	3
2	62.235	2.79	0.076	3.258	0.021	4.770	14.71	3
3	62.653	1.17	0.042	1.497	0.053	7.595	4.63	3
4	64.02	2.96	0.037	2.453	0.064	2.884	1.80	1
5	64.817	0.80	0.095	1.886	0.095	4.310	20.67	3
6	62.642	2.64	0.105	3.058	0.083	4.397	14.68	3
7	63.305	1.23	0.047	2.737	0.073	5.326	34.15	3
8	62.068	1.11	0.126	3.221	0.034	6.600	21.30	3
9	65.543	0.39	0.069	1.635	0.031	4.065	23.33	3
10	62.09	0.97	0.070	2.925	0.095	7.048	44.50	3
11	63.943	1.32	0.037	1.387	0.024	5.758	32.72	3
12	63.108	1.14	0.039	2.357	0.039	6.141	7.71	3
13	59.538	2.04	0.073	6.104	0.080	6.466	8.94	9
14	62.418	0.52	0.099	3.669	0.075	6.244	5.13	3
15	62.12	0.41	0.096	3.155	0.033	7.361	6.21	3
16	63.739	0.52	0.040	1.420	0.148	6.644	23.91	3
17	57.725	0.49	0.095	5.816	3.110	6.932	11.19	5
18	64.591	1.03	0.067	1.721	0.033	4.707	17.59	3
19	62.706	2.49	0.079	2.092	0.018	5.559	21.10	3
20	64.873	0.57	0.043	1.131	0.025	5.423	40.42	3

On a final note, the proposed methodology used for validating the interpreted geological model, based on the calculation of prior and posterior probabilities, and on the definition of heuristic criteria (Sections 5.2.4 and 5.2.5), can be applied not only to the geological scenarios obtained from the simulation and classification of quantitative covariates, as set out in Sections 5.2.2 and 5.2.3, but also to scenarios obtained from any other geostatistical simulation method, e.g., multiple-point, truncated Gaussian or plurigaussian simulation (Xu et al., 2006; Armstrong et al., 2011; Mariethoz and Caers, 2014; Beucher and Renard, 2016).

5.4. Conclusions

The rock type or ore type model interpreted by mining geologists is the basis for mineral resource evaluation, mine planning, and subsequent stages of the mining process. This motivates the design of a method for validating the interpreted category assigned to each grid block and for finding the blocks for which the interpretation is likely to be incorrect. To this end, a geostatistical-based approach has been proposed for constructing a set of simulated geological scenarios and for identifying potentially misinterpreted blocks, assuming that

there is a clear association between geological categories and measured quantitative covariates.

The applicability of the proposal was tested on an iron ore deposit in which there is a clear association between the interpreted rock types and seven quantitative covariates (grades of iron, silica, phosphorus, alumina, manganese, loss on ignition, and granulometry). The proposal combines a change of variables based on a stoichiometric closure formula, PPMT transformation, variogram analysis, turning-bands simulation, post-conditioning cokriging, and decision-tree classification. The potentially misinterpreted blocks are then identified by comparing the prior and posterior rock type probabilities, and by defining heuristic criteria that can be tuned by the user to achieve more or less conservative detections.

The proposed approach can be applied not only in the context of geological modelling, but also in the wider context of geometallurgical modelling, in which there are relatively large volumes of multivariate data of different natures and qualities (e.g., grades, grain sizes, mineralogy, alteration, grindability indices, and metal recoveries), and where the correct identification and interpretation of geometallurgical domains is critical to improving process performance.

Acknowledgments: The first two authors acknowledge funding from the Chilean Commission for Scientific and Technological Research through Project CONICYT/FONDECYT/REGULAR/N°1170101. The authors are grateful to two anonymous reviewers for their constructive comments on a previous version of this work.

Author contributions: Amir Adeli, Xavier Emery, and Peter Dowd conceived and designed the experiments. Amir Adeli performed the experiments. Amir Adeli and Xavier Emery wrote the paper.

Conflicts of interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

5.5. Bibliography

Adeli, A., Emery, X., 2017. A geostatistical approach to measure the consistency between geological logs and quantitative covariates. Ore Geol. Rev. 82, 160–169.

Armstrong, M., Galli, A., Beucher, H., Le Loc'h, G., Renard, D., Doligez, B., Eschard, R., Geffroy, F., 2011. Plurigaussian Simulations in Geosciences. Springer: Berlin/Heidelberg, Germany, p. 176.

Barnett, R.M., Manchuk, J.G., Deutsch, C.V., 2014. Projection pursuit multivariate transform. Math. Geosci. 46, 337–359.

Beucher, H., Renard, D., 2016. Truncated Gaussian and derived methods. C. R. Geosci. 348, 510–519.

Chilès, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 699 pp.

Dowd, P.A., 1994. Geological controls in the geostatistical simulation of hydrocarbon reservoirs. Arab. J. Sci. Eng. 19, 237–247.

Dowd, P.A., 1997. Structural controls in the geostatistical simulation of mineral deposits. In Geostatistics Wollongong'96, Baafi, E.Y., Schofield, N.A., Eds., Kluwer Academic: Dordrecht, The Netherlands, pp. 647–657.

Duke, J.H., Hanna, P.J., 2001. Geological interpretation for resource modelling and estimation. In Mineral Resource and Ore Reserve Estimation—The AusIMM Guide to Good Practice, Edwards, A.C., Ed., Australasian Institute of Mining and Metallurgy: Melbourne, Australia, pp. 147–156.

Emery, X., 2010a. Iterative algorithms for fitting a linear model of coregionalization. Computers & Geosciences 36(9): 1150-1160.

Emery, X., 2010b. Multi-Gaussian kriging and simulation in the presence of an uncertain mean value. Stoch. Environ. Res. Risk Assess. 24, 211–219.

Emery, X., Arroyo, D., Porcu, E., 2016. An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. Stoch. Environ. Res. Risk Assess. 30, 1863–1873.

Emery, X., Robles, L.N., 2009. Simulation of mineral grades with hard and soft conditioning data: Application to a porphyry copper deposit. Comput. Geosci. 13, 79–89.

Emery, X., Silva, D.A., 2009. Conditional co-simulation of continuous and categorical variables for geostatistical applications. Comput. Geosci. 35, 1234–1246.

Friedman, J.H., 1987. Exploratory projection pursuit. J. Am. Stat. Assoc. 82, 249–266.

Friedman, J.H., Tukey, J.W., 1974. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Comput. C-23, 881–890.

Goulard, M., Voltz, M., 1992. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. 24, 269–286.

Guillen, A., Calcagno, P., Courrioux, G., Joly, A., Ledru, P., 2008. Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion. Phys. Earth Planet. Inter. 171, 158–169.

Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? Math. Geol. 21, 715–739.

Knödel, K., Lange, G., Voigt, H.J., 2007. Environmental Geology: Handbook of Field Methods and Case Studies. Springer, Berlin, 1357 pp.

Larrondo, P., Leuangthong, O., Deutsch, C.V., 2004. Grade estimation in multiple rock types using a linear model of coregionalization for soft boundaries. In Proceedings of the 1st International Conference on Mining Innovation; Magri, E., Ortiz, J., Knights, P., Henríquez, F., Vera, M., Barahona, C., Eds.; Gecamin Ltd.: Santiago, Chile, pp. 187–196.

Lelièvre, P.G., 2009. Integrating Geologic and Geophysical Data through Advanced Constrained Inversions. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada.

Leuangthong, O., Deutsch, C.V., 2003. Stepwise conditional transformation for simulation of multiple variables. Math. Geol. 35, 155–173.

Leuangthong, O., McLennan, J.A., Deutsch, C.V., 2004. Minimum acceptance criteria for geostatistical realizations. Nat. Resour. Res. 13, 131–141.

Maleki, M., Emery, X., 2015. Joint simulation of grade and rock type in a stratabound copper deposit. Mathematical Geosciences 47: 471-495.

Maleki, M., Emery, X., 2017. Joint simulation of stationary grade and non-stationary rock type for quantifying geological uncertainty in a copper deposit. Comput. Geosci. 109, 258–267.

Maleki, M., Emery, X., Cáceres, A., Ribeiro, D., Cunha, E., 2016. Quantifying the uncertainty in the spatial layout of rock type domains in an iron ore deposit. Comput. Geosci. 20, 1013–1028.

Mariethoz, G., Caers, J., 2014. Multiple-Point Geostatistics: Stochastic Modeling with Training Images. Wiley: New York, NY, USA, p. 376.

Marjoribanks, R., 2010. Geological Methods in Mineral Exploration and Mining. Springer, Berlin, 238 pp.

Mery, N., Emery, X., Cáceres, A., Ribeiro, D., Cunha, E., 2017. Geostatistical modeling of the geological uncertainty in an iron ore deposit. Ore Geol. Rev. 88, 336–351.

Ortiz, J.M., Emery, X., 2006. Geostatistical estimation of mineral resources with soft geological boundaries: A comparative study. J. S. Afr. Inst. Min. Metall, 106, 577–584.

Papoulis, A., 1984. Probability, Random Variables, and Stochastic Processes. McGraw-Hill: New York, NY, USA.

Pawlowsky-Glahn, V., Buccianti, A., 2011. (Eds.) Compositional Data Analysis: Theory and Applications. Wiley: New York, NY, USA, p. 400.

Rossi, M.E., Deutsch, C.V., 2014. Mineral Resource Estimation. Springer, Heidelberg, 332 pp.

Séguret, S.A., 2013. Analysis and estimation of multi-unit deposits: Application to a porphyry copper deposit. Math. Geosci. 45, 927–947.

Silva, D.S.F., Deutsch, C.V., 2016. Multivariate data imputation using Gaussian mixture models. Spat. Stat. doi:10.1016/j.spasta.2016.11.002.

Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation. Cambridge University Press, Cambridge, 400 pp.

Van den Boogaart, K.G., Mueller, U., Tolosana-Delgado, R., 2016. An affine equivariant multivariate normal score transform for compositional data. Math. Geosci. 49, 231–251.

Vargas-Guzmán, J.A., 2008. Transitive geostatistics for stepwise modeling across boundaries between rock regions. Math. Geosci. 40, 861–873.

Wackernagel, H., 2003. Multivariate Geostatistics: an Introduction with Applications. Springer, Berlin, 387 pp.

Witten, I.H., Frank, E., Hall, M., Pal, C., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Elsevier: Burlington, MA, USA, p. 654.

Xu, C., Dowd, P.A., Mardia, K.V., Fowell, R.J., 2006. A flexible true plurigaussian code for spatial facies simulations. Comput. Geosci. 32, 1629–1645.

Chapter 6: Discussion, conclusions and perspectives

6.1. General discussion

There is an increasing need in the mining industry for high performance of mineral resources and ore reserves models. A portion of the model deviations are due to mislogged samples, for which the logged value of a petrophysical attribute such as the lithology, alteration or mineralogical assemblage is erroneous. Considering the regionalized nature of petrophysical attributes and their dependence relationships with quantitative variables from geochemical analyses or metallurgical tests, a geostatistical approach based on a transitional boundary model and on leave-one-out cross-validation has been proposed for identifying possible mislogged samples. The proposal aims to calculate, for each sample, a measure of consistency between the logged classes and the quantitative covariates, allowing the practitioner to include additional criteria (*p*-value limits and geographical criteria) to detect suspicious samples. Also, because some samples may have all their measures of consistency smaller than the chosen *p*-value limit and may therefore not be classified into any of the logged classes, the proposed methodology cannot be used blindly as a substitute for the original logs.

Another source of deviations of mineral resources and ore reserves models comes from the misinterpretation of the spatial layout of lithological, alteration or mineralogical domains corresponding to rock types or ore types. Accordingly, there is a need for validating the interpreted category assigned to each area or block of the deposit and for finding the blocks for which the geological interpretation is likely to be incorrect. To this end, a geostatistical approach based on a soft boundary model has been proposed for constructing a set of simulated geological scenarios and for identifying potentially misinterpreted blocks.

Each of the two models on which the proposed methodologies are based has been designed for a specific purpose. In the transitional boundary model (model 2 defined in Chapter 3), the quantitative variables are "subordinated" to the geological categories defined by the logged classes, whereas the reverse happens for the soft boundary model (model 3 defined in Chapter 3), where the quantitative variables are defined prior to geological domaining. Despite this difference in their assumptions, both models rely on the idea that there is a strong association between the geological categories and the measured quantitative covariates. Also, both of them allow a spatial correlation of the quantitative variables across the geological boundaries, which is particularly suitable for the modeling of disseminated deposits where there are no clear-cut discontinuities in the values of quantitative variables obtained from geochemistry or from metallurgical tests when crossing a geological boundary.

In contrast, with a hard boundary model (model 1 of Chapter 3), the quantitative covariates would be useless for validating the geological interpretation of the deposit, insofar as these variables would not bring any information on the geological domain present at an unsampled location.

6.2. Conclusions

Different geostatistical models can be elaborated to describe the dependence relationships between qualitative regionalized properties (geological categories defined from geological core logging or from geological interpretations) and quantitative covariates available from geochemical analyses or metallurgical tests. In the case of disseminated deposits for which there are no clear-cut discontinuities of the quantitative variables when crossing a geological boundary, a transitional boundary model or a soft boundary model can be defined.

The first (transitional boundary) model has been used to design a methodology for validating and reclassifying geological logs, based on leave-one-out cross-validation and the definition of a consistency measure between the logs and the quantitative covariates. In turn, the second (soft boundary) model has been used to design a methodology for validating geological interpretations, based on simulation and decision-tree classification. Both methodologies use data analysis and multivariate geostatistical tools, such as joint Gaussian transformation, coregionalization modeling, cokriging and joint Gaussian simulation. In order to illustrate their applicability, a case study from an iron ore deposit has been presented, where ten logged rock types are strongly correlated with seven quantitative variables (grades of iron, silica, phosphorus, alumina, manganese, loss on ignition and granulometry) measured on the same set of exploration drill holes. An important practical aspect of both proposals is the possibility to incorporate heuristic criteria that can be tuned by the user in order to achieve more or less conservative detections.

6.3. Perspectives

The methodological proposals presented in this thesis may be the basis for future works, in particular:

1) The proposed approaches can be applied not only in the context of geological modeling, but also in the context of geometallurgical modeling, by using ore mineralogical data from QEMSCAN analyses for finding mineral zones, or gangue mineralogical data derived from spectral information for recognizing alterations. In geometallurgical studies, there are relatively large volumes of multivariate data of different natures and qualities (e.g., grades, grain sizes, mineralogy, alteration, grindability indices, and metal recoveries), and the correct identification and interpretation of geometallurgical domains is critical for improving the overall performance of the value chain in the mining business.

2) One can enhance the soft boundary model (model 3 suggested in Chapter 3) by focusing on the scales of variations of main interest for geological domaining. This enhancement could be based on the decomposition of the random fields representing the quantitative variables into components acting at different spatial scales via coregionalization analysis. Indeed, instead of integrally simulating these variables, it could be interesting to filter out the noisy or short-scale components (associated with the nugget effect or with the short-range structures of the coregionalization model, which can relate to measurement errors), in order to obtain a simulation of the large-scale structures (nested structures with a large range of correlation). This model enhancement is the topic of ongoing research.

Chapter 7: Bibliography

[1] Adeli, A., Emery, X., 2017. A geostatistical approach to measure the consistency between geological logs and quantitative covariates. Ore Geol. Rev. 82, 160–169.

[2] Agterberg, F.P., 1990. Automated stratigraphic correlation. Elsevier, Amsterdam, 423 pp.

[3] Alabert, F.G., 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Mathematical Geology 19(5), 369-386.

[4] Armstrong, M., Galli, A., Beucher, H., Le Loc'h, G., Renard, D., Doligez, B., Eschard, R., Geffroy, F., 2011. Plurigaussian Simulations in Geosciences. Springer: Berlin/Heidelberg, Germany, p. 176.

[5] Arroyo, D., Emery, X., Peláez, M., 2012. An enhanced Gibbs sampler algorithm for non- conditional simulation of Gaussian random vectors. Computers & Geosciences 46, 138-148.

[6] Barnett, R.M., Manchuk, J.G., Deutsch, C.V., 2014. Projection pursuit multivariate transform. Math. Geosci. 46, 337–359.

[7] Beucher, H., Renard, D., 2016. Truncated Gaussian and derived methods. C. R. Geosci. 348, 510–519.

[8] Bourgine, B., Lasseur, E., Leynet, A., Badinier, G., Ortega, C., Issautier, B., Bouchet, V., 2015. Building a geological reference platform using sequence stratigraphy combined with geostatistical tools. Geophysical Research Abstracts 17 EGU2015-8292

[9] Bourgine, B., Prunier-Leparmentier, A.M., Lembezat, C., Thierry, P., Luquet, C., Robelin, C., 2008. Tools and methods for constructing 3D geological models in the urban environment. The Paris case. In: Ortiz, J.M., Emery, X. (eds.) Proceedings of the Eighth International Geostatistics Congress. Gecamin Ltda, Santiago, pp. 951-960

[10] Cáceres, A., Emery, X., 2013. Geostatistical validation of geological logging. In: Ambrus, J., Beniscelli, J., Brunner, F., Cabello, J., Ibarra, F. (eds.) Proceedings of the Third International Seminar on Geology for the Mining Industry. Gecamin Ltda, Santiago, pp. 73-80

[11] Carrasco, P., Ibarra, F., Rojas, R., Le Loc'h, G., Seguret, S., 2007. Application of the Truncated Gaussian Simulation Method to a Porphyry Copper Deposit. In: Magri, Ed., APCOM 2007, 33rd International Symposium on Application of Computers and Operations research in the Mineral industry, pp. 31-39.

[12] Carrasco, P.C., 2010. Nugget effect, artificial or natural?. The journal of the Southern African institute of mining and metallurgy 110(6), 299-305.

[13] Casella, G., George, E.I., 1992. Explaining the Gibbs sampler. American Statistician 46(3), 167-174.

[14] Cherubini, C., Giasi, C.I., Musci, F., Pastore, N., 2009. Checking simulations of a geolithological model obtained by means of nested truncated bigaussian method. In: International journal of mathematical model and methods in applied sciences 3(2), 152-161.

[15] Chilès, J.P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. Wiley, New York, 699 pp.

[16] Contreras, R., Ortiz, J., Bisso, C., 2010. Mine planning considering uncertainty in grades and work index. Proceedings of the 4th International Conference on mining innovation. Santiago, pp. 129-136.

[17] Cressie, N., 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17(5), 563-586.

[18] Cressie, N., 1993. Statistics for spatial data, revised ed. Wiley, New York.

[19] David, M., 1977. Geostatistical ore reserve estimation, Elsevier, Amsterdam, The Netherlands, 364 pp.

[20] Davis, M., 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Mathematical Geology 19(2), 91-98.

[21] De Beer, J.E., Hammond, A.J., Robertson, A.M., van Schalkwyk, A. and Weaver, J.M., 1976. A guide to core logging for rock engineering. In: Proceedings of the

Symposium on Exploration for Rock Engineering, Johannesburg, Nov. 1976, pp. 71-86.

[22] Deraisme, J., Field, M., 2006. Geostatistical simulations of kimberlite orebodies: application to sampling optimisation. Proceedings of the 6th international mining geology conference, Darwin, NT, Australia, 21-23 Aug 2006, pp. 193-203.

[23] Desassis, N., Renard, D., 2011. Automatic variogram modeling by iterative least squares—Univariate and multivariate cases. Technical Report, Center of Geosciences and Geoengineering, MINES ParisTech, Fontainebleau, France.

[24] Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York, 369 pp.

[25] Deutsch, J.L., Palmer, K., Deutsch, C.V., Szymanski, J., Etsell, T.H., 2016. Spatial modeling of geometallurgical properties: techniques and a case study. Natural Resources Research 25: 161-181.

[26] Dimitrakopoulos, R., 1997. Conditional simulations: tools for modelling uncertainty in open pit optimisation. Optimizing with Whittle. Whittle Programming Pty Ltd, Perth, pp. 31-42.

[27] Dimitrakopoulos, R., 2009. Stochastic mine planning – methods, example and value in an uncertain word. In: Proceedings of orebody modelling and strategic mine planning, AusIMM, pp. 13-19.

[28] Dimitrakopoulos, R., Godoy, M., and Chou, C.L., 2009. Resource/reserve classification with integrated geometric and local grade variability measures. In: Proceedings of orebody modelling and strategic mine planning, AusIMM, pp. 207-214.

[29] Dorr, J.V.N., 1964. Supergene iron ores of Minas Gerais, Brazil. Economic Geology 59(7): 1203-1240.

[30] Dowd, P.A., 1994. Geological controls in the geostatistical simulation of hydrocarbon reservoirs. Arab. J. Sci. Eng. 19, 237–247.

[31] Dowd, P.A., 1997. Structural controls in the geostatistical simulation of mineral deposits. In Geostatistics Wollongong'96, Baafi, E.Y., Schofield, N.A., Eds., Kluwer Academic: Dordrecht, The Netherlands, pp. 647–657.

[32] Dowd, P.A., Pardo-Iguzquiza, E., Xu, C., 2003. Plurigau: a computer program for simulating spatial facies using the truncated plurigaussian method. Computers & Geosciences 29(2), 123-141.

[33] Duke, J.H., Hanna, P.J., 2001. Geological interpretation for resource modelling and estimation. In Mineral Resource and Ore Reserve Estimation—The AusIMM Guide to Good Practice, Edwards, A.C., Ed., Australasian Institute of Mining and Metallurgy: Melbourne, Australia, pp. 147–156.

[34] Dunn, K.J., Bergman, D.J., Latorraca, G.A., 2002. Nuclear Magnetic Resonance – Petrophysical and Logging Applications. Pergamon, Amsterdam.

[35] Emery, X., 2004. Testing the correctness of the sequential algorithm for simulating Gaussian random fields. Stochastic Environment Research and Risk Assessment 18(6), 401-413.

[36] Emery, X., 2005. Variograms of Order Omega: a Tool to Validate a Bivariate Distribution Model. Mathematical Geosciences 37(2), 163-181.

[37] Emery, X., 2007. Simulation of geological domains using the plurigaussian model: New developments and computer programs. Computers & Geosciences 33(9), 1189-1201.

[38] Emery, X., 2008a. A turning bands program for conditional co-simulation of cross-correlated Gaussian random fields. Computers & Geosciences 34(12), 1850-1862.

[39] Emery, X., 2008b. Uncertainty modeling and spatial prediction by multi-Gaussian kriging: Accounting for an unknown mean value. Computers & Geosciences 34(11): 1431–1442.

[40] Emery, X., 2009. The kriging update equations and their application to the selection of neighboring data. Computational Geosciences 13(3): 269-280.

[41] Emery, X., 2010a. Iterative algorithms for fitting a linear model of coregionalization. Computers & Geosciences 36(9): 1150-1160.

[42] Emery, X., 2010b. Multi-Gaussian kriging and simulation in the presence of an uncertain mean value. Stoch. Environ. Res. Risk Assess. 24, 211–219.

[43] Emery, X., 2017. Course Notes on Statistical and Geostatistical Data Analysis. University of Chile.

[44] Emery, X., Arroyo, D., Porcu, E., 2016. An improved spectral turning-bands algorithm for simulating stationary vector Gaussian random fields. Stoch. Environ. Res. Risk Assess. 30, 1863–1873.

[45] Emery, X., González, K.E., 2007a. Incorporating the uncertainty in geological boundaries into mineral resources evaluation. Journal of the Geological Society of India 69(1), 29-38.

[46] Emery, X., González, K.E., 2007b. Probabilistic modelling of mineralogical domains and its application to resources evaluation. The Journal of the Southern African Institute of Mining and Metallurgy 107(12), 803-809.

[47] Emery, X., Lantuéjoul, C., 2006. TBSIM: A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method. Computers & Geosciences 32(10), 1615-1628.

[48] Emery, X., Ortiz, J.M., Cáceres, A.M., 2008. Geostatistical modeling of rock type domains with spatially varying proportions: application to a porphyry copper deposit. The Journal of the Southern African Institute of Mining and Metallurgy 108(5), 285-292.

[49] Emery, X., Ortiz, J.M., Rodríguez, J.J., 2006. Quantifying uncertainty in mineral resources by use of classification schemes and conditional simulations. Mathematical Geology 38(4), 445-464.

[50] Emery, X., Peláez, M., 2012. Reducing the number of orthogonal factors in linear coregionalization modeling. Computers & Geosciences 46, 149-156.

[51] Emery, X., Robles, L.N., 2009. Simulation of mineral grades with hard and soft conditioning data: Application to a porphyry copper deposit. Comput. Geosci. 13, 79–89.

[52] Emery, X., Silva, D.A., 2009. Conditional co-simulation of continuous and categorical variables for geostatistical applications. Comput. Geosci. 35, 1234–1246.

[53] Erickson, A.J., Padgett, J.T., 2011. Chapter 4.1: Geological Data Collection. In P. Darling, SME Mining Engineering Handbook, pp. 145-159.

[54] Ewusi, A., Kuma, J.S., 2011. Calibration of shallow borehole drilling sites using the electrical resistivity imaging technique in the granitoids of Central Region, Ghana. Natural Resources Research 20: 67-63.

[55] Fontaine, L., Beucher, H., 2006. Simulation of the Muyumkum uranium roll front deposit by using truncated plurigaussian method. In: Proceedings of the 6th international mining geology conference, Darwin, NT Australia, 21-23 Aug 2006.

[56] Friedman, J.H., 1987. Exploratory projection pursuit. J. Am. Stat. Assoc. 82, 249–266.

[57] Friedman, J.H., Tukey, J.W., 1974. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Comput. C-23, 881–890.

[58] Galli, A., Beucher, H., Le Loc'h, G., Doligez, B., 1994. The pros and cons of the truncated Gaussian method. In: Armstrong, M., et al. (eds.) Geostatistical simulations. Kluwer, Dordrecht, pp. 217-233.

[59] Galli, A., Gao, H., 2001. Rate of convergence of the Gibbs sampler in the Gaussian case. Mathematical Geology 33(6), 653-677.

[60] Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721-741.

[61] Gentle, J.E., 2009. Computational statistics. Springer, New York.

[62] Glacken, I.M., 1996. Change of support by direct conditional block simulation. Unpublished MSc Thesis, Stanford University.

[63] Glacken, I.M., Snowden, D.V., 2001. Mineral resource estimation. In: Edwards, A.C. (ed.) Mineral Resource and Ore Reserve Estimation - The AusIMM Guide to Good Practice. The Australasian Institute of Mining and Metallurgy, Melbourne, pp. 189-198.

[64] Godoy, M., 2009. A risk analysis based framework for strategic mine planning and design- method and application, AusIMM, pp. 21-27.

[65] Gómez-Hernández, J.J., Cassiraga, E.F., 1994. Theory and practice of sequential simulation. In: M. Armstrong and P.A. Dowd (eds.), Geostatistical simulations, Kluwer Academic Publishers, Dordrecht, Holland, pp. 111-124.

[66] Gómez-Hernández, J.J., Journel, A.G., 1992. Joint sequential simulation of multigaussian fields. Geostatistics Troia 1, Kluwer, 85-94.

[67] Goovaerts, P., 1992. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. Journal of Soil Sciences 43 (4), 597-619.

[68] Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, Oxford University Press, New York, 480 pp.

[69] Goulard, M., 1989. Inference in a coregionalization model. In: Geostatistics, Vol. 1, M., Armstrong, ed. Kluwer, Dordrecht, pp. 397-408.

[70] Goulard, M., Voltz, M., 1992. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. 24, 269–286.

[71] Gringarten, E., Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. Mathematical Geology 33(4), 507-534.

[72] Guardiano, F.B., Parker, H.M., Isaaks, E.H., 1995. Prediction of recoverable reserves using conditional simulation: a case study for the Fort Knox Gold Project, Alaska. Unpublished Technical Report, Mineral Resource Development, Inc.

[73] Guillen, A., Calcagno, P., Courrioux, G., Joly, A., Ledru, P., 2008. Geological modelling from field data and geological knowledge: Part II. Modelling validation using gravity and magnetic data inversion. Phys. Earth Planet. Inter. 171, 158–169.

[74] Haldar, S.K., 2013. Mineral Exploration: Principles and Applications. Elsevier, Oxford, 334 pp.

[75] Hassibi, M., Ershaghi, I., Aminzadeh, F., 2003. High resolution reservoir heterogeneity characterization using recognition technology. Developments in Petroleum Science 51: 289-307.

[76] Hearst, J.R., Nelson, P.H., Paillet, F.L., 2000. Well Logging for Physical Properties. Wiley, Chichester.

[77] Hoyle, I.B., 1986. Computer techniques for the zoning and correlation of well-logs. Geophysical prospecting 34(5): 648-664.

[78] Isaaks, E.H., 1990. The application of Monte Carlo methods to the analysis of spatially correlated data. PhD Thesis, Stanford University, Stanford, CA.

[79] Isaaks, E.H., 1999. SAGE2001 User's Manual, Software license and documentation. http://www.isaaks.com

[80] Isaaks, E.H., Srivastava, R.M., 1989. An introduction to applied geostatistics. Oxford University Press, New York, 561 pp.

[81] Jewbali, A., Dimitrakopoulos, R., 2009. Stochastic mine planning – example and value from integrating long - and short-term mine planning through simulated grade control, sunrise dam, Western Australia. In: Proceedings of orebody modelling and strategic mine planning, AusIMM, pp. 321-327.

[82] Journel, A.G., 1974. Geostatistics for conditional simulation of ore bodies. Econom. Geology 69, 673-687.

[83] Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics: Academic Press, London, 600 pp.

[84] Journel, A.G., Kyriakidis, P., 2004. Evaluation of mineral reserves, a simulation approach. Oxford University Press, 216 pp.

[85] Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? Math. Geol. 21, 715–739.

[86] Killeen, P.G., Elliott, B.E., 1997. Surveying the path of boreholes: a review of developments and methods since 1987. In Gubins, A.G. (ed.), Proc. of fourth decennial international conference on mineral exploration, Prospectors and Developers Assoc., Toronto, pp. 709-712.

[87] Knödel, K., Lange, G., Voigt, H.J., 2007. Environmental Geology: Handbook of Field Methods and Case Studies. Springer, Berlin, 1357 pp.

[88] Lantuéjoul, C., 2002. Geostatistical Simulation, Models and Algorithms. Springer, Berlin, 256 pp.

[89] Lantuéjoul, C., Dessassis, N., 2012. Simulation of a Gaussian random vector: a propagative version of the Gibbs sampler. In: Proceedings of the 9th International Geostatistics Congress, Oslo, Norway, June, 2012, 174–181.

[90] Larocque, G., Dutilleul, P., Pelletier, B., Fyles, J.W., 2006. Conditional Gaussian co-simulation of regionalized components of soil variation. Geoderma 134(1-2), 1-16

[91] Larrondo, P., Leuangthong, O., Deutsch, C.V., 2004. Grade estimation in multiple rock types using a linear model of coregionalization for soft boundaries. In Proceedings of the 1st International Conference on Mining Innovation; Magri, E., Ortiz, J., Knights, P., Henríquez, F., Vera, M., Barahona, C., Eds.; Gecamin Ltd.: Santiago, Chile, pp. 187–196.

[92] Le Loc'h, G., Beucher, H., Galli, A., Doligez, B., 1994. Improvement in the truncated Gaussian method: combining several Gaussian functions. In: ECMOR IV, Fourth European Conference on the Mathematics of Oil Recovery. Røros, Norway.

[93] Lelièvre, P.G., 2009. Integrating Geologic and Geophysical Data through Advanced Constrained Inversions. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada.

[94] Leuangthong, O., Deutsch, C.V., 2003. Stepwise conditional transformation for simulation of multiple variables. Math. Geol. 35, 155–173.

[95] Leuangthong, O., Hodson, T., Rolley, P., Deutsch, C.V., 2006. Multivariate geostatistical simulation at Red Dog Mine, Alaska, USA, Canadian Institution of Mining, Metallurgy, and Petroleum 99, No. 1094.

[96] Leuangthong, O., Khan, K.D., Deutsch, C.V., 2008. Solved problems in Geostatistics. John Wiley & Sons, Hoboken, NJ, 208 pp.

[97] Leuangthong, O., McLennan, J.A., Deutsch, C.V., 2004. Minimum acceptance criteria for geostatistical realizations. Nat. Resour. Res. 13, 131–141.

[98] Luster, G.R., 1985. Raw materials for Portland cement: applications of conditional simulation of coregionalization. PhD Thesis. Stanford University, Stanford, 531pp.

[99] Luthi, S.M., 2001. Geological Well Logs: Their Use in Reservoir Modeling. Springer, Berlin, 373 pp.

[100] Luthi, S.M., Bryant, I.D., 1997. Well-log correlation using a back-propagation neural network. Mathematical Geology 29(3): 413-425.

[101] Madani, N., Emery, X., 2018. A comparison of search strategies to design the cokriging neighborhood for predicting coregionalized variables. Stochastic Environmental Research and Risk Assessment, in press.

[102] Maleki, M., Emery, X., 2015. Joint simulation of grade and rock type in a stratabound copper deposit. Mathematical Geosciences 47: 471-495.

[103] Maleki, M., Emery, X., 2017. Joint simulation of stationary grade and non-stationary rock type for quantifying geological uncertainty in a copper deposit. Comput. Geosci. 109, 258–267.

[104] Maleki, M., Emery, X., Cáceres, A., Ribeiro, D., Cunha, E., 2016. Quantifying the uncertainty in the spatial layout of rock type domains in an iron ore deposit. Comput. Geosci. 20, 1013–1028.

[105] Maleki, M., Emery, X., Mery, N., 2017. Indicator Variograms as an Aid for Geological Interpretation and Modeling of Ore Deposits. Minerals 7, 241.

[106] Manchuk, J.G., Deutsch, C.V., 2012. Applications of data coherency for data analysis and geological zonation. In: Abrahamsen, P., Hauge, R., Kolbjørnsen, O. (eds.) Geostatistics Oslo 2012. Springer, pp. 173-184.

[107] Mantoglou, A., Wilson, J.L., 1982. The turning bands method for simulation of random fields using line generation by a spectral method. Water resources research 18(5), 1379-1394.

[108] Mariethoz, G., Caers, J., 2014. Multiple-Point Geostatistics: Stochastic Modeling with Training Images. Wiley: New York, NY, USA, p. 376.

[109] Mariethoz, G., Renard, P., Cornaton, F., Jaquet. O., 2009. Truncated plurigaussian simulations of aquifer heterogeneity. Ground Water 47(1), 13-24.

[110] Marjoribanks, R., 2010. Geological Methods in Mineral Exploration and Mining. Springer, Berlin, 238 pp.

[111] Matheron, G., 1973. The intrinsic random functions and their applications. Advances applied probability 5, 439-468.

[112] Mejia, J.M., Rodríguez-Iturbe, I., 1974. On the synthesis of random field sampling from the spectrum: An application to the generation of hydrologic spatial processes. Water Resources Research 10(4), 705-711.

[113] Mery, N., Emery, X., Cáceres, A., Ribeiro, D., Cunha, E., 2017. Geostatistical modeling of the geological uncertainty in an iron ore deposit. Ore Geol. Rev. 88, 336–351.

[114] Mitchell, P., 2006. Guidelines for Quality Assurance and Quality Control in Surface Water Quality Programs in Alberta. Pub. No: T/884, ISBN: 0-7785-5081-8 (Printed Edition) 57 pp.

[115] Moon, C.J., Whateley, M.K.G., Evans, A.M., 2006. Introduction to Mineral Exploration. Blackwell Scientific Publications, Oxford, 481 pp.

[116] Myers, D.E., 1991. Pseudo-cross variograms, positive-definiteness, and cokriging. Mathematical Geology 23, 805-816.

[117] Olea, R., editor. 1991. Geostatistical Glossary and Multilingual Dictionary. Oxford University Press, New York.

[118] Ortiz, J.M., Emery, X., 2006. Geostatistical estimation of mineral resources with soft geological boundaries: A comparative study. J. S. Afr. Inst. Min. Metall, 106, 577–584.

[119] Papoulis, A., 1984. Probability, Random Variables, and Stochastic Processes. McGraw-Hill: New York, NY, USA.

[120] Pawlowsky-Glahn, V., Buccianti, A., 2011. (Eds.) Compositional Data Analysis: Theory and Applications. Wiley: New York, NY, USA, p. 400.

[121] Ripley, B.D., 1987. Stochastic Simulation. John Wiley & Sons, New York.

[122] Roberts, G.O., Sahu, S.K., 1997. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. Journal of the Royal Statistical Society B59(2), 291-317.

[123] Rondon, O., 2009. A look at the Plurigaussian simulation for a nickel laterite deposit. Presented at the 7th international mining & geology conference, Perth.

[124] Rossi, M.E., 1999. Optimizing grade control: a detailed case study. In: Proceedings of the 101st annual meeting of the Canadian Institute of Mining, Metallurgy, and Petroleum (CIM), Calgary (May 2-5).

[125] Rossi, M.E., Camacho, V.J., 2001. Applications of geostatistical conditional simulations to assess resource classification schemes. Proceedings of the 102nd annual meeting of the Canadian Institute of Mining, Metallurgy, and Petroleum (CIM), Quebec City (April 29-May 2).

[126] Rossi, M.E., Deutsch, C.V., 2014. Mineral Resource Estimation. Springer, Heidelberg, 332 pp.

[127] Séguret, S.A., 2013. Analysis and estimation of multi-unit deposits: Application to a porphyry copper deposit. Math. Geosci. 45, 927–947.

[128] Serra, J., 1968. Les structures gigognes: morphologie mathématique et interprétation métallogénique. Mineralium Deposita 3, 135-154.

[129] Shinozuka, M., 1971. Simulation of multivariate and multidimensional random processes. The Journal of the Acoustical Society of America 49(1B), 357-367.

[130] Silva, D.S.F., Deutsch, C.V., 2016. Multivariate data imputation using Gaussian mixture models. Spat. Stat. doi:10.1016/j.spasta.2016.11.002.

[131] Sinclair, A.J., Blackwell, G.H., 2002. Applied Mineral Inventory Estimation. Cambridge University Press, Cambridge, 400 pp.

[132] Sinclair, A.J., Vallée, M., 1993. Improved sampling control and data gathering for improved mineral inventories and production control. In Dimitrakopoulos, R., (ed.), Proc. Symp. On "Geostatistics for the next century", June 3–5, Montreal, Kluwer Acad. Pub., The Netherlands, pp. 323-329.

[133] Skvortsova, T., Armstrong, M., Beucher, H., Forkes, J., Thwaites, A., Turner, R., 2000. Applying plurigaussian simulations to a granite-hosted orebody. In: Kleingeld, W., Krige, D.G., (eds.) Geostats 2000 Cape Town. Proceedings of the 6th international Geostatistics Congress, held in Cape Town 10-14 April 2000, pp. 904-911.

[134] Skvortsova, T., Armstrong, M., Beucher, H., Forkes, J., Thwaites, A., Turner, R., 2002. Simulating the geometry of a granite-hosted uranium orebody. In: Armstrong, M., et al., (eds.) Geostatistics Rio 2000. Kluwer, Dordrecht, Holland, pp. 85-100.

[135] Soleimani, M., Shokri, B.H., Rafiei, M., 2016. Integrated petrophysical modeling for a strongly heterogeneous and fractured reservoir, Sarvak Formation, SW Iran. Natural Resources Research, doi:10.1007/s11053-016-9300-9.

[136] Soltani, S., Hezarkhani, A., 2011. Determination of realistic and statistical value of the information gathered from exploratory drilling. Natural Resources Research 20: 207-216.

[137] Spies, B.R., 1996. Electrical and Electromagnetic Borehole Measurements: A Review. Surveys in Geophysics 17: 517-556.

[138] Statistics Canada website:
http://www.statcan.ca/english/edu/power/ch3/quality/quality.htm
http://www.statcan.gc.ca/

[139] Taylor, G.R., 2000. Mineral and lithology mapping of drill core pulps using visible and infrared spectrometry. Natural Resources Research 9: 257-268.

[140] Theys, P., 1999. Log Data Acquisition and Quality Control. Editions Technip, Paris, 480 pp.

[141] USEPA (United States Environmental Protection Agency), 1996. The volunteer monitor's guide to Quality Assurance Project Plans, EPA: 841-B-96-003. http://www.epa.gov/owow/monitoring/volunteer/qappcovr.htm

[142] Vallée, M., 1992. Guide to the evaluation of gold deposits. Canadian Institute of Mining, Metallurgy and Petroleum, Special v. 45, 299 pp.

[143] Vallée, M., 1998. Quality assurance, continuous quality improvement and standards. Exploration and Mining Geology 7, 1-14.

[144] Van den Boogaart, K.G., Mueller, U., Tolosana-Delgado, R., 2016. An affine equivariant multivariate normal score transform for compositional data. Math. Geosci. 49, 231–251.

[145] Vargas-Guzmán, J.A., 2008. Transitive geostatistics for stepwise modeling across boundaries between rock regions. Math. Geosci. 40, 861–873.

[146] Verly, G., 1983. The multigaussian approach and its application to the estimation of local reserves. Mathematical Geology 15(2): 259-286.

[147] Verly, G., 2005. Grade control classification of ore and waste: A critical review of estimation and simulation based procedures. Mathematical Geology 37(5), 451- 475.

[148] Wackernagel, H., 2003. Multivariate Geostatistics: an Introduction with Applications. Springer, Berlin, 387 pp.

[149] Wackernagel, H., 2013. Course Notes on Geostatistics. Mines Paristech.

[150] Witten, I.H., Frank, E., Hall, M., Pal, C., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Elsevier: Burlington, MA, USA, p. 654.

[151] Xu, C., Dowd, P.A., Mardia, K.V., Fowell, R.J., 2006. A flexible true plurigaussian code for spatial facies simulations. Comput. Geosci. 32, 1629–1645.

[152] Xu, W., Tran, T.T., Srivastava, R.M., and Journel. A.G., 1992. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In: Proceedings of the 67th annual technical conference of the society of petroleum engineers, Washington, no. 24742 in SPE, pp. 833-842.

[153] Yunsel, T.Y., Ersoy, A., 2011. Geological modeling of gold deposit based on grade domaining using plurigaussian simulation technique. Natural Resources Research 20: 231-249.

[154] Yunsel, T.Y., Ersoy, A., 2013. Geological modeling of rock type domains in the Balya (Turkey) lead-zinc deposit using plurigaussian simulation. Central European Journal of Geosciences 5(1), 77-89.